Originally published as:

Toboldt, A., Tietze, E., Helmuth, R., Junker, E., Fruth, A., Malorny, B.
Population structure of Salmonella enterica serovar 4,[5],12:b-strains and likely sources of human infection
(2013) Applied and Environmental Microbiology, 79 (17), pp. 5121-5129.

DOI: 10.1128/AEM.01735-13

Population structure of *Salmonella enterica* serovar 4,[5],12:b:- strains and likely sources of human infection

Anne Toboldt[1,2], Erhard Tietze[3], Reiner Helmuth[1], Ernst Junker[1], Angelika Fruth[3], Burkhard Malorny[1]*

[1]*Federal Institute for Risk Assessment, National Reference Laboratory for Salmonella, Max-Dohrn-Str 8-10, D-10589 Berlin, Germany*

[2]*Free University Berlin, Department of Biology, Chemistry and Pharmacy, Takustr. 3, 14195 Berlin, Germany*

[3]*Robert-Koch Institute, Wernigerode Branch, Division Enteropathogenic Bacteria and Legionella, National Reference Centre for Salmonellae and other Bacterial Enteric Pathogens, Burgstr. 37, 38855 Wernigerode, Germany*

*Corresponding author:

Dr. Burkhard Malorny

Federal Institute for Risk Assessment,

National Reference Laboratory for Salmonella

Diedersdorfer Weg 1

D-12277 Berlin, Germany

Tel: +49 30 18412 2237 Fax: +49 30 18412 2953

E-mail: burkhard.malorny@bfr.bund.de

Running title: Characterization of *Salmonella enterica* serovar 4,[5],12:b:-

ABSTRACT

Salmonella enterica serovar 4,[5],12:b:- is a monophasic serovar not able to express the second phase flagellar antigen (H2-antigen). In Germany, the serovar is occasionally isolated from poultry, reptiles, fish, food and humans. In this study a selection of 67 epidemiologically unrelated *Salmonella enterica* serovar 4,[5],12:b:- strains isolated in Germany between 2000 and 2011 from the environment, animal, food, and humans was investigated by phenotypic and genotypic methods to better understand the population structure and to identify potential sources of human infections. Strains of this monophasic serovar were highly diverse. Within the 67 strains analyzed we identified 52 different pulsed-field gel electrophoresis XbaI profiles, twelve different multilocus sequence types and 18 different pathogenicity array types. The relatedness of strains based on the pathogenicity gene repertoire (102 markers tested) was in good agreement with grouping by MLST. *S. enterica* serovar 4,[5],12:b:- is distributed across multiple unrelated eBurst groups and consequently highly polyphyletic. Two sequence types (ST88 and ST127) were linked to *S. enterica* serovar Paratyphi B (D-tartrate +), two single locus variants of ST1583 were linked to *S. enterica* serovar Abony and one sequence type (ST1484) was associated with *S. enterica* serovar Mygdal, a recently defined new serovar. From the characterization of clinical isolates and those of non-human origin it can be concluded that the potential sources of sporadic human infections with *S. enterica* serovar 4,[5],12:b:- most likely are mushrooms, shellfish/fish or poultry.

INTRODUCTION

Salmonella enterica subsp. *enterica* is one of the leading causes of zoonotic food-borne disease worldwide. The main reservoir of *Salmonella enterica* is the intestinal tract of various animal species. The pathogen is transmitted to humans mainly by contaminated food causing gastroenteritis and occasionally systemic infections (1). Globally, approximately 93.8 million human cases with 155,000 deaths annually have been estimated (2). The species is subdivided based on serological classification according to the White-Kaufmann-Le Minor scheme into almost 2600 serovars (3). They are defined by an antigenic formula based on the presence of one somatic (O-antigen) and two flagellar antigens (H1- and H2-antigens). The monophasic *S. enterica* serovar 4,[5],12:b:- does not express phase-2 flagellar antigen. It was reported as fluctuating serovar in broiler flocks (4) and recognized in Spanish and Danish poultry slaughterhouses (5, 6). Occasionally the serovar is isolated from human cases of gastroenteritis (7). Such cases were associated with exposure to turtles and considered to be possibly a specific variant of the biphasic D-tartrate fermenting (dT+) *S. enterica* serovar Paratyphi B (also called var. Java) with seroformula 4,[5],12:b:1,2 (8). Initial characterisation using multilocus sequence typing showed that *S. enterica* serovar 4,[5],12:b:- strains grouped in various eBurst groups (eBGs), some of them together with the biphasic *S. enterica* serovars Paratyphi B, Dublin and Enteritidis (7, 9). DNA microarray analysis of seven 4,5,12:b:- strains related to Danish human cases supports clustering in two separate branches, one together with *S. enterica* serovar Dublin and Enteritidis and another with *S. enterica* serovar Paratyphi B dT+ (7).

The aim of our study was to identify potential sources of human infections caused by D-tartrate positive *S. enterica* serovar 4,[5],12:b:- and to gain a better understanding of the population structure and genetic relatedness within this serovar. In addition we compared it to other serovars, especially to those with O-antigens 4,[5],12. For this purpose we selected from

our strain collections 67 *S. enterica* serovar 4,[5],12:b:- strains isolated from poultry, reptiles, shellfish/fish, different food and humans in Germany during the years 2000 to 2011 and investigated them by phenotypic and genotypic methods. Furthermore, the pathogenicity gene repertoire was determined and compared to estimate the potential health risk for humans.

MATERIAL AND METHODS

Strain selection. Monophasic *S. enterica* serovar 4,[5],12:b:- strains used were selected from the strain collections of the National Reference Laboratory for Salmonella at the Federal Institute for Risk Assessment, Berlin, Germany (NRL-Salm) and the National Reference Centre for *Salmonella* and other Enterics, Wernigerode, Germany (NRZ-RKI). All isolates were received from public and private diagnostic laboratories for serotyping (Table 1). To distinguish the *S. enterica* serovar 4,[5],12:b:- strains from *S. enterica* serovar Schleißheim (seroformula 4,12,27:b:-) or *S. enterica* subspecies *salamae* serovar 4,[5],12,[27]:b:[e,n,x] the following biochemical tests were performed with all monophasic strains (in brackets expected results for *S. enterica* serovar 4,[5],12:b:-/Schleißheim/4,[5],12,[27]:b:[e,n,x]): fermentation of dulcitol (+/-/+), malonate (-/-/+) and gelatinase (-/+/+). Altogether 39 *S. enterica* serovar 4,[5],12:b:- strains isolated from humans suffering from gastroenteritis and 28 strains isolated from livestock, shellfish/fish, mushrooms, food, reptiles and the environment were chosen for molecular typing (Table 2). The origins and sources of the strains cover various geographically distinct regions in Germany. All isolates were obtained between the years 2000 and 2011 and there was no obvious epidemiological link between them, i.e. not isolated at the same place or time or from the same individual.

A subset of 29 *S. enterica* serovar 4,[5],12:b:- strains was chosen to study their pathogenicity gene repertoire using DNA microarrays. It was selected in order to reflect the diversity of PFGE XbaI-profiles and MLST analysis observed among all of the 67 epidemiologically unrelated strains.

Serotyping. Serotyping was performed according to the White-Kauffmann-Le Minor scheme (3) by slide agglutination with O- and H-antigen specific sera (Sifin Diagnostics, Berlin, Germany). The H:z91 antiserum was purchased from Medco Diagnostika GmbH (München, Germany).

Antimicrobial susceptibility testing. Susceptibility of strains was tested against 14 antimicrobials by determining the minimum inhibitory concentration (MIC) using the CLSI broth micro dilution method (10) in combination with the semi-automatic Sensititre system (TREK Diagnostic Systems, Cleveland, Ohio). Cut-off values (mg/l) to be used to determine susceptibility to 10 antimicrobials were applied as described in the Commission Decision on a harmonised monitoring of antimicrobial resistance in poultry and pigs, ((EG) 2007/407), namely cefotaxime (FOT, >0.5), nalidixic acid (NAL, >16), ciprofloxacin (CIP, >0.06), ampicillin (AMP, >4), tetracycline (TET, >8), chloramphenicol (CHL, >16), gentamicin (GEN, >2), streptomycin (STR, >32), trimethoprim (TMP, >2), sulfamethoxazole (SMX, >256). Cut-off values for the remaining 4 antimicrobials were adopted from the European Committee on Antimicrobial Susceptibility Testing, 2011 (http://www.eucast.org/clinical_breakpoints/), namely colistin (COL, >2), florfenicol (FFN, >16), kanamycin (KAN, >32), and ceftazidime (TAZ, >2).

Genomic DNA purification. DNA for PCRs and DNA microarray experiments was isolated from strains grown in Luria-Bertanti broth (Merck, Darmstadt, Germany) at 37°C for 16-18 h. A 1.6 ml aliquot was carried out for purification using the RTP Bacteria DNA Mini Kit (STRATEC Molecular GmbH, Berlin, Germany) according to the manufacturer´s protocol with one additional step. After cell lysis step at 95°C for 5 to10 min and cooling samples for 5 min, a 5-µl aliquot of RNase (100 mg/ml) (Qiagen GmbH, Hilden, Germany) was added and incubated at room temperature for 30 min. The quality and quantity of DNA preparations was determined spectrophotometrically. For DNA-labelling with fluorophors a minimum of 4 µg DNA and, for PCRs a 1-ng/µl dilution in TE buffer were used.

Multilocus sequence typing (MLST). MLST was carried out as previously described including partial sequences of the seven housekeeping genes *aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA* (11). Alleles and sequence types were assigned according to the MLST scheme available at http://mlst.ucc.ie/mlst/dbs/Senterica. Unknown alleles were submitted to

129 the website and newly named. The analysis was carried out in BioNumerics v6.6.4 (Applied
130 Maths, Sint-Martens-Latem, Belgium). The comparisons were made by advanced cluster
131 analysis using the analysis template Maximum Spanning Tree (MST) for categorical data on
132 merged sequences of the seven genes. Complexes were designated as eBurst groups (eBGs)
133 and a new eBG (eBG242) was assigned according to the definition by Achtman et al. (9).

134 **Pulsed-field gel electrophoresis (PFGE).** PFGE was performed according to the Pulse-
135 Net protocol (12) using the restriction enzyme XbaI for digestion of genomic DNA. The
136 analyses of the gel images were carried out in BioNumerics v6.6.4. The comparisons were
137 made by cluster analysis using Dice coefficient and UPGMA with a position tolerance of
138 1.5% and optimization of 1.0%. Fragments that were smaller than 25 kb were not considered
139 for cluster analysis.

140 **DNA microarray analysis.** The DNA microarray used in this study was applied as
141 previously described (13). Altogether 80 pathogenicity gene markers, 22 fimbrial gene
142 markers and 49 resistance gene markers were analysed for the 29 *S. enterica* serovar
143 4,[5],12:b:- strains representing the diversity of PFGE profiles. Moreover, markers for the
144 flagellar genes were investigated (*fljA*, *fljB*, *hin*, *fliC*). Analysis of raw data was performed as
145 previously described (13). After normalisation (presence/absence of gene) the data for each
146 strain were imported in BioNumerics v6.6.4 as character value. For comparison, a cluster
147 analysis with the simple matching binary coefficient, by using the UPGMA dendrogram type
148 was applied on the basis of the 80 pathogenicity and 22 fimbrial markers.

149 **PCRs.** Testing for the presence of the *oafA* gene responsible for O:5-antigen expression in
150 *Salmonella* was according to Hauser et al. (2011) (14) using oligonucleotides P-439 and P-
151 440 amplifying a 433 bp PCR product. To detect a 7 bp tandem repeat within the open
152 reading frame oligonucleotides P-439 and P-1072 were used resulting in a PCR product of
153 170 bp. In case of loss of one 7 bp repeat no PCR product was obtained indicating a non-
154 functional *oafA* gene. The ability of *S. enterica* serovar 4,[5],12:b:- strains to ferment D-

tartrate was determined according to the PCR protocol described by Malorny et al. (15). Moreover, all serotypically monophasic *S. enterica* serovar 4,[5],12:b:- strains were tested by PCR with specific oligonucleotides for the presence of the *fljB*-1,2 (H:1,2 antigen) according to Lim et al. (16).

DNA sequencing. From two strains (11-02483 and 11-02464) belonging to ST127, one strain (11-02467) belonging to ST88 and one strain (11-02485) belonging to ST42 the *fljA*, *fljB*, *hin* and *iroB* genes were sequenced. The region was amplified by three overlapping PCRs using the following oligonucleotides: (i) P-120 (CAGCGTAGTCCGAAGACGTGA) and P-915 (ACGAATGGTACGGCTTCTGTAACC) resulting in a 1881 bp fragment for all four strains; (ii), P-1386 (GTCAGTAGCAACGTTAACTT) and P-1387 (ATGAGGTAAACGTACCGACA) resulting in a 2129 bp fragment for strain 11-02467 and 11-02485 and resulting in a 1126 bp fragment for strains 11-02483 and 11-02464. (iii) P-1392 (CGGAAAGTCTGCACAGAAC) and P-1551 (GCGAACTATCCAGGCACGA) resulting in a 1647 bp fragment for all four strains. All PCR products were sequenced by Qiagen GmbH sequencing service (Hilden, Germany). Single DNA sequence runs were assembled and analysed using Lasergene software package (version 8.1; DNASTAR, Madison, WI). Oligonucleotides used for sequencing can be obtained on request. Sequence comparisons were performed using BLAST search at NCBI (http://blast.ncbi.nlm.nih.gov/).

The *fliC* gene was amplified from two ST1484 strains, 10-01745 and 11-02460, using oligonucleotides P-314 (AAGGAAAAGATCATGGCA) and P-60 (CCGTGTTGCC CAGGTTGGTAAT). The 1286 bp PCR products were sequenced with oligonucleotides P-314 and P-60 by Qiagen GmbH.

Statistical methods. The Simpson's index of diversity (ID) and the 95% confidence intervals (CI) were calculated using the Comparing Partitions website (http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Tool).

The Chi-square test with a confidence interval of 95% ($p<0.05$) was applied (http://www.daten-consult.de/frames/statrechnen.html) to check the significance in the number of strains attributed to each isolation source category.

Nucleotide sequence accession numbers. The *fljB* surrounding DNA sequences were deposited in GenBank under accession numbers: HG003856 (strain 11-02483), HG003858 (11-02485) and HG003857 (11-02467). Partial *fliC* DNA sequences of strains 10-01745 and 11-02460 were deposited under accession numbers HG003859 and HG0003860.

187 **RESULTS**

188

189 **Prevalence of *S. enterica* serovar 4,[5],12:b:-.** Between 2000 and 2011 the National
190 Reference Laboratory for Salmonella (NRL-Salm) received for diagnostic serotyping 47,238
191 *Salmonella* strains, isolated by public and private diagnostic laboratories across Germany
192 from livestock, reptiles, shellfish/fish, food, feed, and the environment. Of these, 0.08% (40
193 strains) were assigned to *S. enterica* serovar 4,[5],12:b:- (Table 1). Strains were mostly
194 isolated from poultry (12 strains), food (8 strains), and reptile (5 strains) and were
195 sporadically isolated from shellfish/fish (3 strains), sheep (2 strains), pig (1 strain) and some
196 other sources in Germany (Table1). Twenty one strains (52.5%) exhibited the O:5 antigen in
197 addition to the O:4,12 antigen. O:5 antigen negative strains were mainly isolated from
198 poultry. Likewise, between 2000 and 2011 the National Reference Centre for *Salmonella* and
199 other Enterics (NRZ-RKI), received 50,705 *Salmonella* strains isolated from humans with
200 *Salmonella* infection in Germany. Of these, 96 strains (0.2%) were serotyped as *S. enterica*
201 serovar 4,[5],12:b:- (Table 1). Most of the human *S. enterica* serovar 4,[5],12:b:- strains (93
202 strains, 95%) expressed the O:5 antigen. Data indicate a misbalance with respect to the
203 frequency of strains isolated from humans expressing the O:5 antigen and such strains
204 isolated from non-human origin, especially poultry.

205 Molecular analysis of the *oafA* gene (encoding O:5 antigen factor) in O:5-antigen negative
206 strains showed that such strains either harbored a non-functional *oafA* gene due to a seven
207 basepair deletion within the ORF or showed that the *oafA* gene was completely absent,
208 especially in strains belonging to sequence type ST1484 (Table 2).

209 From the 136 *S. enterica* serovar 4,[5],12:b:- strains identified in both collections
210 altogether 39 strains isolated from humans and 28 strains isolated from non-human origin
211 were selected for extended molecular typing (Table 2). All the three strains isolated from
212 humans and not expressing the O:5 antigen were included. The remaining 36 strains (92%)

213 isolated from humans and expressing the O:5 antigen were randomly selected, at least two
214 strains from each year (2000-2011). All non-human strains were chosen without obvious
215 epidemiological link.

216 **Antimicrobial resistance.** Sixty-two of the 67 *S. enterica* 4,[5],12:b:- strains (93%) were
217 susceptible to all 14 antimicrobials tested. Three strains were monodrug-resistant to
218 sulfamethoxazole (SMX) and two strains were multidrug-resistant to four or more
219 antimicrobials (Table 2). The strain 07-01889-2 (quail isolate) was resistant to NAL SMX
220 STR TET and the strain 11-02483 (human isolate) was resistant to AMP CHL KAN SMX
221 STR TET. By DNA microarray analysis we have found the following resistance genes:
222 $bla_{\text{TEM-1like}}$ encoding AMP resistance, *aadA1 and aadA2,3,8* encoding STR/SPE resistance,
223 *cmlA1*-like encoding CHL resistance, *sul3* encoding SUL resistance, and *tet*(A) encoding TET
224 resistance. Further markers indicated that specific antibiotic resistance genes are organized
225 within class 1 and class 2 integrons in strain 11-02483.

226 **Typing by PFGE.** Fifty-two different XbaI profiles could be distinguished among the 67
227 strains analyzed (ID, 0.987 [95% CI, 0.976 to 0.998]) (Fig. 1). They were classified into seven
228 clusters (A, B, C, D, E, F and G). The similarity coefficient (*F* value) ranged for cluster A
229 from 0.75 to 0.97, and values were similar for cluster C (0.80 to 0.97), E (0.75 to 0.96) and G
230 (0.89 to 0.96). Cluster A contained 21 of 67 strains (31%), cluster C 26 strains (39%), cluster
231 E eleven strains (16%) and cluster G six strains (9%). Only one strain (1.5%) belonged to
232 clusters B, D and F. Related clusters F and G included all O:5 antigen negative strains with
233 complete absence of the *oafA* gene. These strains were isolated from poultry (four strains),
234 bird (one strain), feed (one strain) and human (one strain). Clusters A to E contained only O:5
235 antigen positive strains with five exceptions, all encoding a non-functional *oafA* gene due to a
236 seven basepair deletion within the ORF (Table 2). Strains isolated from humans were
237 distributed over all seven clusters. In respect to non-human strains we observed a preference
238 for strains isolated from mushrooms in cluster A and for strains isolated from poultry in

239 cluster G. However, the observation was not significant.

240 **MLST.** Twelve different sequence types (STs) were identified (ID, 0.830 [95%CI, 0.787 –
241 0.874]). The most prominent STs were ST42 (27% of strains), ST423 (25%), ST127 (13%),
242 ST135 (12%), and ST1484 (10%). Two strains (3%) belonged to ST1583 and one strain each
243 (1.5%) belonged to the ST88, ST679, ST1578, ST1582, ST1588, and ST1589, respectively.
244 STs were categorized in three main complexes comprising more than one ST (Fig. 2). The
245 founder of the largest complex (eBG32) was ST42. ST42 differed from ST423 in one
246 nucleotide in the *thrA* allele, from ST1582 in one nucleotide in the *hemD* allele and from
247 ST1588 in one nucleotide in the *sucA* allele. The founder of the second complex is ST135
248 with single locus variants ST1589 and ST1578 (eBG 242). Both sequence types differed from
249 ST135 only in one nucleotide each in another allele (*hisD* in ST1589 and *dnaA* in ST1578).
250 ST42 differed from ST135 in four alleles, *aroC* (five nucleotides), *dnaA* (five nucleotides),
251 *sucA* (one nucleotide) and *thrA* (nine nucleotides). The third complex (eBG19) consists of
252 ST127 and ST88. ST127 differed from ST88 in one nucleotide in the *dnaA* allele. ST135 and
253 ST127 differed in six alleles and share only an identical *aroC* allele. There are three unique
254 STs among the strains under investigation of which two (ST1583 and ST1484) were not yet
255 assigned to any eBG and one ST (ST679) that was assigned to eBG155. They have none or at
256 maximum two common alleles (Fig. 2).

257 The eBurst groups or single STs were associated to some extend with strains from a
258 common source. Four STs (ST423, ST42, ST135, and ST127) were associated with more than
259 six strains isolated from humans. ST127 was with one exception (00-02320, sheep)
260 exclusively associated with human strains. Strains from poultry or poultry meat were mainly
261 assigned to ST1484, also sharing the XbaI PFGE clusters F and G. Three out of the five
262 strains isolated from mushrooms belonged to ST135. All four *S. enterica* serovar 4,5,12:b:-
263 strains isolated from shellfish/fish belonged to eBG32.

264 We have compared the newly identified *S. enterica* serovar 4,[5],12:b:- ST1484, ST1578,

265 ST1582, ST1583, ST1588, and ST1589 with all STs publicly available in the Salmonella-
266 MLST database (http://mlst.ucc.ie/mlst/dbs/Senterica) in order to identify related STs in any
267 other *S. enterica* serovars. Two single locus variants (ST273 and ST442) of ST1583 were
268 associated with *S. enterica* serovar Abony. Furthermore, ST1484 was a single locus variant of
269 ST252 observed in a *S. enterica* serovar Mygdal strain (4,12:z_{91}:-).

270 **Determination of pathogenicity genes.** Eighteen different pathogenicity array types
271 (PATs) were identified among the 29 strains tested (Table 2). PATs differed in up to 24 of
272 102 markers tested (Fig. 3). We have observed certain relatedness between PATs and specific
273 eBGs. ST42 strains, the founder of eBG32, are linked to six different PATs, and ST423
274 strains to three different PATs. All those PATs are closely related. PATs associated with
275 eBG242 cluster together with PATs of eBG32. For ST127 and ST88 strains (eBG19) three
276 similar but distinct PATs were found. The remaining PATs (13, 14 and 18) are subdivided in
277 specific branches in accordance to their specific STs (ST679, ST1583 and ST1484).

278 A number of markers targeting the *Salmonella* Pathogenicity Islands (SPIs) SPI-1, SPI-2
279 and SPI-3 were absent in certain strains or eBGs. All strains belonging to eBG32 (PAT 1-6, 8-
280 10) lacked the *avrA* gene located in SPI-1 (encoding a protein inhibiting the key
281 proinflammatory immune response) and *rhuM* and *sugR* genes, both located in SPI-3. In
282 PATs 11, 12 and 15 to 18 the *avrA* gene was present, but the *rhuM* and *sugR* genes (SPI-3)
283 were missing. The six PATs (PAT 1-3, 6, 8 and 9) linked with ST42 differed from each other
284 in one to five of the seven markers for genes *gtgA*, *sodC1*, *sseI*, *irsA*, *sopE1*, *srfJ* or *tcfA*.
285 Three PATs (PAT2, 4 and 5) belonged to ST423. These PATs differed in single markers for
286 *gtgA*, *sopE1* or *srfJ*. ST127 (PAT 15 and PAT 16) and ST88 (PAT17) strains belonging to
287 eBG19 differed in four pathogenicity gene markers (*sseI*, *sspH1*, *sopE1* and *sopD2*). Markers
288 for *msgA* (SsrB regulator) and *pagK* (PhoPQ activated protein) were only absent in PAT18
289 (ST1484).

290 **Serotype marker genes in *S. enterica* 4,[5],12:b:-.** Genes encoding the repressor for

291 phase-1 flagellin (*fljA*), the structural phase-2 flagellin unit (*fljB*), and a DNA Invertase (*hin*)
292 are ordered consecutively in the *Salmonella* genome. Three different combinations for these
293 markers were found within the 29 *S. enterica* 4,[5],12:b:- strains tested. DNA microarray
294 probes indicating the presence of *fljA*, *fljB_1,x* and *hin* genes were negative in 22 strains. In
295 five strains (all belonging to ST127) the probes for genes *fljA* and the *fljB_1,x* were present
296 but *hin* was absent. Two strains (11-02467, ST88 and 11-02485, ST42) were positive for all
297 three probes *fljA*, *fljB_1,x* and *hin*.

298 We have sequenced the *fljB* region of two ST127 strains (11-02464 and 11-02483), one
299 ST88 strain (11-02467) and one ST42 strain (11-02485) to identify the genetic background
300 leading to non-functional phase-2 H-antigen expression. Both ST127 strains lacked the *hin*
301 gene exactly between the recombination sites *hixL* and *hixR* (17) but genes *fljA* and *fljB* were
302 present. In the ST88 strain genes *fljA* and *fljB* as well as the *hin* gene were present in a regular
303 arrangement. However, compared to the DNA sequence of *S. enterica* serovar Typhimurium
304 LT2 (GeneBank accession no. NC_003197) there was a one basepair deletion at position
305 2915827 within a non-coding region downstream to *hin*. Sequence comparison revealed 100%
306 identity with the flagellin encoding gene *fljB* of *S. enterica* serovars Newport, Hissar,
307 Litchfield, Stanley and Schottmuelleri whereas the identity to *fljB*-1,2 encoded in *S. enterica*
308 serovars Typhimurium, Paratyphi B dT+ or Saintpaul strains was 99%. All these serovars
309 express the phase-2 H:1,2-antigen. In the monophasic ST42 strain all three genes *fljA*, *fljB* and
310 *hin* were present. However, there were several polymorphic sites compared to the ST88 strain.
311 The ST42 *fljB* gene differed from the ST88 *fljB* gene in six nucleotides and it was 100%
312 identical to *fljB* gene of *S. enterica* serovar Paratyphi B dT+ str. SPB7 (GeneBank accession
313 no. CP000886.1). The DNA sequence revealed no hints which could explain a non-functional
314 phase-2 H-antigen expression of the strain.

315 The phase-1 flagellin gene *fliC* is chromosomally located apart from *fljB*. The *fliC*-b
316 (encoding H:b structural phase-1 flagellin unit) marker was positive in all 29 *S. enterica*

4,[5],12:b:- strains tested by microarray analysis. We have also sequenced the *fliC* gene of two ST1484 strains and found only one (10-01745) or two (11-02460) non-synonymous polymorphic sites compared to the unique published *fliC*-z91 gene sequence of *S. enterica* serovar Mygdal (GeneBank accession no. GQ280905.1). Agglutination using H:z91 antiserum was positive for all strains belonging to ST1484 but negative for representative *S. enterica* 4,[5],12:b:- strains belonging to all other MLST types. This indicates that a single amino acid exchange in the H:b structural phase-1 flagellin protein leads to a positive agglutination reaction with H:z91 antiserum besides with H:b antiserum (G in ST1484 strains → D in *S. enterica* serovar Mygdal). Usually, *fliC*-z91 of *S. enterica* serovar Mygdal differs from *fliC*-b in 15 (GeneBank accession no. DQ838210.1) to 48 nucleotides (CP000886.1).

DISCUSSION

In this study the population structure and pathogenicity gene repertoire of *S. enterica* serovar 4,[5],12:b:-, as well as potential infection sources for humans, were investigated. The monophasic serovar is not frequently isolated from humans and animals but occasional cases were reported (8). MLST data of 67 strains analyzed showed that *S. enterica* serovar 4,[5],12:b:- is highly polyphyletic, which is in contrast to another monophasic *S. enterica* serovar, the worldwide expanding *S. enterica* serovar Typhimurium variant 4,[5],12:i:- (18, 19). Altogether 12 different STs were found clustering in four eBurst groups and two additional unrelated STs. The genetic diversity was supported by PFGE data assigning 52 XbaI profiles with a discrimination index of 0.987. This clearly indicates that the serovar is composed of various phylogenetic lineages lacking a common ancestor.

Relatedness of *S. enterica* serovar 4,[5],12:b:- sequence types to other serovars. Some of the STs observed in *S. enterica* serovar 4,[5],12:b:- were identical to or single locus variants of STs associated with other *S. enterica* serovars characterized by the same somatic and H1-antigen but expressing in addition the H2:1,2-antigen (serovar Paratyphi B dT+) or the H2:e,n,x,-antigen (serovar Abony). To our surprise, a single locus variant (ST252) of our ST1484 *S. enterica* serovar 4,[5],12:b:- strains, which are strongly associated with poultry, was found in a *S. enterica* serovar Mygdal strain (4,12:z_{91}:-). *S. enterica* serovar Mygdal was newly identified and recently added to the White-Kauffmann-Le Minor scheme (20). The original reference strain was isolated from swine faeces in 2003 in Denmark. DNA sequencing of the *fliC* gene in two ST1484 strains supports the close relationship to ST252. The *fliC*-z91 gene of *S. enterica* serovar Mygdal and *fliC*-b gene of *S. enterica* serovar 4,[5],12:b:- ST1484 strains differed only in up to two non-synonymous nucleotides leading to a positive slide agglutination with H:z91 and H:b antisera. From our data we conclude that certain *S. enterica* serovar 4,[5],12:b:- strains are monophasic variants of closely related

353 biphasic serovars with which they share the O- and H1-antigen. Others are closely related to
354 monophasic serovars (e.g. Mygdal) sharing highly similar H1-antigens. However, for most
355 STs belonging to *S. enterica* serovar 4,[5],12:b:- related serovars have not been identified.
356 With the increasing generation of MLST data possibly certain further genetic relationships to
357 other serovars will be found.

358 Molecular screening for presence of the *fljB* gene in *S. enterica* serovar 4,[5],12:b:- was
359 with two exceptions in accordance with the monophasic phenotype. The exceptional strains
360 (11-02467, 11-02485) were assigned to ST88 and ST42. ST88 occurred also in biphasic *S.*
361 *enterica* serovar Paratyphi B dT+ (21). Both types of ST88 strains possess *fljA*, *fljB* and *hin*
362 genes. In the monophasic ST88 strain a single basepair deletion in the intergenic region
363 between *hin* and *iroB* might be the cause for a non-functional phase-2 flagellin expression.
364 However, we speculate that the *hin* gene is by any reasons locked and unable to switch in
365 "on" orientation for co-transcription of *fljA* and *fljB*. The *fljB* sequence of the ST88 strain was
366 100% identical to the *fljB* gene of *S. enterica* serovars Newport, Hissar, Litchfield, Stanley
367 and Schottmuelleri probably imported by horizontal gene transfer. Whether the same *fljB*
368 allele occurs in biphasic ST88 *S. enterica* serovar Paratyphi B dT+ strains has to be
369 elucidated. ST127 strains lacked the *hin* gene but *fljA* and *fljB* genes were identical compared
370 to the monophasic ST88 strain. This supports the close relationship of ST127 and ST88
371 strains as also indicated by MLST.

372 **Potential sources of human infection.** The sources of *S. enterica* serovar 4,[5],12:b:-
373 strains of non-human origin isolated in Germany between 2000 and 2011 (40 strains) were
374 diverse. Most strains were isolated from poultry (30%), mushrooms (12.5%) or shellfish/fish
375 (10%) but due to the genetic variability and low prevalence the serovar could not be
376 associated to a specific source of human infection. However, MLST analysis revealed one
377 specific type which is strongly connected to poultry. Strains isolated from poultry belonged
378 predominantly to ST1484 and one out of the 39 human strains investigated clustered into this

379 ST. This indicates that there is human exposure to poultry contaminated with this *S. enterica*
380 serovar 4,[5],12:b:- type but clinically apparent infections are rare. We have previously
381 described two similar examples in *S. enterica* serovar Paratyphi B dT+ strains belonging to
382 ST28 and in *S. enterica* serovar 4,12:d:- strains (ST279). These strains were also highly
383 associated with poultry but rarely isolated from humans (13, 21). However, because of
384 completely different alleles *S. enterica* serovar 4,12:b:- ST1484 strains do not represent a
385 monophasic variant of *S. enterica* serovar Paratyphi B dT+ ST28 strains as might be
386 concluded from seroformula or are related to *S. enterica* serovar 4,12:d:- strains.. Of higher
387 risk for humans seems to be the exposure of mushrooms contaminated with *S. enterica*
388 serovar 4,[5],12:b:-. Three out of five strains isolated from mushrooms and five strains from
389 humans belonged exclusively to ST135 and, therefore, likely connected to the consumption of
390 mushrooms. The source of two mushroom strains (ST135) was linked to an import from Asia.
391 Most of the *S. enterica* serovar 4,[5],12:b:-strains (55%) investigated belonged to eBG32.
392 The group was highly diverse in respect to PFGE-profiles and PATs. Predominantly, all four
393 strains isolated from shellfish/fish were found in this group along with human strains
394 indicating a seafood associated subtype which is able to infect humans. However, single
395 strains isolated from a number of various other sources were also found in this lineage. The
396 two eBGs32 and eBGs242 showed variation in the pathogenicity gene repertoire but could be
397 clearly differentiated from eBG19 and other singleton STs. Some of these differences might
398 have an influence on the virulence in humans and animals.
399 Human salmonellosis caused by *S. enterica* serovar 4,[5],12:b:- has been reported from
400 Spain (8). Contact with turtles has been identified as the potential source of infection. We
401 have observed in our study an identical PFGE XbaI profile in two strains isolated from human
402 and feed (XbaI profile no. 38), respectively, as previously described by Hernandez et al.
403 (2012) (8). If the feed (dried fish) was intended for reptiles is unknown. However, in our
404 study, two strains isolated from reptiles were distantly related between each other and to the

405 one described previously. Therefore, based on our data we cannot conclude that reptiles
406 contaminated with *S. enterica* serovar 4,[5],12:b:- are a major source of human infection.

407 In conclusion, *S. enterica* serovar 4,[5],12:b:- is a polyphyletic serovar which can be
408 isolated from many different sources. The serovar is represented by several phylogenetic
409 lineages and is not generally a monophasic variant of *S. enterica* serovar Paratyphi B dT+.
410 Consumption of contaminated mushrooms and shellfish/fish are potential sources of human
411 infection. The pathogenicity gene repertoire of the different phylogenetic lineages indicates
412 that some lineages might be more virulent for humans than others. Currently, the serovar
413 seems not to pose a major threat for humans.

ACKNOWLEDGEMENTS

This work was funded by the Bundesministerium für Bildung und Forschung (BMBF), FBI-Zoo (01KI1012I (BFR) and 01KI 1012F (RKI)). We thank Istvan Szabo, Andreas Schroeter, Manuela Jaber, Martha Brom and Katharina Thomas, for serotyping and antimicrobial resistance testing of strains.

REFERENCES

1. **Thorns, C. J.** 2000. Bacterial food-borne zoonoses. Rev Sci Tech **19:**226-239.
2. **Majowicz, S. E., J. Musto, E. Scallan, F. J. Angulo, M. Kirk, S. J. O'Brien, T. F. Jones, A. Fazil, R. M. Hoekstra, and S. International Collaboration on Enteric Disease 'Burden of Illness.** 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis **50:**882-889.
3. **Grimont, P. A. D. F.-X. W.** 2007. Antigenic formulae of the *Salmonella* serovars. , 9th ed. WHO Collaborating Centre for Reference and Research on *Salmonella*, Institut Pasteur, Paris, France. Available from:http://www.pasteur.fr/ip/portal/action/WebdriveActionEvent/oid/01s-000036-089. Last access April 16, 2013.
4. **Chadfield, M., M. Skov, J. Christensen, M. Madsen, and M. Bisgaard.** 2001. An epidemiological study of *Salmonella enterica* serovar 4, 12:b:- in broiler chickens in Denmark. Vet Microbiol **82:**233-247.
5. **Carramiñana, J. J., J. Yangüela, D. Blanco, C. Rota, A. I. Agustín, and A. Herrera.** 1997. Potential virulence determinants of *Salmonella* serovars from poultry and human sources in Spain. Vet Microbiol **54:**375-383.
6. **Olsen, J. E., D. J. Brown, M. Madsen, and M. Bisgaard.** 2003. Cross-contamination with *Salmonella* on a broiler slaughterhouse line demonstrated by use of epidemiological markers. J Appl Microbiol **94:**826-835.
7. **Litrup, E., M. Torpdahl, B. Malorny, S. Huehn, H. Christensen, and E. M. Nielsen.** 2010. Association between phylogeny, virulence potential and serovars of *Salmonella enterica*. Infect Genet Evol **10:**1132-1139.
8. **Hernández, E., J. L. Rodriguez, S. Herrera-León, I. García, V. de Castro, and N. Muniozguren.** 2012. *Salmonella* Paratyphi B var Java infections associated with

exposure to turtles in Bizkaia, Spain, September 2010 to October 2011. Euro Surveill **17**.

9. **Achtman, M., J. Wain, F. X. Weill, S. Nair, Z. Zhou, V. Sangal, M. G. Krauland, J. L. Hale, H. Harbottle, A. Uesbeck, G. Dougan, L. H. Harrison, S. Brisse, and S. E. M. S. Group.** 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog **8:**e1002776.

10. **Clinical and Laboratory Standards Institute** 2006. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically. Approved Standard 7th ed. (M7-A7) vol. 23. Clinical and Laboratory Standards Institute, Wayne, USA.

11. **Kidgell, C., U. Reichard, J. Wain, B. Linz, M. Torpdahl, G. Dougan, and M. Achtman.** 2002. *Salmonella typhi*, the causative agent of typhoid fever, is approximately 50,000 years old. Infect Genet Evol **2:**39-45.

12. **Ribot, E. M., M. A. Fair, R. Gautom, D. N. Cameron, S. B. Hunter, B. Swaminathan, and T. J. Barrett.** 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. Foodborne Pathog Dis **3:**59-67.

13. **Huehn, S., C. Bunge, E. Junker, R. Helmuth, and B. Malorny.** 2009. Poultry-associated *Salmonella enterica* subsp. *enterica* serovar 4,12:d:- reveals high clonality and a distinct pathogenicity gene repertoire. Appl Environ Microbiol **75:**1011-1020.

14. **Hauser, E., E. Junker, R. Helmuth, and B. Malorny.** 2011. Different mutations in the oafA gene lead to loss of O5-antigen expression in *Salmonella enterica* serovar Typhimurium. J Appl Microbiol **110:**248-253.

15. **Malorny, B., C. Bunge, and R. Helmuth.** 2003. Discrimination of d-tartrate-fermenting and -nonfermenting *Salmonella enterica* subsp. *enterica* isolates by genotypic and phenotypic methods. J Clin Microbiol **41:**4292-4297.

16. **Lim, Y. H., K. Hirose, H. Izumiya, E. Arakawa, H. Takahashi, J. Terajima, K. Itoh, K. Tamura, S. I. Kim, and H. Watanabe.** 2003. Multiplex polymerase chain reaction assay for selective detection of *Salmonella enterica* serovar Typhimurium. Jpn J Infect Dis **56:**151-155.

17. **Hughes, K. T., P. Youderian, and M. I. Simon.** 1988. Phase variation in *Salmonella*: analysis of Hin recombinase and hix recombination site interaction in vivo. Genes Dev **2:**937-948.

18. **Switt, A. I., Y. Soyer, L. D. Warnick, and M. Wiedmann.** 2009. Emergence, distribution, and molecular and phenotypic characteristics of *Salmonella enterica* serotype 4,5,12:i. Foodborne Pathog Dis **6:**407-415.

19. **Hauser, E., E. Tietze, R. Helmuth, E. Junker, K. Blank, R. Prager, W. Rabsch, B. Appel, A. Fruth, and B. Malorny.** 2010. Pork contaminated with *Salmonella enterica* serovar 4,[5],12:i:-, an emerging health risk for humans. Appl Environ Microbiol **76:**4601-4610.

20. **Guibourdenche, M., P. Roggentin, M. Mikoleit, P. I. Fields, J. Bockemühl, P. A. Grimont, and F. X. Weill.** 2010. Supplement 2003-2007 (No. 47) to the White-Kauffmann-Le Minor scheme. Res Microbiol **161:**26-29.

21. **Toboldt, A., E. Tietze, R. Helmuth, A. Fruth, E. Junker, and B. Malorny.** 2012. Human infections attributable to the D-tartrate-fermenting variant of *Salmonella enterica* serovar Paratyphi B in Germany originate in reptiles and, on rare occasions, poultry. Appl Environ Microbiol **78:**7347-7357.

493 **TABLE 1** Number and source of *S. enterica* serovar 4,[5],12:b:- isolates in Germany received by the NRL-Salm and NRZ-RKI

Year of isolation	No. all isolates		No. of 4,[5],12:b:- isolates (source)	
	NRL	NRZ	4,5,12:b: -	4,12:b: -
2000	3915	6696	7 (human), 2 (sheep), 1 (spice)	1 (human),
2001	3605	7635	3 (human), 1 (poultry meat), 1 (spice)	2 (reptile)
2002	4411	6300	8 (human), 3 (dried mushroom), 1 (reptile)	1 (shellfish)
2003	3630	3930	4 (human), 1 (fish/fish product), 1 (dried mushroom)	–
2004	3604	3691	2 (human), 1 (shellfish)	3 (poultry)
2005	4090	3655	6 (human), 1(pet bird)	1 (poultry)
2006	3887	3333	7 (human), 1 (dried mushroom), 2 (others[1])	–
2007	3955	3855	2 (human), 2 (others[1])	1 (poultry), 3 (others[1]),
2008	3606	3205	4 (human), 1 (pig)	1 (human), 2 (poultry)
2009	4111	3646	21 (human)	–
2010	4631	2320	12 (human), 1 (shellfish)	1 (human), 4 (poultry), 1 (other[1])
2011	3793	2439	17 (human), 1 (reptile)	1 (reptile)
Total	47238	50705	93 (human), 5 (dried mushroom), 1 (pet bird), 1 (pig), 1 (poultry meat), 2 (reptile), 3 (shellfish/fish), 2 (sheep), 2 (spice), 4 (others[1])	3 (human), 11 (poultry), 3 (reptile), 1 (shellfish), 4 (others[1])

494 [1] non animal origin

495 **TABLE 2** *S. enterica* serovar 4,[5],12:b:- strains used for phenotypic and molecular analysis in this study

Strain no.	Year of isolation	Origin	Resistance[5]	PFGE cluster	PFGE profile no.	MLST	Microarray (PAT)	O-antigen
01-02861	2001	Food, spicery	SMX	A	13	423	5	4,5,12
02-00002	2001	Food, dried mushroom	susceptible	A	6	135	NT	4,5,12
02-00059	2002	Food, dried mushroom	susceptible	A	9	42	2	4,5,12
03-01178	2003	Food, dried mushroom	susceptible	A	6	135	NT	4,5,12
04-01012	2003	Shellfish, shrimps	susceptible	A	15	1582	6	4,5,12
06-03656	2005	Sludge	susceptible	A	7	42	NT	4,5,12
06-02764	2006	Food, dried mushroom	susceptible	A	5	135	7	4,5,12
10-00322	2009	Shellfish, Black Tiger Prawn	susceptible	A	16	42	NT	4,5,12
11-00612	2011	Reptile, turtle, faeces	susceptible	A	18	1484	NT	4,12[3]
11-02445	2000	Human	susceptible	A	11	135	NT	4,12[3]
11-02470	2000	Human	susceptible	A	3	423	NT	4,5,12
11-02473	2001	Human	susceptible	A	10	1589	12	4,5,12
11-01473	2002	Human	susceptible	A	14	42	6	4,5,12
11-02476	2003	Human	susceptible	A	19	1588	10	4,5,12
11-02482	2005	Human	susceptible	A	12	135	NT	4,5,12
11-02486	2006	Human	susceptible	A	1	135	NT	4,5,12
11-02487	2007	Human	susceptible	A	4	135	NT	4,5,12
11-02491	2008	Human	susceptible	A	11	135	NT	4,5,12
11-02493	2009	Human	susceptible	A	17	42	NT	4,5,12
11-02497	2009	Human	susceptible	A	2	42	9	4,5,12
11-02501	2010	Human	susceptible	A	8	42	NT	4,5,12
01-02664	2001	Reptile, organ	susceptible	B	20	1583	14	4,12[3]
00-02409	2000	Food, spicery	susceptible	C	40	42	NT	4,5,12
01-00189	2000	Chicken, meat	susceptible	C	21	423	NT	4,5,12
02-00052	2002	Food, dried mushroom	susceptible	C	33	423	NT	4,5,12
02-04446	2002	Reptile, *Boiga dendrophila*	susceptible	C	37	42	1	4,5,12
02-04643	2002	Shellfish	susceptible	C	23	42	3	4,12[3]
03-03172	2003	Fish or fish product	susceptible	C	23	42	NT	4,5,12
11-01531	2004	Food, vegetable	susceptible	C	24	423	4	4,5,12
07-03684	2007	Feed	susceptible	C	38	42	8	4,5,12
07-03824	2007	Fertilizer	susceptible	C	32	42	NT	4,5,12
08-00880	2008	Pig	susceptible	C	35	42	NT	4,5,12
11-02471	2001	Human	susceptible	C	25	423	2	4,5,12
11-01471	2002	Human	susceptible	C	27	1578	11	4,5,12

11-02475	2002	Human	susceptible	C	28	423	NT	4,5,12
11-02479	2004	Human	susceptible	C	22	423	NT	4,5,12
11-02480	2004	Human	susceptible	C	34	423	2	4,5,12
11-02485[1,2]	2006	Human	susceptible	C	26	42	9	4,5,12
11-02488	2007	Human	susceptible	C	29	423	NT	4,5,12
11-02490	2008	Human	susceptible	C	29	423	2	4,5,12
11-02494	2009	Human	susceptible	C	31	423	NT	4,5,12
11-02496	2009	Human	susceptible	C	39	42	NT	4,5,12
11-02498	2010	Human	susceptible	C	29	423	NT	4,5,12
11-02499	2010	Human	susceptible	C	38	42	NT	4,5,12
11-02500	2010	Human	susceptible	C	30	423	NT	4,5,12
11-02502	2011	Human	susceptible	C	36	423	2	4,5,12
11-02503	2011	Human	susceptible	C	22	423	NT	4,5,12
11-02504	2011	Human	susceptible	C	29	423	NT	4,5,12
11-02505	2011	Human	susceptible	D	41	679	13	4,5,12
00-02320	2000	Sheep	SMX	E	45	127	16	4,5,12
05-00829	2005	Pet bird, faeces	susceptible	E	47	42	NT	4,5,12
11-02467[1,2]	2000	Human	susceptible	E	42	88	17	4,5,12
11-02474[1]	2002	Human	susceptible	E	45	127	NT	4,5,12
11-02478[1]	2003	Human	susceptible	E	43	127	15	4,5,12
11-02483[1,2]	2005	Human	AMP CHL KAN SMX STR TET	E	44	127	16	4,5,12
11-02489[1]	2008	Human	susceptible	E	45	127	16	4,5,12
11-02492[1]	2009	Human	susceptible	E	45	127	NT	4,5,12
11-02495[1]	2009	Human	susceptible	E	46	127	NT	4,5,12
11-02464[1,2]	2010	Human	susceptible	E	45	127	16	4,12[3]
11-02465[1]	2010	Human	susceptible	E	45	127	NT	4,5,12
07-01889-2	2007	Bird, quail, faeces	NAL SMX STR TET	F	48	1484	NT	4,12[4]
04-02058	2004	Turkey	SMX	G	52	1484	18	4,12[4]
07-01980	2007	Feed	susceptible	G	51	1583	NT	4,12[4]
08-02676	2008	Chicken, environment	susceptible	G	52	1484	NT	4,12[4]
10-01745	2010	Turkey, meat	susceptible	G	49	1484	18	4,12[4]
10-01843	2010	Turkey	susceptible	G	50	1484	NT	4,12[4]
11-02460	2008	Human	susceptible	G	52	1484	18	4,12[4]

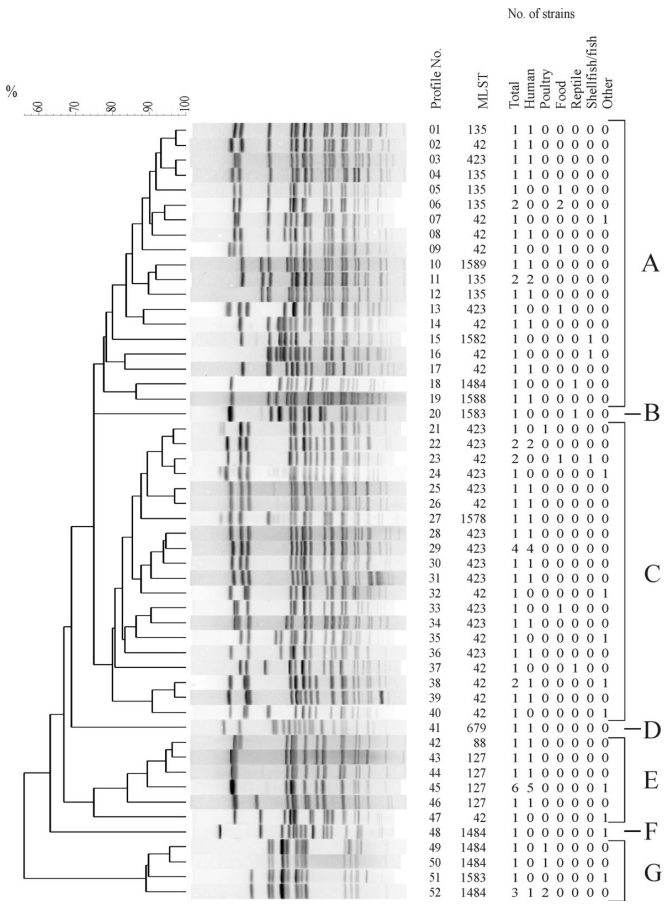
[1] positive for *fljB*-1,2 according to Lim et al. 2003 (16).
[2] *fljB* region sequenced. GenBank accession no. see Material and Methods.
[3] 7 bp deletion in *oafA* gene leading to non-functional O:5 antigen.
[4] *oafA* gene complete absence.
[5] Abbreviations: ampicillin (AMP), chloramphenicol (CHL), kanamycin (KAN), nalidixic acid (NAL), sulfamethoxazole (SMX), streptomycin (STR), tetracycline (TET).

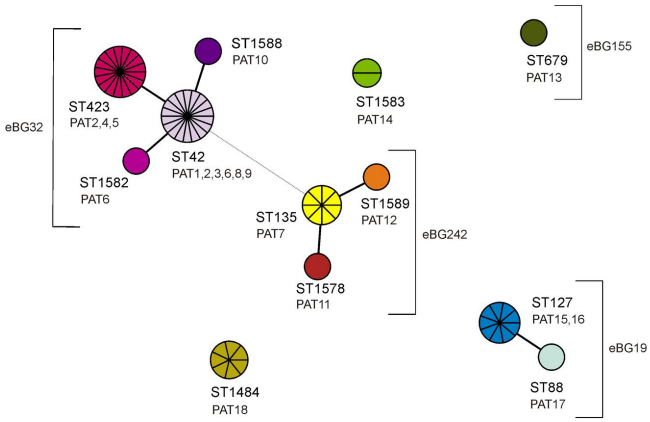
FIGURE LEGENDS

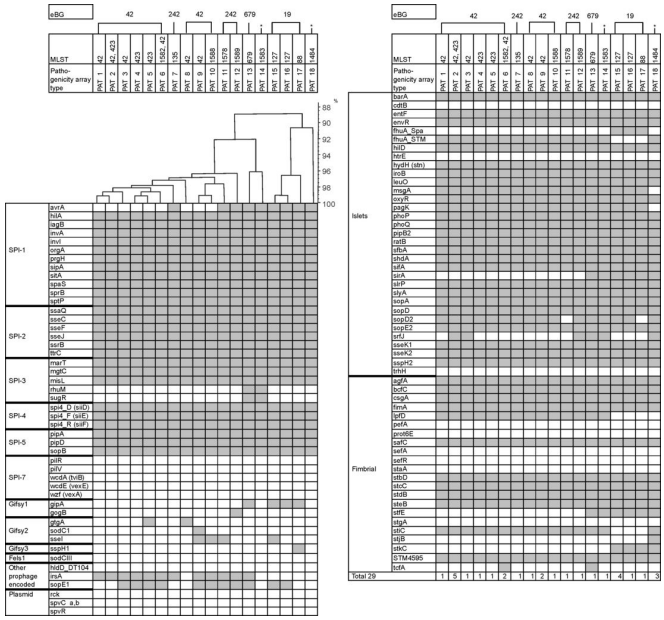
FIG. 1. UPGMA dendrogram of PFGE profiles identified in 67 *S. enterica* serovar 4,[5],12:b:- strains after digestion with XbaI. Profiles were numbered serially from 1 to 52. The number of strains belonging to each source (total, human, poultry, food, reptile, shellfish/fish, and other) and corresponding MLSTs are shown on the right side. Assigned clusters A to G are indicated by square brackets.

FIG. 2. Minimal spanning tree of MLST data on 67 *S. enterica* serovar 4,[5],12:b:- isolates. Each circle refers to one ST subdivided into one pie slice per strain. STs that share six identical alleles are linked by a black line. STs sharing three alleles are linked by a grey dashed line. Based on their similarity, STs were grouped in three complexes, of these two were already described as eBG groups according to the nomenclature of Achtman et al. (9). Pathogenicity array types (PATs) found in each ST are shown below designations.

FIG. 3. Virulence determinants microarray for 29 *S. enterica* serovar 4,[5],12:b:- strains analyzed. On the left side the analyzed genes are indicated and grouped according to their particular genomic location (SPI-1 to SPI-7; Prophages Gifsy-1, Gifsy-2, Gifsy-3 and Fels-1; plasmids and islets) or function (fimbrial). At the top assigned pathogenicity array types (PATs) and corresponding eBGs are indicated. The asterisks show the STs that have not yet belong to an eBG. The hybridization result of each type is shown by row. A white box indicates the absence and a grey box indicates the presence of the target sequence. SPI, *Salmonella* Pathogenicity Island.


%
60
70
80
90
100
No. of strains
Profile No.
MLST
Total
Human
Poultry
Food
Reptile
Shellfish/fish
Other
01 135 1 1 0 0 0 0 0
02 42 1 1 0 0 0 0 0
03 423 1 1 0 0 0 0 0
04 135 1 1 0 0 0 0 0
05 135 1 0 0 1 0 0 0
06 135 2 0 0 2 0 0 0
07 42 1 0 0 0 0 0 1
08 42 1 1 0 0 0 0 0
09 42 1 0 0 1 0 0 0
10 1589 1 1 0 0 0 0 0
11 135 2 2 0 0 0 0 0
12 135 1 1 0 0 0 0 0
13 423 1 0 0 1 0 0 0
14 42 1 1 0 0 0 0 0
15 1582 1 0 0 0 0 1 0
16 42 1 0 0 0 0 1 0
17 42 1 1 0 0 0 0 0
18 1484 1 0 0 0 1 0 0
19 1588 1 1 0 0 0 0 0
20 1583 1 0 0 0 1 0 0
21 423 1 0 1 0 0 0 0
22 423 2 2 0 0 0 0 0
23 42 2 0 0 1 0 1 0
24 423 1 0 0 0 0 0 1
25 423 1 1 0 0 0 0 0
26 42 1 1 0 0 0 0 0
27 1578 1 1 0 0 0 0 0
28 423 1 1 0 0 0 0 0
29 423 4 4 0 0 0 0 0
30 423 1 1 0 0 0 0 0
31 423 1 1 0 0 0 0 0
32 42 1 0 0 0 0 0 1
33 423 1 0 0 1 0 0 0
34 423 1 1 0 0 0 0 0
35 42 1 0 0 0 0 0 1
36 423 1 1 0 0 0 0 0
37 42 1 0 0 0 1 0 0
38 42 2 1 0 0 0 0 1
39 42 1 1 0 0 0 0 0
40 42 1 0 0 0 0 0 1
41 679 1 1 0 0 0 0 0
42 88 1 1 0 0 0 0 0
43 127 1 1 0 0 0 0 0
44 127 1 1 0 0 0 0 0
45 127 6 5 0 0 0 0 1
46 127 1 1 0 0 0 0 0
47 42 1 0 0 0 0 0 1
48 1484 1 0 0 0 0 0 1
49 1484 1 0 1 0 0 0 0
50 1484 1 0 1 0 0 0 0
51 1583 1 0 0 0 0 0 1
52 1484 3 1 2 0 0 0 0
A
B
C
D
E
F
G


ST1588
PAT10
ST679
PAT13
eBG155
ST1583
PAT14
ST423
PAT2,4,5
eBG32
ST42
PAT1,2,3,6,8,9
ST1582
PAT6
ST135
PAT7
ST1589
PAT12
eBG242
ST1578
PAT11
ST127
PAT15,16
eBG19
ST1484
PAT18
ST88
PAT17


eBG
42
242
42
242
679
19
MLST
Patho-genicity array type
PAT 1
PAT 2
PAT 3
PAT 4
PAT 5
PAT 6
PAT 7
PAT 8
PAT 9
PAT 10
PAT 11
PAT 12
PAT 13
PAT 14
PAT 15
PAT 16
PAT 17
PAT 18
42
42, 423
42
423
423
1562, 42
135
42
42
1568
1578
1569
679
1563
127
127
88
1464
%
88
90
92
94
96
98
100
SPI-1
avrA
hilA
sprB
hilA
invA
invI
orgA
prgH
spaA
sitA
spaS
sprB
spiP
SPI-2
ssaQ
sseC
sseF
sseJ
ssrB
ssrA
SPI-3
mgtF
mgtC
misL
rhuM
sugR
SPI-4
spi4_D (siiD)
spi4_F (siiE)
spi4_R (siiF)
SPI-5
pipA
pipD
sopB
SPI-7
pilR
pilV
wcdA (tviB)
wcaE (vexE)
wzf (vexA)
Gifsy1
gipA
gogB
Gifsy2
gtgA
sodC1
sseI
Gifsy3
ssph1
Fels1
sodCIII
Other prophage encoded
fidD_DT104
rsA
sopE1
Plasmid
trk
spvC a,b
spvR
Islets
barA
cdtB
entF
envR
fhuA_Spa
fhuA_STM
hilD
hlyE
hydH (stn)
iroB
leuO
msgA
oxyR
pagK
phoP
phoQ
prpB2
ratB
sfbA
shdA
sifA
sirA
slrP
slyA
sopA
sopD
sopD2
sopE2
srfJ
sseK1
sseK2
sspH2
tlrhH
Fimbrial
sgfA
bcfC
csgA
fimA
safD
pefA
prot6E
safC
safA
safR
stiaA
stbD
stcC
stdB
stxB
stfE
stgA
stkC
stjB
stkC
STM4595
bfA
Total 29
1
5
1
1
1
2
1
1
2
1
1
1
1
4
1
1
3