

Original Article

An evaluation of the species and subspecies of the genus *Salmonella* with whole genome sequence data: Proposal of type strains and epithets for novel *S. enterica* subspecies VII, VIII, IX, X and XI

Madison E. Pearce^{a,b}, Gemma C. Langridge^c, A.C. Lauer^d, Kathie Grant^e, Martin C.J. Maiden^{a,b}, Marie A. Chattaway^{e,*}

^a Department of Zoology, University of Oxford, Peter Medawar Building for Pathogen Research, South Parks Road, Oxford OX1 3SY, United Kingdom

^b National Institute for Health Research, Health Protection Research Unit, Gastrointestinal Infections, University of Oxford, United Kingdom

^c Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, United Kingdom

^d Centers for Disease Control and Prevention, Enteric Diseases Laboratory Branch, 1600 Clifton RD NE, Atlanta, GA 30329, USA

^e Public Health England, Gastrointestinal Bacteria Reference Unit, 61 Colindale Avenue, London NW9 5EQ, United Kingdom

ARTICLE INFO

Keywords:

Salmonella
Genomics
Londinensis
Brasilensis
Hibernicus
Essexiensis
Reptilium

ABSTRACT

Species and subspecies within the *Salmonella* genus have been defined for public health purposes by biochemical properties; however, reference laboratories have increasingly adopted sequence-based, and especially whole genome sequence (WGS), methods for surveillance and routine identification. This leads to potential disparities in subspecies definitions, routine typing, and the ability to detect novel subspecies. A large-scale analysis of WGS data from the routine sequencing of clinical isolates was employed to define and characterise *Salmonella* subspecies population structure, demonstrating that the *Salmonella* species and subspecies were genetically distinct, including those previously identified through phylogenetic approaches, namely: *S. enterica* subspecies *londinensis* (VII), subspecies *brasilensis* (VIII), subspecies *hibernicus* (IX) and subspecies *essexiensis* (X). The analysis also identified an additional novel subspecies, *reptilium* (XI). Further, these analyses indicated that *S. enterica* subspecies *arizonae* (IIIa) isolates were divergent from the other *S. enterica* subspecies, which clustered together and, on the basis of ANI analysis, subspecies IIIa was sufficiently distinct to be classified as a separate species, *S. arizonae*. Multiple phylogenetic and statistical approaches generated congruent results, suggesting that the proposed species and subspecies structure was sufficiently biologically robust for routine application. Biochemical analyses demonstrated that not all subspecies were distinguishable by these means and that biochemical approaches did not capture the genomic diversity of the genus. We recommend the adoption of standardised genomic definitions of species and subspecies and a genome sequence-based approach to routine typing for the identification and definition of novel subspecies.

1. Introduction

Defining prokaryotic species has long been controversial [1]. The most widely accepted approach is for organisms within a species to have (i) at least one consistent distinguishable phenotypic trait and (ii) at least 70% DNA-DNA hybridisation (DDH) [2], combining both the phenotype and genotype of isolates. Within the genus *Salmonella*, isolates are classified through a range of tests that determine biochemical properties [3]; however, conventional *Salmonella* nomenclature and taxonomy has presented problems [4,5]. Some clarity was achieved in

2005 through a judicial opinion [6] and an accompanying commentary [7], which split the genus into two species, *Salmonella bongori* (*S. bongori*) [6,8], (previously known as *S. enterica* subspecies V) and *Salmonella enterica* (*S. enterica*) [6]. At the time of writing, *S. enterica* was divided into six subspecies [7]: *Salmonella enterica* subspecies I (*enterica*); *Salmonella enterica* subspecies II (*salamae*); *Salmonella enterica* subspecies IIIa (*arizonae*); *Salmonella enterica* subspecies IIIb (*diarizonae*); *Salmonella enterica* subspecies IV (*houtenae*); and *Salmonella enterica* subspecies VI (*indica*) [7]. The phenotypic distinctions among the subspecies have been further confirmed by molecular [8,9] and genetic [5,9] tests.

* Corresponding author.

E-mail addresses: gemma.langridge@quadram.ac.uk (G.C. Langridge), ALauer@cdc.gov (A.C. Lauer), Martin.Maiden@zoo.ox.ac.uk (M.C.J. Maiden), marie.chattaway@phe.gov.uk (M.A. Chattaway).

<https://doi.org/10.1016/j.ygeno.2021.07.003>

Received 18 December 2020; Received in revised form 25 June 2021; Accepted 2 July 2021

Available online 7 July 2021

0888-7543/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Prior to this opinion [6] and commentary [7], many of the *Salmonella* subspecies and serovars had at times been considered separate species [6,7] or alternatively grouped together as one subspecies including *Salmonella arizonae* and *diarizonae* [10].

A multilocus enzyme electrophoresis (MLEE) analysis [9] demonstrated the presence of a genetically distinct group, *Salmonella enterica* subspecies VII, which was later confirmed by genetic analyses [5]. As biochemical approaches identified isolates from this group as belonging to subspecies IV [9], subspecies VII has not been formally recognised [3]. A further 3 novel subspecies have also been identified, using a core genome SNPs analysis (novel A, B, and C), but at the time of writing no biochemical data had been published on these possible subspecies to ensure backward compatibility [11]. Evolutionary analysis of whole genome sequencing (WGS) data have confirmed the existence of distinct species and subspecies groups and demonstrated that extensive within-genus hybridisation has occurred during the emergence of these groups, such that their evolution cannot be modelled solely with bifurcating phylogenies [12]; however, there remains a need for reference laboratories to have a robust widely-recognised nomenclature system.

The lack of formal recognition of subspecies VII highlights the problems of a typing scheme based on biochemical characteristics [3]. It is now recognised that biochemical and microbiological subspecies identification methods are incapable of representing the genetic relationships accurately among strains [5,13]. For example, such methods require microbiological expertise across a wide range of biochemical tests and are low throughput [14]. Additionally, reference laboratories, such as Public Health England (PHE) in the UK and the US Centres for Disease Control and Prevention (CDC), have increasingly replaced biochemical approaches with whole genome sequencing (WGS), which allows greater resolution of the genetic relationships among isolates [15–21]. The WGS approach of core genome multilocus sequence typing (cgMLST) has been adopted as the subtyping replacement for Pulsed-field gel electrophoresis by PulseNet [22], so that global monitoring of foodborne disease can rely on this technology [22].

Whilst the move toward WGS-based subtyping allows for increased understanding and more accurate reporting, the majority of research and publications have focused on subspecies I, which is responsible for most *Salmonella* infections in humans [15,23]. However, the other subspecies are also responsible for cases of human disease [24], being opportunistic pathogens [25] that infect humans under certain conditions. Most reported cases have occurred within vulnerable individuals, such as the young [24], elderly [26], and immunocompromised [27]. Non-subspecies I *Salmonella* have been isolated from animal sources including exotic reptilian pets [28], wild [29,30] and zoo animals [31], chickens [32], livestock [33,34], and farmed reptiles [35]. The primary hosts for non-subspecies I isolates are thought to be reptiles, as isolates have been found in snake, crocodile, lizard, and turtle species [28,35,36]. Reptile farms from across the globe, including within the United Kingdom, have had non-subspecies I salmonellae isolated from them [37].

When humans are in close contact with reptiles, the chances of transmission of non-subspecies I serovars increases [30]. In French Guiana, a correlation between *Salmonella* serotypes in patients and those found in local reptile populations was found, with two-thirds of reptile serotypes also isolated from patients [30]. A total of 16.7% of reptiles were positive for subspecies IV strains, which are responsible for ~10% of human salmonellosis cases. Reptiles are thought to be a natural reservoir for non-subspecies I organisms [36,38] and its estimated that more than 1% of households in the USA and Europe are now home to either a reptile or amphibian [31,39]. This increase in exotic pet ownership could lead to an increase in cases of salmonellosis from members of subspecies II – VI. For example, one study examining 16 pet snakes found 13 of them were infected with subspecies IIb strains [38] and another found subspecies II strains in 8 out of 80 pet turtles, in the majority of animals more than one strain was identified [36].

The need for the characterisation of the whole of the *Salmonella*

genus is accentuated by the focus on subspecies I, which skews analyses of the other subspecies [40], and by the difficulties of conventional biochemical approaches to subspecies definition and identification [13] within *Salmonella*. However, with reference laboratories adopting WGS [15,41] and increased sequencing sharing through databases [11,42], a large scale examination of the complete *Salmonella* genus is now possible. Here, a wide-ranging analysis of isolates from across the *Salmonella* genus was undertaken, with the aim of identifying and defining a genome-based approach to *Salmonella* species and subspecies definition and identification. All isolates in this study were analysed in a search for previously unidentified novel subspecies, using both genetic and biochemical approaches. Additionally, all confirmed species and subspecies as well as previously identified novel subspecies (VII, novel A, B, and C) were re-examined using statistical and phylogenetic approaches.

2. Materials and methods

2.1. Isolate collection and curation

2.1.1. Non-subspecies I isolates

A total of 1530 previously published, confirmed non-subspecies I isolates [40], were used in this analysis. This dataset contained: 278 subspecies II; 308 subspecies IIIa; 442 subspecies IIIb; 326 subspecies IV; 28 subspecies IV; 22 subspecies VII; 3 subspecies VIII; 48 subspecies IX; 19 subspecies X; and 56 *S. bongori* (Table S1) [40]. The subspecies of these isolates were confirmed using phenotypic data and phylogenetic clustering or reanalysis within a reference laboratory. For all of these isolates the subspecies provided for the metadata stored in Enterobase by the original submitter matched with their genetic clustering, with the exception of SAL_MA5841AA (Enterobase ID) [40].

All genomic data used in the study is available either available in NCBI BioProject PRJNA248792 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA248792>), the European Nucleotide Archive (ENA Browser (ebi.ac.uk)) or Enterobase (Enterobase (warwick.ac.uk)).

2.1.2. Subspecies I isolates

As subspecies I is overrepresented within the available genome databases, a representative subset was identified for analysis. This subset consisted of the twenty *S. enterica* subspecies I serovars most commonly identified in the UK at Public Health England (PHE). To assemble this dataset, eBurst Groups (eBGs), based on multilocus sequence typing (MLST) [43] were used. This approach linked identical sequence types and single-locus variants to create groups of genetically closely related isolates [43]. If a single eBG group contained isolates from only one serovar and represented more than 10% of the isolates within that serovar, then up to 10 isolates from that eBG were randomly chosen from PHE isolates. PHE isolates were used as they had been validated by a reference laboratory and they were available for reanalysis if in case of any discrepancies among the various datasets. If there were fewer than 10 isolates available from PHE, then as many as were available were used. This resulted in a subset of 271 subspecies I representative isolates (Table S1).

2.1.3. Identification of novel subspecies

The 1530 non-subspecies I isolates and the 271 subspecies I isolates were used to construct rMLST [44] and cgMLST [40] minimal spanning trees (MStrees) in GrapeTree [45], indexing allelic differences. These MStrees were used to identify potential novel subspecies, defined as groups of three or more isolates that did not merge with any other subspecies when the branches in the minimum spanning trees were collapsed. Groups of three or more isolates were chosen to ensure the presence of a real subspecies and not the result of a mixed sample or a single highly mutated strain. To identify novel subspecies, branches were collapsed until the nodes of two previously identified subspecies merged or multiple phylogenetically identified subspecies merged with

a confirmed subspecies, creating a single node. When a potential new subspecies was identified, Enterobase [11] was interrogated for other isolates that were closely related to these novel subspecies, which may have been uploaded without metadata or incorrect metadata and were therefore missed in the initial searches for non-subspecies I isolates. All novel subspecies were assigned the next available Roman numeral; previously detected novel subspecies A, B, and C [11] were therefore assigned subspecies VIII, IX, and X and any further detection of novel subspecies in this study were assigned XI onwards.

2.2. Datasets

Dataset A comprised 569 isolates representative of the diversity of *Salmonella* [6,7] and were used to reconstruct rMLST and cgMLST neighbouring joining trees [46] (NJTree), and a rMLST Maximum Likelihood (ML) phylogeny [47], and for the STRUCTURE analysis of the rMLST loci. The dataset consisted of: 29 *S. bongori* isolates; 100 *S. arizonae*; 95 subspecies I; 100 subspecies II; 100 subspecies IIIb; 100 subspecies IV; 16 subspecies VI; 5 subspecies VII [9]; 3 subspecies VIII; 8 subspecies IX; 10 subspecies X; and 3 subspecies XI. Isolates were chosen from the 1801 isolates to represent the ribosomal sequence types (rSTs) within each species or subspecies. Up to 100 isolates or as many as were available, all with different rSTs, were randomly chosen or as many as were available, with the exception of subspecies XI where two isolates with the same rST were used as only three isolates were available in total

(Table S2).

Dataset B comprised 77 isolates and was used for cgMLST STRUCTURE analysis. The 77 isolates represented seven randomly chosen examples of each suspected subspecies, except subspecies VIII, and XI, for which only three isolates were available. Isolate SAL_BA7507AA (Enterobase ID) was also included in this analysis, as it clustered with different subspecies when phylogenetic trees were reconstructed using rMLST and cgMLST sequence data. The subspecies I isolates were chosen to represent diversity, with three random examples from clades A and B [43] and 2 from the Typhi/Paratyphi A lineage [48,49], one Typhi and one Paratyphi A) (Table S2).

Dataset C comprised 25 isolates which were chosen at random for ANI comparisons. For the confirmed subspecies, type strains approved by the Judicial Commission of the International Committee on Systematics of Prokaryotes [6] were used for between species and subspecies comparisons. This included the type strain for *S. bongori*, V (NCTC12419), the type strain for *S. arizonae*, IIIa (NCTC8297) and type strains for each of the currently defined subspecies within the species *S. enterica*: *S. enterica* Serovar Typhimurium, I (NCTC 12416), *S. salamae*, II (NCTC5773), *S. diarizonae*, IIIb (NCTC10060), *S. houtenae*, IV (NCTC 12418) and *S. indica*, VI (NCTC12420) which had previously been defined as type strains and deposited in the NCTC culture collection. Strains of *S. enterica* subspecies VII, VIII, IX, X, and XI were also included and one of each subspecies was proposed as a type strain (Table 1).

Three isolates were used for subspecies I, one isolate each from

Table 1
Overview of Type strains used and proposed in this study with subspecific epithets.

PubMLST_id	Reference Name	Enterobase_ID	NCTC accession number	Other collections reference ^a	Year of isolation ^a	Country of origin ^a	Source of isolation ^a	Species/subspecies No.	Species epithet	Subspecies epithet/proposed epithet
17,191	M328/ NCTC20044	SAL_HA8868AA	NCTC 12416	ATCC 43971; CIP 60.62; DSM 17058; NCIMB 11450, JCM 32817	<1960	New York, USA	Unknown	I	<i>enterica</i>	<i>enterica</i> (serovar Typhimurium)
16,394	NCTC5773	SAL_HA8861AA	NCTC 5773	ATCC 43972; CIP 82.29; DSM 9220, CCUG 30039	1939	Copenhagen, Denmark	Unknown	II	<i>enterica</i>	<i>salamae</i>
16,389	NCTC10060	SAL_HA8897AA	NCTC 10060	ATCC 43973; CCUG 30040; CIP 82.31; DSM 14847; PESO M11, CCUG 30040	1947	Kentucky, USA	Unknown	IIIb	<i>enterica</i>	<i>diarizonae</i>
16,391	NCTC12418	SAL_HA8832AA	NCTC 12418	ATCC 43974; DSM 9221; CCUG 30041, CIP 82.32	<1982	Unknown	Unknown	IV	<i>enterica</i>	<i>houtenae</i>
16,393	NCTC12420	SAL_HA8831AA	NCTC 12420	ATCC 43976; CCUG 30038; CIP 102501; DSM 14848; K1240	1986	Unknown	Unknown	VI	<i>enterica</i>	<i>indica</i>
16,108	267,055	SAL_MA2040AA	NCTC 10415	*	1964	London	Unknown	VII	<i>enterica</i>	<i>londinensis</i>
16,111	85–0120	SAL_JA5200AA	NCTC 14236	*	1984	Brazil	Water	VIII	<i>enterica</i>	<i>brasiliensis</i>
15,996	304,467	SAL_MA5929AA	NCTC 14237	*	2016	Ireland	Human	IX	<i>enterica</i>	<i>hibernicus</i>
16,160	242,410	SAL_MA5883AA	NCTC 14239	*	2016	Essex	Human	X	<i>enterica</i>	<i>essexiensis</i>
16,148	267,042	SAL_MA5907AA	NCTC 10436	*	1965	South Africa	Lizard	XI	<i>enterica</i>	<i>reptilium</i>
16,406	NCTC8297	SAL_HA8865AA	NCTC 8297	ATCC 13314; CIP 8230; CN 4247; DC5; DSM 9386	<1953	Arizona, USA	Unknown	IIIa	<i>arizonae</i>	–
16,392	NCTC12419	SAL_HA8836AA	NCTC12419	ATCC 43975; CCUG 30042; CIP 82.33; DSM 13772	<1982	USA	Human	V	<i>bongori</i>	–

Table containing the type strains of the isolates used in this study. For the novel subspecies, type strains were recommended and submitted to NCTC.

^a Information of previously defined strains obtained from culture collection data sheets. *also submitted to DSMZ (German Collection of Microorganisms and Cell Cultures) at time of publication

clades A and B [48] and one from the Typhi/Paratyphi A lineage [48,49], enabling an average to be taken for within subspecies I comparisons (Table S2).

Dataset D comprised 25 randomly chosen isolates used for ANI analysis. Two isolates of *S. bongori*, *S. arizonae*, subspecies II, IIIB, IV, VI, VII, VIII, IX, X, and XI respectively were chosen at random. Three isolates of subspecies I were randomly chosen, one representative each from clades A and B [48] and one from the Paratyphi A/Typhi [48,49] lineage (Table S2).

Dataset E comprised 22 representative genomes chosen from the PHE archives for a phenotypic analysis of all of the species and subspecies. Two isolates were used for each of *S. bongori*, *S. arizonae*, *S. enterica*, subspecies I, II, IIIB, IV, VI, IX, and X, with one isolate was used for subspecies VII, VIII, and XI given that only one example of these subspecies were available within PHE archives. Isolate SAL_MA5841AA, which grouped phylogenetically with subspecies I but which was typed as subspecies II using microbiological approaches, was also included for further analysis (Table S2).

2.3. Phylogenetic analyses

The rMLST alleles of the isolates in dataset A (Table S2) were used to reconstruct a phylogeny using the Neighbour joining algorithm [46]. The NJtree was generated using the Genome Comparator tool [42] from the Bacterial Isolate Genome Sequence database (BIGSdb) [42] available through the PubMLST website. The tree was rooted using the *S. bongori* species as an outgroup and annotations were performed using the Interactive Tree of Life (ITOL) online tool [50]. For comparison an NJtree [46] using the SalmcgmLSTv1.0 [40] was also created. In order to confirm the accuracy of the NJtree [46] the rMLST sequences were used to reconstruct a ML phylogeny [47]. The rMLST sequences of dataset A (Table S2) were exported from BIGSdb [42] after being aligned using MAFFT [51]. The ML phylogeny [47] was reconstructed using MEGA7 [52] with 100 bootstrap replications calculated using the Kimura 2-parameter model [53]. The tree was rooted using the *S. bongori* species as an outgroup and annotations were performed using the ITOL online tool [50].

2.4. Statistical analyses

STRUCTURE [54,55]: The STRUCTURE algorithm can infer population structure accurately when compared with expected phylogenetic results [54], identifying populations from and assigning an individual to the population to which it is most similar. The algorithm analyses differences in the distribution of variants within allelic profiles of a dataset to create genetic clusters. Individuals that share similar variation patterns are grouped by a Bayesian iterative algorithm. STRUCTURE applies Markov Chain Monte Carlo (MCMC) estimations, which initially assigning individuals to a random group and then reassigns them based on estimations on variant frequency [54]. STRUCTURE was applied to both rMLST [44] and cgMLST [11] [11] data, using the allele frequencies independent model, as the subspecies represented genetically diverse and distinct populations [9,11]. For both analyses, the maximum number of populations assumed (K) was increased from twelve until no further populations were observed and until lower order taxonomical groups were observed, such as the splitting of subspecies I into its respective clades [48,49].

The rMLST STRUCTURE analysis was performed on dataset A (Table S2), based on 51/53 ribosomal genes. The rMLST loci BACT000060 (*rpmE*) and BACT000065 (*rpmJ*), were excluded due to the existence of paralogous loci. The rMLST analysis was run with a burn-in of 100,000 and 150,000 MCMC repetitions. The output was edited and annotated using the DISTRUCT [56] [51] software. The cgMLST analysis was based on 2746/2750 core genes. Four (STMMW_36391, STMMW_00801, STMMW_44771 and STMMW_05221) were removed from the analysis, due to the presence of paralogs in more than three of the genomes

analysed. If a locus were paralogous in two or fewer genomes then the paralogous loci were replaced with the value for a missing locus. The cgMLST analysis was conducted on dataset B (Table S2), with a burn-in of 100,000 and 200,000 MCMC repetitions.

Average Nucleotide Identity (ANI) [1]: An online ANI calculator with the OrthoANIu [57] algorithm was used. This algorithm was chosen because the ANI scores between two genomes can vary when reciprocal calculations are compared [57]. Instead, OrthoANI fragments both genomic sequences and only uses the orthologous pairs to calculate the nucleotide identities. OrthoANI is highly comparable with traditional ANI but the values are typically approximately 0.1% higher [57].

The ANI values were calculated by comparing isolates from dataset C (Table S2). Pairs of isolates with ANI values higher than 95% [58,59] were considered to be from the same species. ANI analyses to identify species with the 95% criteria were performed on the entire genome sequence. As there was no percentage cut-off for subspecies, this comparison was used to identify the difference in values within and between the subspecies. The analysis was repeated using the same genomes at the levels of cgMLST and rMLST. Analysis of the WGSs of randomly chosen representatives of the subspecies (dataset D [Table S2]) were also performed to confirm the results obtained using type strains. The program OrthoANI was used to analyse isolates SAL_MA5841AA and SAL_BA7507AA against dataset C and dataset D isolates.

2.5. Phenotypic analyses

Metabolic phenotypes of the isolates in dataset E (Table S2), were analysed using a Biolog assay (Technopath). *Salmonella* isolates were streaked onto blood agar plates (Oxoid) and incubated at 37 °C overnight. A single colony pick was inoculated using a flat end cotton swab into type A broth (Technopath) and mixed thoroughly. Each well of a Biolog Gen III 96 -well MicroPlate, which contained metabolites or a negative control (<https://biolog.com/products-portfolio-overview/microbial-identification>) was inoculated using 100 µl of the broth and placed into the Biolog at an incubation temperature of 33 °C. The plate was read every 15 min for 48 h to record metabolic respiration. Runs were repeated independently 3 times.

The results were analysed by calculating the signal value [60] for each run. The signal value equalled the ((average signal over 48 h + maximum signal over 48 h)/2) – (the average signal over the first 2 h) [60]. An average signal value for each subspecies was calculated via RStudio [60] and used in order to create a heatmap of substrate utilisation for metabolic phenotyping and a dendrogram based on metabolic respiration values. The heatmap and dendrogram of metabolic phenotyping (including substrate utilisation) were generated to assess if utilisation of biochemical properties could be used to distinguish all of the *Salmonella* species and subspecies.

3. Results

3.1. Identification of novel subspecies

The phenotypic data for the subspecies I isolates, and phylogenetic clustering were fully congruent among the preliminary analyses as subspecies I (Fig. S1), with the exception of isolate SAL_MA5841AA, which was phenotypically identified as subspecies II, but which clustered with subspecies I isolates phylogenetically. MStrees containing all of these isolates were reconstructed using sequence data from rMLST loci (Fig. S1A) and cgMLST loci (Fig. S1B). Within the rMLST MStree, when branches were collapsed to 20/51 allelic differences (Fig. S1C), the nodes representing subspecies II and IIIB merged, whereas within the cgMLST MS tree (Fig. S1B) at 2640/2750 allelic differences subspecies VII and novel A (VIII) merged with subspecies IV (Fig. S1D). Any cluster of three or more isolates that did not cluster with any other subspecies and was present in both the rMLST and cgMLST trees was identified as a potential novel subspecies. There was one cluster of subspecies II

isolates which met these criteria (isolates SAL_DA4561AA, SAL_E-A2466AA and SAL_MA5907AA), these isolates were therefore identified as a novel subspecies, subspecies XI. Species *S. bongori* and *S. enterica* were distinct, as were subspecies I, II, IIIa, IIIb, IV, VI, VII, VIII, IX, X, and XI; however, subspecies IIIa was identified as a separate species, which we propose be called *Salmonella arizonae*.

3.2. Phylogenetic analyses

Both the rMLST (Fig. 1a) and cgMLST NJTrees [41] (Fig. 1b) of dataset A indicated that *S. bongori* was divergent from all of the other isolates, although the species did not form a single monophyletic sister clade in the cgMLST tree. *S. arizonae* (formerly *S. enterica* subspecies IIIa) was distinct from all other *S. enterica* subspecies, also forming a distinct sister clade (Fig. 1b). These trees showed that all of the *S. enterica* subspecies identified through conventional phenotypic methods (II, IIIb, IV, and VI) and those identified through phylogenetic approaches (VII [MLEE], VIII, IX, X [SNP-based MLtree], and XI [MStree]) formed distinct monophyletic groups. In both of the trees, subspecies IV, VII, and VIII formed a clade of closely related subspecies and all of the subspecies remained monophyletic. These results were highly congruent with previous analyses [5,11,61], demonstrating consistency within the identification of *Salmonella* subspecies, despite the use of different phylogenetic methods. The overall topologies of the two trees were also highly congruent, with both cgMLST and rMLST capable of subspecies discrimination. One isolate (SAL_BA7507AA) did not cluster as expected within the rMLST tree: SAL_BA7507AA was identified through phenotypic methods as a subspecies II isolate but genetically clustered with subspecies XI. Within the cgMLST tree, SAL_BA7507AA clustered with subspecies II, but formed a divergent sister clade.

In order to confirm the results observed within the NJtrees, a MLtree [47] was reconstructed for the rMLST profiles (Fig. S2). The topologies of the NJtree and the MLtree were highly congruent: *S. bongori* was divergent from all other isolates and *S. arizonae* was divergent from all

the *S. enterica* subspecies. Within the rMLST MLtree, all of the subspecies nodes were highly supported with bootstrap values of 0.89 or greater.

3.3. Statistical analyses

STRUCTURE: STRUCTURE analyses using both rMLST and cgMLST loci (Fig. 2) also supported the distinct species status of *S. bongori*, *S. arizonae*, and *S. enterica*, along with the integrity of the *S. enterica* subspecies I, II, IIIb, IV, VI, VII, VIII, IX, X, and XI. All of the species and subspecies formed distinct and consistent clades. STRUCTURE analysis of the rMLST loci of dataset A identified twelve distinct groups (Fig. 2a). The species *S. bongori*, *S. arizonae* and the *S. enterica* subspecies I, II, IIIb, IV, VI, VII, VIII, IX, X, and XI were all identified as distinct populations. When the number of populations (K) was set to twelve, the species *S. bongori*, *S. arizonae*, and *S. enterica* subspecies I, II, IIIb, IV, VI, VII, IX, and X resolved into single populations with very little admixture. In comparison, subspecies VIII and XI were comprised of multiple rMLST populations and as such showed noticeable admixture. *S. enterica* subspecies VIII was comprised of an admixture of subspecies I, IV, and VII that were all present in similar proportions. Subspecies XI was an admixture of subspecies II and a novel population. Within the rMLST STRUCTURE analysis, isolate SAL_BA7507AA clustered with the other subspecies II isolates. At an increased value of K, no further populations were observed, only increased admixture among the twelve subspecies groups was seen. The results of this STRUCTURE analysis supported the presence of the twelve distinct groups, species or subspecies, and the phylogenies created from the same datasets.

By comparison, the STRUCTURE algorithm performed on the cgMLST loci of dataset B isolates identified thirteen distinct groups (Fig. 2b). At a K of eighteen, STRUCTURE identified species *S. bongori*, *S. arizonae* and the *S. enterica* subspecies I, II, IIIb, IV, VI, VII, VIII, IX, X, and XI; however, the *S. enterica* subspecies I isolates from the Typhi/Paratyphi A lineage [43,44] were admixed between the subspecies I population and a distinct population. One isolate of subspecies VIII also exhibited a small

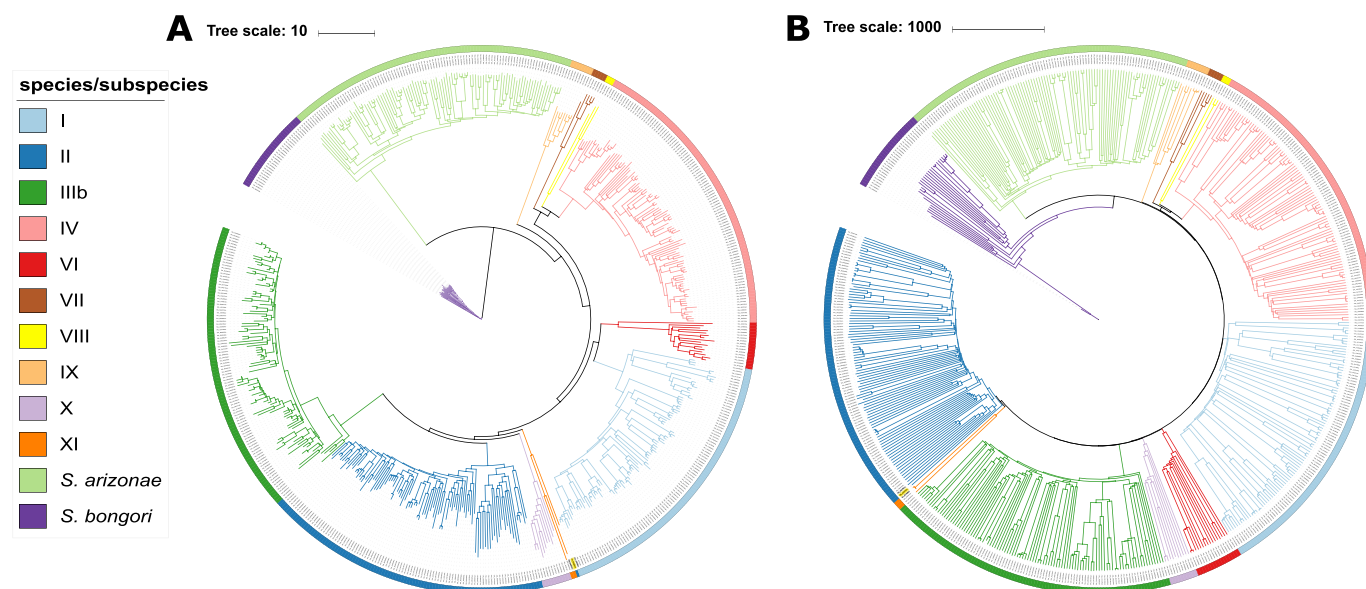


Fig. 1. A and B: Neighbour-Joining Tree based on the rMLST loci of 569 representative isolates (A) Neighbour-Joining Tree based on the cgMLST loci of 569 representative isolates (B).

Neighbour-Joining Trees of 569 isolates rooted using *S. bongori*. These trees were constructed using the BIGSdb Genome Comparator tool and SplitsTree software and edited using the ITOL online tool. Each isolate was randomly selected from the 1801 initial isolates to represent up to 100 different rST profiles per species or subspecies.

A: Neighbour-Joining Tree of 569 isolates, based on rMLST. The rMLST tree clustered all of the species and subspecies into distinct clades. The only isolate that didn't cluster as expected was SAL_BA7507AA, whose subspecies name is highlighted in yellow.

B: Neighbour-Joining Tree of 569 isolates, based on cgMLST. Within the cgMLST tree SAL_BA7507AA (highlighted in yellow) clustered with subspecies II but represented a sister clade to the other subspecies II isolates included in the analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

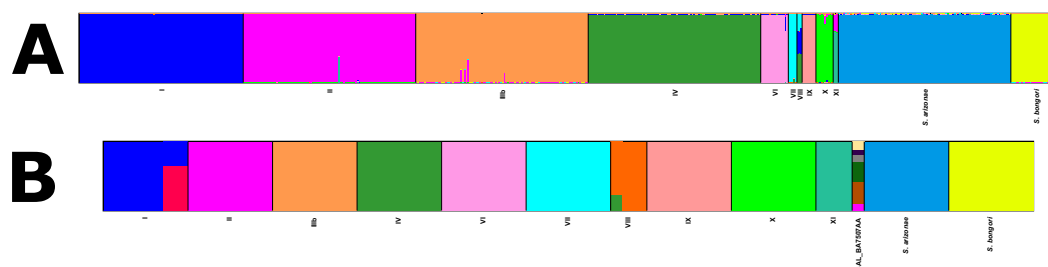


Fig. 2. a and b: STRUCTURE analysis (A) of 569 isolates using rMLST (B) of 57 isolates using cgMLST.

STRUCTURE analyses of representative isolates of the species and subspecies. The STRUCTURE algorithm was run until no further higher-level taxonomical groups were observed. Both of the STRUCTURE outputs were edited using the DISTRICT software.

A: The STRUCTURE algorithm was performed on the rMLST profiles of dataset A (Table S2) isolates with a K of 12. *S. bongori*, *S. arizonae* and subspecies I, II, IIIb, IV, VI, VII, IX, and X showed very little admixture and were all distinct populations. In comparison subspecies VIII and XI were composed of multiple rMLST populations. Subspecies VIII was an admixture of subspecies VII, IV and I, all of which were present in similar proportions in each isolate. Subspecies XI was predominantly a novel population, however there was a high level of admixture from subspecies II. The isolate SAL_BA7507AA grouped with subspecies II.

B: The STRUCTURE algorithm was performed on the cgMLST profiles of Dataset B (Table S2) with a K of 18. Species *S. bongori* and *S. arizonae* and subspecies II, IIIb, IV, VI, VII, IX, X, and XI were all distinct populations with little admixture. Subspecies I serovars Paratyphi A and Typhi also formed a distinct population (red), but showed admixture with the rest of subspecies I. One isolate of subspecies VIII was admixed with subspecies IV. Isolate SAL_BA7507AA was highly admixed with five novel populations and some of subspecies II.

amount of admixture with subspecies IV. Using cgMLST, isolate SAL_BA7507AA also formed a distinct population that was highly admixed and was composed of five novel populations and subspecies II. Higher values of K revealed no further subspecies; subspecies became more admixed and split into lineages.

Average Nucleotide Identity (ANI): The WGS ANI performed on dataset C showed that comparisons of all of the species and subspecies against *S. bongori* scored below 90% and comparisons against *S. arizonae* scored below 95% (Fig. 3a). All of these comparisons were below the 95% cut-off typically used within ANI analyses to define species groups [58,59] and supports the reclassification of *S. enterica* subspecies IIIa into a separate species, *S. arizonae*. The presence of novel subspecies identified through phylogenetic techniques (VII, VIII, IX, X, and XI) was also supported by this ANI analysis. Comparisons among the other *S. enterica* subspecies had ANI values between 92.93% and 97.13%, with an average of 95.15%. Comparisons within the subspecies were all higher, with ANI values above 98%, except for among subspecies VIII isolates, which scored 97.89% (Fig. 3a). While this was higher than all between-species comparisons, this low score may suggest that subspecies VIII is currently comprised of multiple subspecies.

The ANI analyses were repeated using the cgMLST (Fig. 3b) and rMLST loci (Fig. 3c) and the patterns observed within these analyses were highly similar to those observed within the analysis of the WGS. When ANI was performed on the cgMLST loci *S. bongori* and *S. arizonae* were all still lower than the 95% threshold, suggesting the classification of species [58,59]. The greater relatedness within as opposed to among subspecies generally persisted with the use of rMLST loci alone, with the exception of subspecies VIII, where the within-subspecies comparison was lower in subspecies VIII than the comparison between subspecies IV and VIII (Fig. 3c).

ANI analyses on WGS, cgMLST loci, and rMLST loci were repeated using randomly chosen isolates for all of the species and subspecies (dataset D). The results were highly congruent among the chosen type strains and the randomly chosen isolates (Fig. S3). This further supported the presence of the novel subspecies (VII, VIII, IX, X, and XI) and the reclassification of *S. enterica* subspecies IIIa to *S. arizonae*. The pattern of relatedness within ANI analyses using WGS, cgMLST, and rMLST sequence data were consistent throughout the comparisons, demonstrating that subspecies are observable across different genomic levels, despite a marked reduction in the number of loci used per analysis.

Isolate SAL_BA7507AA, which clustered with subspecies XI in rMLST trees and with subspecies II using cgMLST, was analysed using ANI against both datasets C and D. This isolate was closely related to

subspecies II and XI, scoring 97.99% and 97.87% respectively for dataset C and 97.65% and 97.69% respectively for dataset D (Fig. S4a). These values were all lower than comparisons within the subspecies II and subspecies XI isolates, which all scored above 98%. Whereas when isolate SAL_MA5841AA, which was phenotypically typed as subspecies II but clustered with subspecies I isolates phylogenetically, was analysed against datasets C and D it was very closely related to subspecies I, scoring 99.89% and 99.98% respectively (Fig. S4b).

3.4. Phenotypic analyses

Metabolic phenotyping was performed on pairs of isolates from each species and subspecies and the average signal values [60] used to construct a dendrogram based on metabolic respiration values (Fig. 4). This indicated that both *S. bongori* and *S. arizonae* (currently recognised subspecies IIIa) isolates clustered as species. This clustering was the case for the conventionally defined subspecies II, IIIb, IV, and VI and the phenotypically identified subspecies XI also formed an independent branch. The signal values clustered subspecies VII [9] and VIII [11] with subspecies IV. Subspecies I and X isolates did not cluster with each other or other subspecies (Fig. 4). The signal values identified subspecies I, serovar Enteritidis as phenotypically most closely related to the subspecies IV, VII, and VIII clade, while serovar Typhimurium was phenotypically most closely related to the *S. arizonae* clade. One of the subspecies X isolates was divergent from the subspecies II isolates while the other clustered with isolate I/II (SAL_MA5841AA). The biochemical approaches and the PHE k-mer approach identified this isolate as subspecies II, but it had clustered phylogenetically with subspecies I isolates. The subspecies IX isolates did not form a single clade (Fig. 4) but were both divergent from the same clade. Prior to identification and subsequent reclassification as novel subspecies, subspecies IX, X, and XI were all identified as subspecies II with phenotypic data accompanying the isolate; however, the dendrogram produced from the signal values demonstrated that none of these subspecies were in the same clade as subspecies II (Fig. 4). In addition, subspecies IX and XI were not closely related phenotypically to subspecies II.

Proposal of type strains and epithets for novel subspecies

The type strains for subspecies VII, VIII, IX, X, and XI (Table 1) recommended here were submitted to the National Collection of Type Cultures (NCTC) database and have therefore been given NCTC accession numbers as follows: Subspecies VII, SAL_MA2040AA, NCTC 1041; subspecies VIII, SAL_JA5200AA, NCTC 14236; subspecies IX, SAL_MA5929AA & SAL_MA5932AA, NCTC 14237 & NCTC 14238 respectively; subspecies X, SAL_MA5883AA & SAL_MA5962AA, NCTC 14239,

A) Whole Genome Sequences

	I	II	IIIb	IV	VI	VII	VIII	IX	X	XI	<i>S. arizonae</i>	<i>S. bongori</i>
I	98.37											
II	96.15	98.3										
IIIb	95.16	96.04	99.53									
IV	95.01	95.42	95	99.49								
VI	95.62	96	95.19	94.82	99.99							
VII	93.68	93.96	93.96	96.63	93.61	99.51						
VIII	95.82	95.44	94.71	97.13	94.77	96.45	97.89					
IX	94.57	95.2	94.78	94.15	94.61	92.93	94.08	99.79				
X	95.05	95.86	95.67	94.98	95.45	93.83	94.82	94.79	99.41			
XI	95.72	97.04	96.05	95.36	95.81	94.02	95.16	95.09	95.96	98.79		
<i>S. arizonae</i>	93.35	93.88	93.83	93.37	93.16	93.04	93.24	92.85	93.47	93.66	99.33	
<i>S. bongori</i>	89.79	89.91	89.78	89.54	89.58	89.38	89.62	89.5	89.54	89.72	89.31	99.39

B) cgMLST

	I	II	IIIb	IV	VI	VII	VIII	IX	X	XI	<i>S. arizonae</i>	<i>S. bongori</i>
I	98.79											
II	96.94	99.01										
IIIb	96.13	96.89	99.69									
IV	95.97	96.47	96.07	99.7								
VI	96.42	96.98	96.33	96.03	100							
VII	94.63	95.21	95.05	96.86	94.76	99.54						
VIII	96.62	96.22	95.86	97.26	95.88	96.54	98.79					
IX	95.64	96.25	95.86	95.32	95.78	94.3	95.17	99.88				
X	96.37	96.92	96.66	95.98	96.33	94.94	95.82	95.91	99.77			
XI	96.7	97.71	96.9	96.43	96.7	95.13	96.05	96.14	96.85	99.2		
<i>S. arizonae</i>	94.33	94.77	94.89	94.58	94.21	94.26	94.42	94.07	94.46	94.71	99.4	
<i>S. bongori</i>	91.1	90.89	91.11	90.8	90.7	90.53	90.73	90.94	90.97	90.91	90.33	99.54

C) rMLST

	I	II	IIIb	IV	VI	VII	VIII	IX	X	XI	<i>S. arizonae</i>	<i>S. bongori</i>
I	99.82											
II	99.34	99.88										
IIIb	99.13	99.47	99.92									
IV	98.83	99.09	99.05	99.9								
VI	99.13	99.37	99.25	98.89	100							
VII	98.53	98.72	98.65	99.15	98.56	99.84						
VIII	99.01	99.03	98.94	99.47	98.81	99.02	99.42					
IX	98.76	98.94	98.82	98.72	98.73	98.44	98.74	99.93				
X	99.25	99.43	99.29	98.93	99.17	98.54	98.94	98.82	99.89			
XI	99.06	99.33	99.16	98.81	99.13	98.48	98.77	98.72	99.16	99.79		
<i>S. arizonae</i>	98.56	98.61	98.42	98.42	98.39	98.3	98.42	98.26	98.46	98.41	99.82	
<i>S. bongori</i>	98.2	98.4	98.34	98.07	98.36	98	98.12	98.06	98.3	98.29	98	99.95

Fig. 3. A, B and C: ANI analysis of (A) type strains and recommended type strains using WGS (B) ANI analysis of type strains and recommended type strains using cgMLSTv1.0 (C) ANI analysis of type strains and recommended type strains using rMLST.

ANI analyses of isolates from dataset C (Table S2) where each isolate was compared to all of the other species or subspecies isolates and with a secondary isolate that belonged to the same species or subspecies using the OrthoANI algorithm.

A: ANI analysis of type strains and recommended type strains using WGS. Within this analysis all comparisons of species and subspecies against *S. bongori* and *S. arizonae* scored below 95%. Comparisons between the other subspecies ranged between 92.93 and 97.13%, with an average of 95.15%. Comparisons within the subspecies ranged from 97.89 to 99.99%.

B: ANI analysis of type strains and recommended type strains using cgMLST. *S. bongori* and *S. arizonae* scored below 95%. Comparisons between the other subspecies were between 94.3 and 97.71%, with an average of 96.10% and comparisons within subspecies were between 98.79 and 100%.

C: ANI analysis of type strains and recommended type strains using the rMLST loci. The patterns observed within the rMLST analysis was very similar to that observed within the cgMLST and WGS analyses. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

NCTC 14240 respectively; and subspecies XI, SAL_MA5907AA, NCTC 10436. Strains were also submitted to the German Collection of Micro-organisms and Cell Cultures (DSMZ) at time of publication and currently awaiting DSMZ accession numbers.

Type strains and epithets proposed in this study for recently defined

subspecies included *S. londonensis*, VII (NCTC10415), *S. brasiliensis*, VIII (NCTC14236), *S. hibernicus*, IX (NCTC14237), *S. essexiensis*, X (NCTC14239) and *S. reptilium* XI (NCTC10436) (Table 1). Naming of the subspecies was undertaken in line with the International Code of Nomenclature of Prokaryotes [62,63].

Phylogeny based on Metabolic Phenotyping

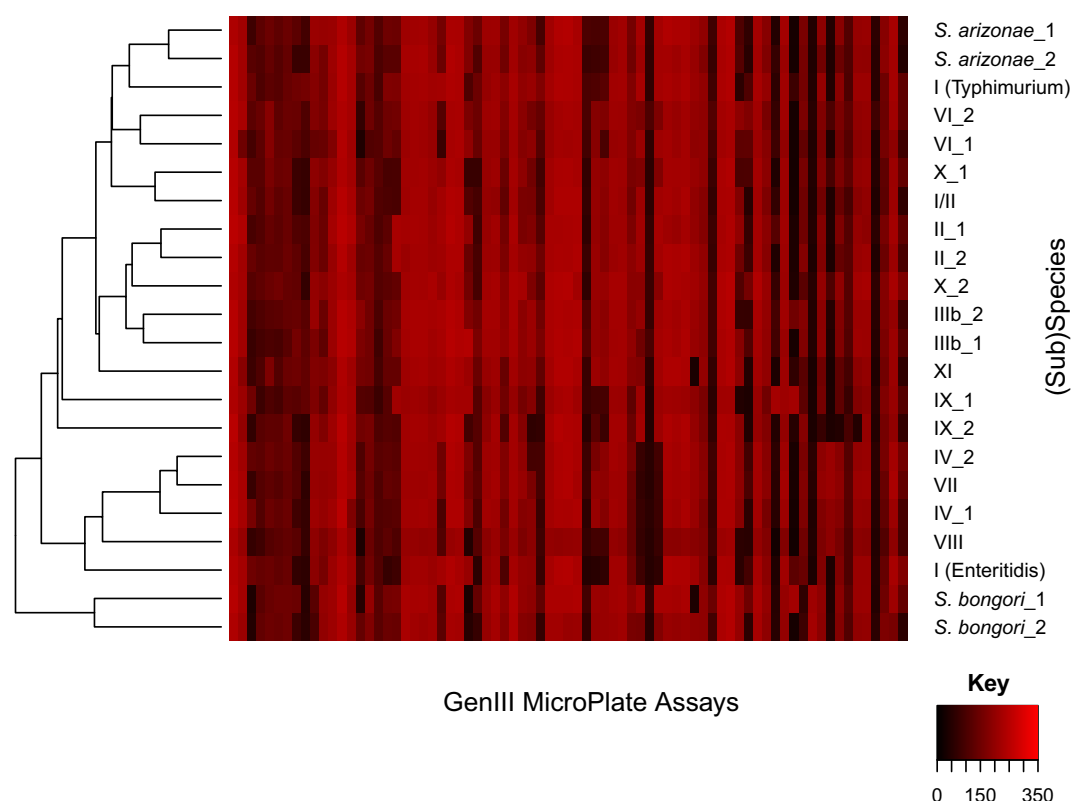


Fig. 4. Dendrogram based on metabolic phenotyping.

Pairs of isolates from each (sub)species underwent metabolic phenotyping using the biolog GenIII MicroPlate Assays. Runs were conducted three times and an average of the three runs was used to create signal values. These values were used to construct a dendrogram based on metabolic respiration values.

4. Discussion

The availability of very large numbers of bacterial WGSs provides alternative approaches to the definition and identification of bacterial species and subspecies. At the time of writing, isolate characterisation of *Salmonella* remained primarily undertaken by morphological and metabolic characterisation, despite the widespread uptake of WGS by reference laboratories, including those of Public Health England (PHE) isolates [15–20,41] and the U.S. Centers for Disease Control and Prevention (CDC) [21,64,65]. The phenotypic approaches were prone to result in misclassifications [40] with a lack of clarity in the identification of species and subspecies, as biochemical properties were often shared among distinct organisms [5,13]. Furthermore, the adoption by PulseNet International of cgMLST [22,65], a WGS-based method, for the global monitoring of foodborne disease outbreaks [15] and analysis [22] including for *Salmonella*, necessitated the harmonisation of approaches to species and subspecies definition and identification. WGS data have the additional advantage of being suitable for evolutionary analysis [12].

Using WGS-based approaches, four new *S. enterica* subspecies were confirmed and named with the next available roman numerals, in accordance with current taxonomy [4]: subspecies VII [9]; VIII; IX; and X. An additional novel subspecies was also identified, subspecies XI. Furthermore, *S. enterica* subspecies IIIa was confirmed as a separate species, *Salmonella arizonae*. The name *S. arizonae* was chosen as *arizonae* was the historical name for subspecies IIIa before the subspecies were reclassified with roman numerals [4]. Until 1973, isolates from this group were considered to form a separate species [10]. The identification, characterisation, and reclassification of these *Salmonella* species

and subspecies was consistent across different genomic levels, using multiple phylogenetic and statistical approaches [54,57] which were highly congruent. As different subspecies and serovars of *Salmonella* have different antimicrobial resistance profiles [66,67], virulence properties, and primary hosts [66,67], the identification and accurate analyses of these different subspecies, such as with the cgMLST genus scheme (SalmcgMLSTv1.0) [40], is important.

WGS-based methods consistently identified the subspecies correctly and within the phylogenies the species and subspecies formed distinct clades irrespective of the phylogenetic approach used. Trees reconstructed using the Neighbour-Joining algorithm, based on the 51 rMLST loci [44] and the 2750 cgMLST [40] loci (Fig. 1), and a rMLST phylogeny reconstructed using the Maximum Likelihood algorithm [46] (Fig. S2) were highly congruent. The consistencies between the MLtree and NJtree demonstrated that the increase in time taken to produce a higher resolution tree provided limited further insight into the relationships among the salmonellae and was therefore unnecessary for the purpose of species and subspecies classification. The phylogenetic clades were also highly congruent with the results from ANI [57] (Fig. 3) and STRUCTURE [54] analyses (Fig. 2), where all species and subspecies were identified. All these analyses supported the classification of *S. bongori*, *S. arizonae* and *S. enterica* as separate species and of the division of *S. enterica* into subspecies I – XI. In addition, these groupings were consistent with those defined using alternative evolutionary approaches, which in addition showed evidence for hybridisation among the species and subspecies during their emergence [12].

In all the reconstructed phylogenies, *S. bongori* was the most divergent of the *Salmonella* species. Furthermore, *S. arizonae* (currently recognised as subspecies IIIa) was divergent from the remaining *S. enterica*

subspecies in all phylogenies, consistent with it forming another species and being a *S. enterica* subspecies. Results from the ANI [57] analyses also demonstrated that both *S. bongori* and *S. arizonae* were separate species, as defined by the 95% ANI threshold for species definition. Comparisons with *Salmonella* isolates from the other species and subspecies consistently scored below 95% ANI [1], *S. bongori* consistently scored below 90% when compared against all other *Salmonella* species or subspecies and *S. arizonae* scored below 94%.

The genetic differences among the subspecies were also conserved, as were the relationships among the species and subspecies, observable within phylogenies when trees were constructed using cgMLST [40] or rMLST [44], within ANI analyses and using the STRUCTURE algorithm. The ANI values were consistently lower among subspecies comparisons than within subspecies comparisons. When using STRUCTURE, the rMLST analysis identified all twelve of the species and subspecies at a population of twelve, while cgMLST required a population of eighteen. As anticipated, in the cgMLST analysis, serovars Typhi and Paratyphi A were split into their own population, which was admixed with the subspecies I population. It is likely this distinct population occurred due to the host-restricted nature of these serovars [68,69], the reduction in their genome size [70], and hybridisation between the two serovars [69,70]. Furthermore, one isolate of subspecies VIII also demonstrated a degree of admixture with subspecies IV, this is likely because these subspecies are closely related [11].

Of the large number of isolates examined here, only two isolates were identified that demonstrated incongruences. When analysed using phylogenetic approaches isolate SAL_BA7507AA clustered with subspecies XI using rMLST [44] and subspecies II using cgMLST [40]. Isolate SAL_BA7507AA was also problematic when the cgMLST loci were analysed using the STRUCTURE algorithm [54] (Fig. 2b), showing evidence of being highly admixed. ANI analysis of this isolate (Fig. S4a) against the type strains (Table 1) and randomly chosen isolates (Datasets C and D) found that it was on the threshold for clustering with both subspecies II (97.65% and 97.99% respectively) and XI (97.69% and 97.87% respectively). The ambiguity to which subspecies isolate SAL_BA7507AA belongs in all these analyses suggests that it could represent a novel subspecies; however, this isolate could also be an admixed sample or an isolate representing a midpoint between subspecies II and XI. Additional isolates closely related to SAL_BA7507AA would need to be identified to confirm the classification of this isolate.

All of the WGS-based approaches used in the present study identified isolate SAL_MA5841AA as a subspecies I organism, despite it being identified as a subspecies II using conventional approaches. Reanalysis using Biolog (Technopath) indicated that this isolate clustered more closely with isolates identified as subspecies II. This discrepancy suggested that occasional conflicts between biochemical and phylogenetic approaches will occur, but that these appear to be rare. This further emphasised the need for a WGS-based approach to species definition and identification, demonstrating the misclassifications [40] that can occur with phenotypic biochemical properties, which are frequently shared [5,13]. Furthermore, dendrograms produced from the metabolic respiration values demonstrated that phenotypic typing is not always capable of delineating the *Salmonella* subspecies. The large-scale analyses carried out on a selection of 22 *Salmonella* isolates using WGS-based approaches are not scalable to phenotypic approaches. These 22 isolates were representative of all of the *Salmonella* species and subspecies identified through both conventional and phylogenetic approaches. The metabolic respiration [60] of these isolates was examined using Biolog (Technopath) (Fig. 4). While some of the *Salmonella* species and subspecies were identified through these metabolic respiration profiles, the dendrogram produced demonstrated that phenotypic typing was not consistent in distinguishing *Salmonella* subspecies. This shortcoming was true for both conventionally-identified subspecies, such as subspecies I and for phylogenetically identified subspecies such as IX and X [11]. The representative isolates of these subspecies did not form subspecies-specific clades within the dendrogram of the metabolic respiration

values. The isolates from phylogenetically identified subspecies VII [9] and VIII [11] clustered with the subspecies IV isolates, which was expected, as subspecies VII and VIII isolates were classified as subspecies IV prior to the present analyses. While there were phenotypic differences between most of the subspecies, shared biochemical properties [5,13] resulted in the inability for biochemical subspecies identification methods to represent the genetic relationships among isolates accurately [5,13] and consistently distinguish all of the subspecies, demonstrating the need for a shift in how subspecies are identified [1]. Furthermore, these methods require microbiological expertise across a wide range of biochemical tests and are low throughput [14].

A number of studies have been published which identify the need for a genome-based species identification methods [71–73] and recommended guidelines have been published [72,74,75] in support of this shift away from conventional methods. The use of genome-based methods for the identification of species is becoming more common and species within other genera, such as *Neisseria* [76] and *Acinetobacter* [77] have been identified using WGS-based approaches. In some cases, with open sharing of genomic data, similar studies have overlapped [11,12]. Even though genomic approaches were different, the same novel groups have been detected but named in different ways, for example novel subspecies VIII (novel subsp. A, Houtenae B), IX (novel subsp. B, VIII) and X (novel subsp. C, Salamae B)) [11,12]. Hence the need for the IJSEM, which ultimately agrees new taxonomy proposals and has now introduced the use of genomic data for publication of new taxa into their scope (www.microbiologyresearch.org/journal/ijsem/scope). Having bespoke genomic methods can be insightful in understanding evolutionary events and detection of novel strains but what standardised methods can be used in reference laboratories to detect and report out these strains in real-time? Routinely, *Salmonella* isolates can be typed using sequence typing methods, such as cgMLST [78] or rMLST [44] and isolates that do not cluster with any known subspecies can be confirmed as novel subspecies using ANI [57], the new ‘gold standard approach’ [59] for species definition and identification. Of course, updated k-mer databases can be used and should be reviewed to add any newly defined species or subspecies. As the cgMLST phylogenies provided limited additional resolution compared to the rMLST phylogenies, and given the facts that rMLST requires less computational power and curation of fewer loci [44,79], rMLST could be used for routine species and subspecies identification. The ribosomal genes are distributed around the genome, offering some stability against horizontal gene transfer [44,79] and are also under stabilising selection for conservation of function [44]. These traits make the ribosomal genes ideal candidates for a typing scheme.

Isolates from the different *S. enterica* subspecies began to cluster together when more than 19 of the rST alleles differed. Therefore, a more conservative number of allelic differences, such as 15, would identify the majority of isolates from the same subspecies, whilst ensuring that conflicts among subspecies classification did not occur. This threshold should, however, be periodically reviewed because as gaps within the diversity of an organism are filled, the number of allelic differences required to identify all members of a species or subspecies can decrease.

For the confirmation of novel subspecies, ANI can be used by comparing test WGS data against reference genomes for each subspecies. To identify organisms as the same species an ANI of 95% or greater is generally recommended, although this varies among species [58], from the data presented here, 95% appears appropriate for the salmonellae. ANI has been shown to be comparable [59] and congruent [80] to DNA–DNA hybridisation (DDH) [2]. Due to the development of online tools [57,81], however, ANI analyses are simpler to perform than DDH, which is too labour-intensive [72], time consuming [72] and error prone [82] for all but a few labs to perform [72,82]. As ANI indexes differs from across the whole genome, it is not reliant on a single or small subset of genes remaining consistent across all isolates within a species or subspecies. Therefore, ANI should not be affected by small changes within a

genome and should be capable of consistently identifying species and subspecies.

Five new *S. enterica* subspecies have been confirmed and named with the next available roman numerals, in accordance with current *Salmonella* taxonomy (VII, VIII, IX, X, and XI) [3,4] with the subspecific epithet in line with the International Code of Nomenclature of Prokaryotes [62,63] (Table 1). Representative isolates have been deposited into the NCTC and DSMZ culture collections (Table 1) as type strains (*S. enterica* subsp. *londinensis*, VII: NCTC 1041, *S. enterica* subsp. *brasiliensis*, VIII: NCTC 14236, *S. enterica* subsp. *hibernicus* IX: NCTC 14238 and NCTC 14237, *S. enterica* subsp. *essexiensis*, X: NCTC 14239 and NCTC 14240, *S. enterica* subsp. *reptilium*, XI: NCTC 10436). *S. enterica* subspecies IIIa was identified as a separate species – *Salmonella arizonae*, as historically suggested [83] which was identified in this study using rMLST and cgMLST NJtrees and confirmed using an MLTree, STRUCTURE analysis and ANI data. It is demonstrated here that the identification of species and subspecies of the genus *Salmonella* can and should be performed using WGS-based approaches. This methodology change would be in line with global surveillance [22] and reference laboratories [15], which are now routinely using WGS-based approaches for *Salmonella* identification and monitoring.

5. Conclusions

A comprehensive analysis of the genus of *Salmonella* with respect to species and subspecies structure, undertaken with WGS data analysed with a range of different approaches, proposed that: (i) five novel subspecies (*S. enterica* subsp. *londinensis* (VII), *S. enterica* subsp. *brasiliensis* (VIII), *S. enterica* subsp. *hibernicus* (IX), *S. enterica* subsp. *essexiensis* (X), *S. enterica* subsp. *reptilium* (XI)) be defined; and (ii) *S. enterica* subspecies IIIa should be reclassified as the separate species, *S. arizonae*. The analyses demonstrated that conventional microbiological approaches to species and subspecies definition and identification were not uniformly consistent with WGS data and could not identify all of the species and subspecies within the genus reliably. It was further shown that routine species and subspecies identification could be achieved with rMLST, which it has a low computational overhead. For those few isolates which were not identified unambiguously with rMLST data alone, ANI using core genome loci confirmed species or subspecies status is required.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2021.07.003>.

Declaration of competing interest

None.

Acknowledgements

Acknowledgements to Dr. Patti Fields and Susan Van Duyne (CDC) for their strain and metadata contributions.

Acknowledgements to the Institut Pasteur for previously providing strains to CDC and François-Xavier Weill for permission to use a strain in this study.

Thank you to Sarah Alexander from National Collection of Type Cultures (NCTC) for provision of strains and historical metadata.

Thank you to Mark Pallen from the Quadram Institute for advice on Latin nomenclature of bacterial species.

This research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with University of East Anglia, University of Oxford and the Quadram Institute (no. 109524). Madison Pearce was based at the University of Oxford. The views expressed are those of the author (s) and not necessarily those of the NHS, the NIHR, the Department of Health, Public Health England, Centers for Disease Control and Prevention or the Quadram Institute.

GCL was funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent project BBS/E/F/000PR10352.

References

- [1] K.T. Konstantinidis, A. Ramette, J.M. Tiedje, The bacterial species definition in the genomic era, *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 361 (1475) (2006) 1929–1940.
- [2] L.G. Wayne, D.J. Brenner, R.R. Colwell, P.A.D. Grimont, O. Kandler, M. I. Krichevsky, et al., Report of the ad hoc committee on reconciliation of approaches to bacterial systematics, *Int. J. Syst. Evol. Microbiol.* 37 (4) (1987) 463–464.
- [3] Patrick A.D. Grimont, François-Xavier Weill, *Antigenic Formulae of the Salmonella Serovars*, 9th ed., Institut Pasteur: WHO Collaborating Centre for Reference and Research on Salmonella, 2007, 166 p.
- [4] F.W. Brenner, R.G. Villar, F.J. Angulo, R. Tauxe, B. Swaminathan, *Salmonella* nomenclature, *J. Clin. Microbiol.* 38 (7) (2000) 2465–2467.
- [5] J.R. McQuiston, S. Herrera-Leon, B.C. Wertheim, J. Doyle, P.I. Fields, R.V. Tauxe, et al., Molecular phylogeny of the salmonellae: relationships among *Salmonella* species and subspecies determined from four housekeeping genes and evidence of lateral gene transfer events, *J. Bacteriol.* 190 (21) (2008) 7060–7067.
- [6] Judicial Commission Of The International Committee On Systematics Of P, The type species of the genus *Salmonella* Lignieres 1900 is *Salmonella enterica* (ex Kauffmann and Edwards 1952) Le Minor and Popoff 1987, with the type strain LT2T, and conservation of the epithet enterica in *Salmonella enterica* over all earlier epithets that may be applied to this species. Opinion 80, *Int J Syst Evol Microbiol* 55 (Pt 1) (2005) 519–520.
- [7] B.J. Tindall, P.A.D. Grimont, G.M. Garrity, J.P. Euzéby, Nomenclature and taxonomy of the genus *Salmonella*, *Int. J. Syst. Evol. Microbiol.* 55 (Pt 1) (2005) 521–524.
- [8] M.W. Reeves, G.M. Evins, A.A. Heiba, B.D. Plikaytis, J.J. Farmer 3rd., Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov., *J. Clin. Microbiol.* 27 (2) (1989) 313–320.
- [9] E.F. Boyd, F.S. Wang, T.S. Whittam, R.K. Selander, Molecular genetic relationships of the salmonellae, *Appl. Environ. Microbiol.* 62 (3) (1996) 804–808.
- [10] J.H. Crosa, D.J. Brenner, W.H. Ewing, S. Falkow, Molecular relationships among the Salmonellae, *J. Bacteriol.* 115 (1) (1973) 307–315.
- [11] N.F. Alikhan, Z. Zhou, M.J. Sergeant, M. Achtman, A genomic overview of the population structure of *Salmonella*, *PLoS Genet.* 14 (4) (2018), e1007261.
- [12] A. Criscuolo, S. Issenuth-Jeanjean, X. Didelot, K. Thorell, J. Hale, J. Parkhill, et al., The speciation and hybridization history of the genus *Salmonella*, *Microbial genomics.* 5 (8) (2019), e000284.
- [13] E.F. Boyd, D.L. Hartl, *Salmonella* virulence plasmid. Modular acquisition of the spv virulence region by an F-plasmid in *Salmonella enterica* subspecies I and insertion into the chromosome of subspecies II, IIIa, IV and VII isolates, *Genetics* 149 (3) (1998) 1183–1190.
- [14] S. Herrera-León, J.R. McQuiston, M.A. Usera, P.I. Fields, J. Garaizar, M.A. Echeita, Multiplex PCR for distinguishing the most common phase-1 flagellar antigens of *Salmonella* spp., *J. Clin. Microbiol.* 42 (6) (2004) 2581–2586.
- [15] M.A. Chattaway, T.J. Dallman, L. Larkin, S. Nair, J. McCormick, A. Mikhail, et al., The transformation of reference microbiology methods and surveillance for *Salmonella* with the use of whole genome sequencing in England and Wales, *Front. Public Health* 7 (2019) 317.
- [16] M.A. Chattaway, G.C. Langridge, J. Wain, *Salmonella* nomenclature in the genomic era: a time for change, *Sci. Rep.* 11 (1) (2021) 7494.
- [17] M. Chattaway, D. Greig, A. Gentle, H. Hartman, T. Dallman, C. Jenkins, Whole-genome sequencing for national surveillance of *Shigella flexneri*, *Front. Microbiol.* 8 (2017).
- [18] M.A. Chattaway, T.J. Dallman, A. Gentle, M.J. Wright, S.E. Long, P.M. Ashton, et al., Whole genome sequencing for public health surveillance of Shiga toxin-producing *Escherichia coli* other than serogroup O157, *Front. Microbiol.* 7 (2016) 258.
- [19] T.J. Dallman, M.A. Chattaway, P. Mook, G. Godbole, P.D. Crook, C. Jenkins, Use of whole-genome sequencing for the public health surveillance of *Shigella sonnei* in England and Wales, 2015, *J. Med. Microbiol.* 65 (8) (2016) 882–884.
- [20] A. Holmes, T.J. Dallman, S. Shabaan, M. Hanson, L. Allison, Validation of whole-genome sequencing for identification and characterization of shiga toxin-producing *Escherichia coli* to produce standardized data to enable data sharing, *J Clin Microbiol* 56 (3) (2018).
- [21] E. Vaughn, Q. Vo, J. Vostok, T. Stiles, A. Lang, C. Brown, et al., Linking epidemiology and whole-genome sequencing to investigate *Salmonella* outbreak, Massachusetts, USA, 2018, *Emerging Infectious Disease journal.* 26 (7) (2020) 1538.
- [22] C. Nadon, I. Van Walle, P. Gerner-Smidt, J. Campos, I. Chinen, J. Concepcion-Acevedo, et al., PulseNet International: vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance, *Eurosurveillance.* 22 (23) (2017) 30544.
- [23] P. Publications, *Salmonella* Data 2007–2016, Public Health England, London, 2018.
- [24] Editorial team C, S. Bertrand, R. Rimhanen-Finne, F.X. Weill, W. Rabsch, L. Thornton, et al., *Salmonella* infections associated with reptiles: the current situation in Europe, *Eurosurveillance* 13 (24) (2008) 18902.

- [25] A. Lamas, J.M. Miranda, P. Regal, B. Vázquez, C.M. Franco, A. Cepeda, A comprehensive review of non-enterica subspecies of *Salmonella enterica*, *Microbiol. Res.* 206 (2018) 60–73.
- [26] S. Nair, J. Wain, S. Connell, E. de Pinna, T. Peters, *Salmonella enterica* subspecies II infections in England and Wales – the use of multilocus sequence typing to assist serovar identification, *J. Med. Microbiol.* 63 (Pt 6) (2014) 831–834.
- [27] J.B. Hoag, C.N. Sessler, A comprehensive review of disseminated *Salmonella* Arizona infection with an illustrative case presentation, *South. Med. J.* 98 (11) (2005) 1123–1129.
- [28] V.O. Wikström, L.L. Fernström, L. Melin, S. Boqvist, *Salmonella* isolated from individual reptiles and environmental samples from terraria in private households in Sweden, *Acta Vet. Scand.* 56 (1) (2014) 7.
- [29] T. Zottola, S. Montagnaro, C. Magnapera, S. Sasso, L. De Martino, A. Bragagnolo, et al., Prevalence and antimicrobial susceptibility of salmonella in European wild boar (*Sus scrofa*); Latium region – Italy, *Comp. Immunol. Microbiol. Infect. Dis.* 36 (2) (2013) 161–168.
- [30] N. Gay, S. Le Hello, F.X. Weill, B. de Thoisy, F. Berger, *Salmonella* serotypes in reptiles and humans, French Guiana, *Vet Microbiol.* 170 (1–2) (2014) 167–171.
- [31] S.B. Romero, P. Kvapil, A. Cizek, Z. Knotek, The prevalence and antimicrobial resistance of *Salmonella* species isolated from captive reptiles at Ljubljana Zoo, *Slov. Vet. Res.* 53 (2016) 43–48.
- [32] L. Duffy, G. Dykes, N. Fegan, A review of the ecology, colonization and genetic characterization of *Salmonella enterica* serovar Sofia, a prolific but avirulent poultry serovar in Australia, *Food Res. Int.* 45 (2012) 770–779.
- [33] A. Rodriguez, P. Pangoli, H.A. Richards, J.R. Mount, F.A. Draughon, Prevalence of *Salmonella* in diverse environmental farm samples, *J. Food Prot.* 69 (11) (2006) 2576–2580.
- [34] K. Sören, M. Lindblad, C. Jernberg, E. Eriksson, L. Melin, H. Wahlström, et al., Changes in the risk management of *Salmonella enterica* subspecies diarizonae serovar 61: (k):1, 5, (7) in Swedish sheep herds and sheep meat due to the results of a prevalence study 2012, *Acta Vet Scand* 57 (1) (2015) 6.
- [35] M. Madsen, Prevalence and serovar distribution of *Salmonella* in fresh and frozen meat from captive Nile crocodiles (*Crocodylus niloticus*), *Int. J. Food Microbiol.* 29 (1) (1996) 111–118.
- [36] A. Nowakiewicz, G. Ziolkowska, P. Zięba, K. Stepniewska, S. Tokarzewski, Russian tortoises (*Agryonemys horsfieldi*) as a potential reservoir for *Salmonella* spp, *Res. Vet. Sci.* 92 (2) (2012) 187–190.
- [37] S. Magnino, P. Colin, E. Dei-Cas, M. Madsen, J. McLauchlin, K. Nöckler, et al., Biological risks associated with consumption of reptile products, *Int. J. Food Microbiol.* 134 (3) (2009) 163–175.
- [38] M. Schröter, P. Roggentin, J. Hofmann, A. Speicher, R. Laufs, D. Mack, Pet snakes as a reservoir for *Salmonella enterica* subsp. diarizonae (Serogroup IIIB): a prospective study, *Appl. Environ. Microbiol.* 70 (1) (2004) 613–615.
- [39] Association AVM, AMVA Pet Ownership and Demographics Sourcebook, USA, 2018.
- [40] M.E. Pearce, M.A. Chattaway, K. Grant, M.C.J. Maiden, A proposed core genome scheme for analyses of the *Salmonella* genus, *Genomics* 112 (1) (2020) 371–378.
- [41] P.M. Ashton, S. Nair, T.M. Peters, J.A. Bale, D.G. Powell, A. Painset, et al., Identification of *Salmonella* for public health surveillance using whole genome sequencing, *PeerJ.* 4 (2016), e1752.
- [42] K.A. Jolley, M.C.J. Maiden, BIGSdb: scalable analysis of bacterial genome variation at the population level, *BMC Bioinformatics.* 11 (1) (2010) 595.
- [43] M. Achtman, J. Wain, F.X. Weill, S. Nair, Z. Zhou, V. Sangal, et al., Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*, *PLoS Pathog.* 8 (6) (2012), e1002776.
- [44] K.A. Jolley, C.M. Bliss, J.S. Bennett, H.B. Bratcher, C. Brehony, F.M. Colles, et al., Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain, *Microbiology (Reading)* 158 (Pt 4) (2012) 1005–1015.
- [45] Z. Zhou, N.F. Alikhan, M.J. Sergeant, N. Luhmann, C. Vaz, A.P. Francisco, et al., GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens, *Genome Res.* 28 (9) (2018) 1395–1404.
- [46] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* 4 (4) (1987) 406–425.
- [47] N. Goldman, Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analyses, *Syst. Biol.* 39 (4) (1990) 345.
- [48] D. Falush, M. Torpdahl, X. Didelot, D.F. Conrad, D.J. Wilson, M. Achtman, Mismatch induced speciation in *Salmonella*: model and data, *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 361 (1475) (2006) 2045–2053.
- [49] X. Didelot, R. Bowden, T. Street, T. Golubchik, C. Spencer, G. McVean, et al., Recombination and population structure in *Salmonella enterica*, *PLoS genetics* 7 (7) (2011) e1002191-e.
- [50] I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, *Nucleic Acids Res.* 44 (W1) (2016) W242–W245.
- [51] K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, *Nucleic Acids Res.* 30 (14) (2002) 3059–3066.
- [52] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, *Mol. Biol. Evol.* 33 (7) (2016) 1870–1874.
- [53] M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, *J. Mol. Evol.* 16 (2) (1980) 111–120.
- [54] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, *Genetics* 155 (2) (2000) 945–959.
- [55] M.J. Hubisz, D. Falush, M. Stephens, J.K. Pritchard, Inferring weak population structure with the assistance of sample group information, *Mol. Ecol. Resour.* 9 (5) (2009) 1322–1332.
- [56] N.A. Rosenberg, Distruct: a program for the graphical display of population structure, *Mol. Ecol. Notes* 4 (1) (2004) 137–138.
- [57] I. Lee, Y. Ouk Kim, S.C. Park, J. Chun, OrthoANI: an improved algorithm and software for calculating average nucleotide identity, *Int. J. Syst. Evol. Microbiol.* 66 (2) (2016) 1100–1103.
- [58] J. Goris, K.T. Konstantinidis, J.A. Klappenbach, T. Coenye, P. Vandamme, J. M. Tiedje, DNA-DNA hybridization values and their relationship to whole-genome sequence similarities, *Int. J. Syst. Evol. Microbiol.* 57 (Pt 1) (2007) 81–91.
- [59] M. Richter, R. Rosselló-Móra, Shifting the genomic gold standard for the prokaryotic species definition, *Proc. Natl. Acad. Sci. U. S. A.* 106 (45) (2009) 19126–19131.
- [60] O.R. Homann, H. Cai, J.M. Becker, S.L. Lindquist, Harnessing natural diversity to probe metabolic pathways, *PLoS Genet.* 1 (6) (2005), e80.
- [61] P.T. Desai, S. Porwollik, F. Long, P. Cheng, A. Wollam, V. Bhonagiri-Palsikar, et al., Evolutionary genomics of *Salmonella enterica* subspecies, *mBio* 4 (2) (2013).
- [62] T.C. Parker, B.J. Tindall, G.M. Garrity, International code of nomenclature of prokaryotes, *Int. J. Syst. Evol. Microbiol.* 69 (1a) (2019) s1–s111.
- [63] J.G.T.B.F. Graesse, H. Plechl, *Orbis Latinus: Lexikon lateinischer geographischer Namen des Mittelalters und der Neuzeit*, 1972.
- [64] G.L. Armstrong, D.R. MacCannell, J. Taylor, H.A. Carleton, E.B. Neuhaus, R. S. Bradbury, et al., Pathogen genomics in public health, *N. Engl. J. Med.* 381 (26) (2019) 2569–2580.
- [65] E.M. Ribot, M. Freeman, K.B. Hise, P. Gerner-Smidt, PulseNet: entering the age of next-generation sequencing, *Foodborne Pathog. Dis.* 16 (7) (2019) 451–456.
- [66] H. Harbottle, D.G. White, P.F. McDermott, R.D. Walker, S. Zhao, Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates, *J. Clin. Microbiol.* 44 (7) (2006) 2449–2457.
- [67] S. Sukhnanand, S. Alcaine, L.D. Warnick, W.L. Su, J. Hof, M.P. Craver, et al., DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella enterica* serotypes, *J. Clin. Microbiol.* 43 (8) (2005) 3688–3698.
- [68] S. Spanò, Mechanisms of *Salmonella Typhi* host restriction, *Adv. Exp. Med. Biol.* 915 (2016) 283–294.
- [69] K.E. Holt, N.R. Thomson, J. Wain, G.C. Langridge, R. Hasan, Z.A. Bhutta, et al., Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi, *BMC Genomics* 10 (1) (2009) 36.
- [70] X. Didelot, M. Achtman, J. Parkhill, N.R. Thomson, D. Falush, A bimodal pattern of relatedness between the *Salmonella Paratyphi* A and Typhi genomes: convergence or divergence by homologous recombination? *Genome Res.* 17 (1) (2007) 61–68.
- [71] J. Chun, F.A. Rainey, Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea, *Int. J. Syst. Evol. Microbiol.* 64 (Pt 2) (2014) 316–324.
- [72] D. Gevers, F.M. Cohan, J.G. Lawrence, B.G. Spratt, T. Coenye, E.J. Feil, et al., Opinion: re-evaluating prokaryotic species, *Nat Rev Microbiol.* 3 (9) (2005) 733–739.
- [73] N.J. Varghese, S. Mukherjee, N. Ivanova, K.T. Konstantinidis, K. Mavrommatis, N. C. Kyrpides, et al., Microbial species delineation using whole genome sequences, *Nucleic Acids Res.* 43 (14) (2015) 6761–6767.
- [74] J. Chun, A. Oren, A. Ventosa, H. Christensen, D.R. Arahal, M.S. da Costa, et al., Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes, *Int. J. Syst. Evol. Microbiol.* 68 (1) (2018) 461–466.
- [75] M. Kim, H.S. Oh, S.C. Park, J. Chun, Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes, *Int. J. Syst. Evol. Microbiol.* 64 (Pt 2) (2014) 346–351.
- [76] J.S. Bennett, K.A. Jolley, S.G. Earle, C. Corton, S.D. Bentley, J. Parkhill, et al., A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*, *Microbiology (Reading)* 158 (Pt 6) (2012) 1570–1580.
- [77] J.Z.M. Chan, M.R. Halachev, N.J. Loman, C. Constantinidou, M.J. Pallen, Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*, *BMC Microbiol.* 12 (1) (2012) 302.
- [78] M.E. Pearce, N.F. Alikhan, T.J. Dallman, Z. Zhou, K. Grant, M.C.J. Maiden, Comparative analysis of core genome MLST and SNP typing within a European *Salmonella enterica* serovar Enteritidis outbreak, *Int. J. Food Microbiol.* 274 (2018) 1–11.
- [79] M.C. Maiden, M.J. Jansen van Rensburg, J.E. Bray, S.G. Earle, S.A. Ford, K. A. Jolley, et al., MLST revisited: the gene-by-gene approach to bacterial genomics, *Nat Rev Microbiol.* 11 (10) (2013) 728–736.
- [80] K.T. Konstantinidis, J.M. Tiedje, Genomic insights that advance the species definition for prokaryotes, *Proc. Natl. Acad. Sci. U. S. A.* 102 (7) (2005) 2567–2572.
- [81] M. Richter, R. Rosselló-Móra, F. Oliver Glöckner, J. Peplies, JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison, *Bioinformatics* 32 (6) (2016) 929–931.
- [82] E. Stackebrandt, W. Frederiksen, G.M. Garrity, P.A.D. Grimont, P. Kämpfer, M.C. J. Maiden, et al., Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology, *Int. J. Syst. Evol. Microbiol.* 52 (Pt 3) (2002) 1043–1047.
- [83] Skerman V.B.D. MV, P.H.A. Sneath, Approved lists of bacterial names (amended), in: Skerman VBD, V. McGowan, Sneath PHA (Eds.), *Approved Lists of Bacterial Names (Amended)*, ASM Press, Washington, DC, 1989. Copyright © 1989, ASM Press.