Available online at www.sciencedirect.com
ScienceDirect
Journal of Microbiological Methods 69 (2007) 23–36
Journal of Microbiological Methods
www.elsevier.com/locate/jmicmeth

MLST-v, multilocus sequence typing based on virulence genes, for molecular typing of *Salmonella enterica* subsp. *enterica* serovars

B. Tankouo-Sandjong [a], A. Sessitsch [a], E. Liebana [b], C. Kornschober [c], F. Allerberger [c], H. Hächler [d], L. Bodrossy [a,*]

[a] *ARC Seibersdorf Research GmbH, Department of Bioresources, Seibersdorf, Austria*
[b] *Department of Food and Environmental Safety, Veterinary Laboratories Agency, Weybridge, UK*
[c] *Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH, Humanmedizin, Graz, Austria*
[d] *NENT, Institute for Medical Microbiology, Kantonsspital, Luzern, Switzerland*

Received 19 July 2006; received in revised form 8 November 2006; accepted 20 November 2006
Available online 8 January 2007

Abstract

Salmonella enterica subsp. *enterica* is one of the main causative agents of food-borne disease in man, and can also be the cause of serious systemic illness. Organisms belonging to this genus have traditionally been classified on the basis of the antigenic properties of the cell-surface lipopolysaccharide and of the phase 1 and phase 2 flagellar proteins. Primary isolation, biochemical identification, and serotyping are laborious and time consuming. Molecular identification based on suitable marker genes could be an attractive alternative to conventional bacteriological and serological methods. We have assessed the applicability of two housekeeping genes, *gyrB*, *atpD*, in combination with the flagellin genes *fliC* and *fljB* in multilocus sequence typing of *Salmonella*. Sequencing and comparative analysis of sequence data was performed on multiple strains from Austria, the United Kingdom, and Switzerland, representing all subspecies and 22 of the more prevalent non-typhoid *S. enterica* subsp. *enterica* serovars. A combination of these four marker genes allowed for a clear differentiation of all the strains analysed, indicating their applicability in molecular typing. The term MLST-v, for multilocus sequence typing based on virulence genes, is proposed to distinguish this approach from MLST based solely on housekeeping genes. An assortative recombination of the *fliC* gene was found in seven of the analysed serovars indicating multiple phylogenetic origin of these serovars.

Keywords: *atpD*; *fliC*; *fljB*; *gyrB*; Multilocus sequence typing; *Salmonella*

1. Introduction

Salmonella is a large genus, and serotyping is widely used to classify isolates into serogroups according to their surface antigen variability. Serotyping is based on the immunological classification of the lipopolysaccharide moieties (O antigen), the flagellar protein (H antigen), and the capsular polysaccharide (Vi antigen). The Kauffmann–White scheme, generally used for the classification of *Salmonella* serotypes, recognises 46 O serogroups and 114 H antigens resulting in 2523 characterized serotypes (Popoff et al., 2003). The genus *Salmonella* is comprised of 2 species: *S. enterica* (with six subspecies — I, II, IIIa, IIIb, IV and VI), and *S. bongori*, now considered as a separate species (Brenner et al., 2000). Different genetic studies have indicated horizontal transfer events of chromosomal genes leading to the emergence of novel serovars (Beltran et al., 1988; Li et al., 1994; Porwollik et al., 2004).

Serotyping is a reliable, epidemiologically congruent, well-established methodology for *Salmonella* typing, but is rather time consuming and requires a high number of specific antisera. Alternative molecular *Salmonella* typing methods are being developed in several laboratories (Echeit et al., 2002; Torpdahl et al., 2005; Kotetishvili et al., 2002; Sukhnanand et al., 2005; Herrera-Leon et al., 2004), aiming for higher sensitivity and specificity, better reproducibility. Multilocus sequence typing is one of the preferred methods, the existing MLST schemes being based on housekeeping genes. While these schemes reflect the

* Corresponding author. Mailing address: ARC Seibersdorf Research GmbH, Department of Bioresources, A-2444 Seibersdorf, Austria. Tel.: +43 50550 3548; fax: +43 50550 3444.
E-mail address: levente.bodrossy@arcs.ac.at (L. Bodrossy).

0167-7012/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2006.11.013

Table 1
Salmonella strains used in this study

Strains	Antigenic formulae				Country of origin	Source
	O	O	H1	H2		
Agona_1.461_03	B	4.12	fgs	–	Austria	Human
Agona_S06192_02	B	4.12	fgs	–	UK	Swine
Agona_S08219_02	B	4.12	fgs	–	UK	Cattle
Anatum_4.313_03	E1	3.1	eh	1.6	Austria	Human
Anatum_8.604_03	E1	3.1	eh	1.6	Austria	Duck
Braenderup_1.324_03	C1	6.7	eh	e,n,z15	Austria	Human
Braenderup_S03179_03	C1	6.7	eh	e,n,z15	UK	Chicken
Braenderup_S06429_03	C1	6.7	eh	e,n,z15	UK	Chicken
Brandenburg_6.229_03	B	1,4,12	lv	e,n,z15	Austria	Turkey
Brandenburg_8.750_03	B	1,4,12	lv	e,n,z15	Austria	Human
Brandenburg_S03387_03	B	1,4,12	lv	e,n,z15	UK	Horse
Brandenburg_S07129_02	B	1,4,12	lv	e,n,z15	UK	Swine
Bredeney_10.181_03	B	1,4,12,27	lv	1.7	Austria	Human
Bredeney_8.146_03	B	1,4,12,27	lv	1.7	Austria	Activated sludge
Bredeney_S06106_03	B	1,4,12,27	lv	1.7	UK	Turkey
Bredeney_S07340_02	B	1,4,12,27	lv	1.7	UK	Chicken
Cerro_S00366_03	K	6,4,18	z4,z23	[1,5]	UK	Feed of vegetable origin
Cerro_S01705_03	K	6,4,18	z4,z23	[1,5]	UK	Human
Derby_8.598_03	B	1,4,12	fg	[1,2]	Austria	Turkey
Derby_9.235_03	B	1,4,12	fg	[1,2]	Austria	Turkey
Derby_S05983_03	B	1,4,12	fg	[1,2]	UK	Dog
Derby_S06116_03	B	1,4,12	fg	[1,2]	UK	Swine
Dublin_1.160_03	D1	1,4,12[Vi]	gp	–	Austria	Human
Dublin_995_03	D1	1,4,12[Vi]	gp	–	Austria	Calf
Dublin_S06124_03	D1	1,4,12[Vi]	gp	–	UK	Cattle
Dublin_S05979_03	D1	1,4,12[Vi]	gp	–	UK	Sheep
Enteritidis_LT_1c	D1	1,9,12	gm	[1,7]	Austria	Human
Enteritidis_LT_1	D1	1,9,12	gm	[1,7]	Austria	Human
Enteritidis_S06001_03	D1	1,9,12	gm	[1,7]	UK	Chicken
Enteritidis_S05954_03	D1	1,9,12	gm	[1,7]	UK	Cattle
Heidelberg_10.311_03	B	1,4,[5],12	r	1.2	Austria	Turkey
Heidelberg_S02109_03	B	1,4,[5],12	r	1.2	UK	Chicken
Infantis_1.615_03	C1	6.7	r	1.5	Austria	Human
Infantis_S09835_02	C1	6.7	r	1.5	UK	Swine
Infantis_S08458_02	C1	6.7	r	1.5	UK	Chicken
Livingstone_8.142_03	C1	6.7	d	l,w	Austria	Human
Livingstone_8.219_03	C1	6.7	d	l,w	Austria	Human
Livingstone_S06210_03	C1	6.7	d	l,w	UK	Duck
Livingstone_S08161_02	C1	6.7	d	l,w	UK	Chicken
London_2.595_03	E1	3.1	lv	1.6	Austria	Chicken
London_4.119_03	E1	3.1	lv	1.6	Austria	Swine
Manhattan_10.299_03	C2	6.8	d	1.5	Austria	Activated sludge
Manhattan_S06179_03	C2	6.8	d	1.5	UK	Swine
Manhattan_S09472_02	C2	6.8	d	1.5	UK	Swine
Mbandaka_8.676_03	C1	6.7	z10	e,n,z15	Austria	Human
Mbandaka_9.781_03	C1	6.7	z10	e,n,z15	Austria	Human
Mbandaka_S07026_03	C1	6.7	z10	e,n,z15	UK	Chicken
Montevideo_S05461_03	C1	6.7	gm[p]s	[1,2,7]	UK	Turkey
Montevideo_S05998_03	C1	6.7	gm[p]s	[1,2,7]	UK	Chicken
Newport_1.343_03	C2	6.8	eh	1.2	Austria	Horn viper
Newport_626_03	C2	6.8	eh	1.2	Austria	Human
Newport_S04530_03	C2	6.8	eh	1.2	UK	Chicken
Newport_S05991_03	C2	6.8	eh	1.2	UK	Turkey
Oranienburg_8.139_03	C1	6.7	mt	–	Austria	Human
Oranienburg_8.283_03	C1	6.7	mt	–	Austria	Human
Oranienburg_S05997_03	C1	6.7	mt	–	UK	Feed of vegetable origin
Oranienburg_S01107_03	C1	6.7	mt	–	UK	Swine
Schwarzengrund_6.906_03	B	1,4,12,27	d	1.7	Austria	Turkey
Schwarzengrund_9.528_03	B	1,4,12,27	d	1.7	Austria	Turkey
Schwarzengrund_S06199_02	B	1,4,12,27	d	1.7	UK	Cattle
Schwarzengrund_S07291_02	B	1,4,12,27	d	1.7	UK	Turkey
Thompson_10.439_03	C1	6.7	k	1.5	Austria	Chicken

Table 1 (*continued*)

Strains	Antigenic formulae				Country of origin	Source
	O	O	H1	H2		
Thompson_8.671_03	C1	6.7	k	1.5	Austria	Human
Thompson_S01835_03	C1	6.7	k	1.5	UK	Sheep
Thompson_S05026_02	C1	6.7	k	1.5	UK	Chicken
Typhimurium_LT_2	B	1,4,5,12	i	1.2	Austria	Human
Typhimurium_LT_41	B	1,4,5,12	i	1.2	Austria	Human
Typhimurium_LT_R/NC	B	1,4,5,12	i	1.2	Austria	Human
Typhimurium_LT_RDNC	B	1,4,5,12	i	1.2	Austria	Human
Typhimurium_S05917_03	B	1,4,5,12	i	1.2	UK	Cattle
Typhimurium_S05984_03	B	1,4,5,12	i	1.2	UK	Cat
Virchow_1.243_03	C1	6.7	r	1.2	Austria	Human
Virchow_S04627_03	C1	6.7	r	1.2	UK	Chicken
Virchow_S07740_02	C1	6.7	r	1.2	UK	Cattle
*S. arizonae*_N2130_04	21	21	g,z51	–	Switzerland	Ball python
*S. arizonae*_4.801_03	56	56	z4,z23	–	Austria	Hawk
*S. arizonae*_N376_04	41	41	Z4,z23	–	Switzerland	Reticulated python
*S. bongori*_N2296_04	48	48	Z35	–	Switzerland	Human
*S .diarizonae*_1.479_03	61	61	i	Z53	Austria	Snake
*S. diarizonae*_1246_04	50	50	lv	Z35	Switzerland	Environment
*S. diarizonae*_N2496_04	50	50	r	–	Switzerland	Human
*S. houtenae*_N2398_03	38	38	Z4, z23	–	Switzerland	Environment
*S .houtenae*_N737_03	43	43	Z4, z23	–	Switzerland	Environment
*S. indica*_N2576_03	6,14,24	6,14,24	a	e,n,x	Switzerland	Environment
*S. indica*_N372_03	6.14	6.14	a	e,n,z15	Switzerland	Environment
*S. salamae*_N2154_03	42	42	b	e,n,x,z15	Switzerland	Human
*S. salamae*_1.345_03	42	42	z	1.5	Austria	Mouse

Antigenic formulae were determined by conventional serotyping at the reference laboratories VLA in the UK, AGES in Austria and NENT, Switzerland, respectively.

true phylogeny of the analysed isolates, including information on the genes directly determining the serotype may confer some advantages and smoothen the interpretation of results in relation to existing epidemiology data from conventional serotyping.

The O antigen is the outermost component of the LPS (lipopolysaccharide), determined by a complex interaction of the enzymes responsible for the synthesis and assembly of the O antigen, encoded by the gene clusters *rfb*, *wzk* and *wzy* (Reeves et al., 1996). Variations of the O antigen are due to different genes of these clusters being present and absent. It is thus not possible to find a single gene which would on its own reflect the O antigenic properties of the carrying strain.

Two genes encode the flagellar antigens. The *fliC* gene that encodes the phase 1 (H1) antigen is located in one of the flagellar biosynthesis operons, is present in all salmonellae, and has a homologue in *Escherichia coli* (Macnab, 1992). The *fljB* gene encoding the phase 2 (H2) antigen is located in a region of the genome that is unique to *Salmonella* and is present in four of the six subspecies. Generally, the central part of flagellin genes is highly variable and is assumed to determine the epitope of the H antigen, whereas the 5′ and 3′ ends are typically conserved (Wei and Joys, 1985). Flagellin genes are considered useful markers for the molecular determination of the flagellar antigen type (McQuiston et al., 2004; Herrera-Leon et al., 2004).

Phylogenetic relationships of prokaryotes are most commonly based on the variability in their ribosomal RNA genes. However, for the classification of closely related species and subspecies rRNA genes are too conserved and alternative marker genes are required. As an alternative, the *gyrB* gene encoding the subunit B protein of DNA gyrase (topoisomerase type II) has been applied for the classification of bacteria including members of the Enterobacteriaceae (Fukushima et al., 2002; Dauga, 2002). Several studies demonstrated that the rate of molecular evolution of the *gyrB* gene is higher than that of ribosomal RNA genes (Dauga, 2002; Yamamoto and Harayama, 1995) enabling the design of species specific oligonucleotide probes for most Enterobacteriaceae. The *atpD* gene, encoding the beta-subunit of bacterial F1F0 type ATP-synthases, has also been shown to reflect the phylogenetic position of bacteria (Paradis et al., 2005), with a similar phylogenetic resolution power to that of the *gyrB* gene.

The objective of the present study was to perform comparative sequence analysis of relevant marker genes and to exploit the obtained information to develop a DNA-based assay for *Salmonella* serotype identification with a potential to replace traditional serotyping. The aim was a method that is reproducible, relatively free of personnel error, and generates results that are fully compatible with those from the traditional serotyping. A modified multilocus sequence typing (Maiden et al., 1998) method (considering also virulence genes, termed MLST-v for multilocus sequence typing based on virulence genes) was developed and tested against 74 *Salmonella enterica* subsp. *enterica* strains belonging to prevalent serovars occurring in Europe, plus representatives of all the other subspecies and species of the genus. In order to cover different geographical regions, strains deposited in the national *Salmonella* strain collections of the UK, Switzerland and Austria were included. Our analysis was based on housekeeping genes to enable a phylogenetic comparison, as well as on flagellin genes in order to identify serovar-specific differences. In addition, the comparative sequence analysis of housekeeping and flagellin

Table 2
PCR and sequencing primers used in this study

Primer designation	Sequence (5′- 3′)	Size of product	Reference	T annealing
PCR and sequencing primers				
atpDF (5737) (*atpD* forward primer)	TAGTTGACGTCGAATTCCCTCAGG	888 bp	(Christensen and Olsen, 1998)	55 °C
atpDR (6625) (*atpD* reverse primer)	GGAGACGGGTCAGTCAAGTCATC			
FSa1 (*fliC* forward primer)	CAAGTCATTAATAC(AC)AACAGCCTGTCGC	1500 bp	(Dauga et al., 1998)	55 °C
rFSa1 (*fliC* reverse primer)	TTAACGCAGTAAAGAGAGGACGTTTTGC			
FSa2[a] (*fljB* forward primer)	GGCACAAGTAATCAACACTAACAGTCTGT	1478 bp		58 °C
rFSa2[b] (*fljB* reverse primer)	CGTAACAGAGACAGCACGTT(CT)TG(CT)G			
PCR primers				
UP1 (*gyrB* forward primer)	GAAGTCATCATGACCGTTCTGCA(CT)GC-(ACGT)GG(ACGT)GG(ACGT)AA(AG)TT(CT)GA	1200 bp	(Yamamoto and Harayama, 1995; Fukushima et al., 2002)	63 °C
UP2r (*gyrB* reverse primer)	AGCAGGGTACGGATGTGCGAGCC(AG)TC(ACGT)-AC(AG)TC(ACGT)GC(AG)TC(ACGT)GTCAT			
Sequencing primers				
UP1S (*gyrB* (forward primer)	GAAGTCATCATGACCGTTCTGCA		(Yamamoto and Harayama, 1995)	
UP2Sr (*gyrB* reverse primer)	AGCAGGGTACGGATGTGCGAGCC			

[a] Fsa2 primer was the unique primer used in this to sequence the *fljB* gene and generated a partial sequence of 535 nucleotides.
[b] rFSa2 was not used in this study to sequence the *fljB* gene.

genes gave insight into the evolutionary relationships of *Salmonella* serotypes and indicated the occurrence of horizontal gene transfer in the flagellin genes of some serovars.

2. Materials and methods

2.1. Bacterial strains and culture method

Salmonella strains used in this study are listed in Table 1. *Salmonella* strains were obtained from national *Salmonella* reference laboratories of Austria, the United Kingdom and Switzerland. Strains were grown aerobically in 10% TSA at 37 °C.

2.2. Genomic DNA extraction

Cells from overnight cultures were harvested and chromosomal DNA for PCR was prepared by using the UltraClean™ Microbial DNA Isolation kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) according to the manufacturer's instructions.

2.3. PCR amplification

PCR amplification of *gyrB*, *atpD*, *fliC* and *fljB* from chromosomal DNA was performed by using the PCR primers UP1 and UP2r (Yamamoto and Harayama, 1995; Fukushima et al., 2002), atpDF(5737) and atpDR(6625) (Christensen and Olsen, 1998), FSa1 and rFSa1, FSa2 and rFSa2 (Dauga et al., 1998), respectively. For each target, PCR reactions of 50 µl volume each, consisting of 1x PCR buffer, 1.5 mM $MgCl_2$, 50 nM for each four dNTPs, 15 pmoles of both primers, 1 ng genomic DNA as template, and 1 U of Taq polymerase (Invitrogen, Carlsbad, CA, USA), were performed in a Hybaid Combi Thermal Reactor TR2 using *Taq* DNA polymerase in accordance with the manufacturer's instructions. Amplification parameters applied were those described in the corresponding publications, shown in Table 2.

2.4. DNA sequencing

For sequencing, *gyrB*, *atpD*, *fliC* and *fljB* genes were PCR amplified using the conditions described above. PCR products were purified using the High Pure PCR product purification kit (Roche, Basel, Switzerland) according to the manufacturer's instructions and used as templates in sequencing reactions. Sequencing of *atpD*, *fliC* and *fljB* was performed by using the respective primer pair used for PCR amplification. The *gyrB* gene was sequenced by using the primers UP1S and UP2Sr (targeting the 5′ sequencing tags on the PCR primers), due to the high ambiguity of the PCR primers themselves (Yamamoto

Table 3
Characteristics of the marker genes used in this study

Gene	Size for analysis (bp)[a]	*Salmonella typhimurium* LT_2 position[b]	G+C%	No. of alleles	No. of sequences analysed[c]	No. of polymorphic sites (%)
atpD	763	105–867	55	45	74	53 (6.9)
gyrB	809	500–1308	55	38	74	84 (10.4)
fliC	1321 (1063–1314)	90–1371	47	27	74	923 (69.9)
fljB	535 (529–535)	95–630	52	16	51	105 (19.6)[d]

[a] Numbers in parentheses indicate the effective lengths not counting alignment gaps.
[b] Nucleotide positions on the corresponding gene of *Salmonella* Typhimurium LT_2, Acc.No. NC_003197.
[c] Only *Salmonella enterica* subsp. *enterica* serovars were included in the analysis.
[d] Only a partial sequence of the *fljB* gene has been included in the analysis, excluding some of the more variable regions.

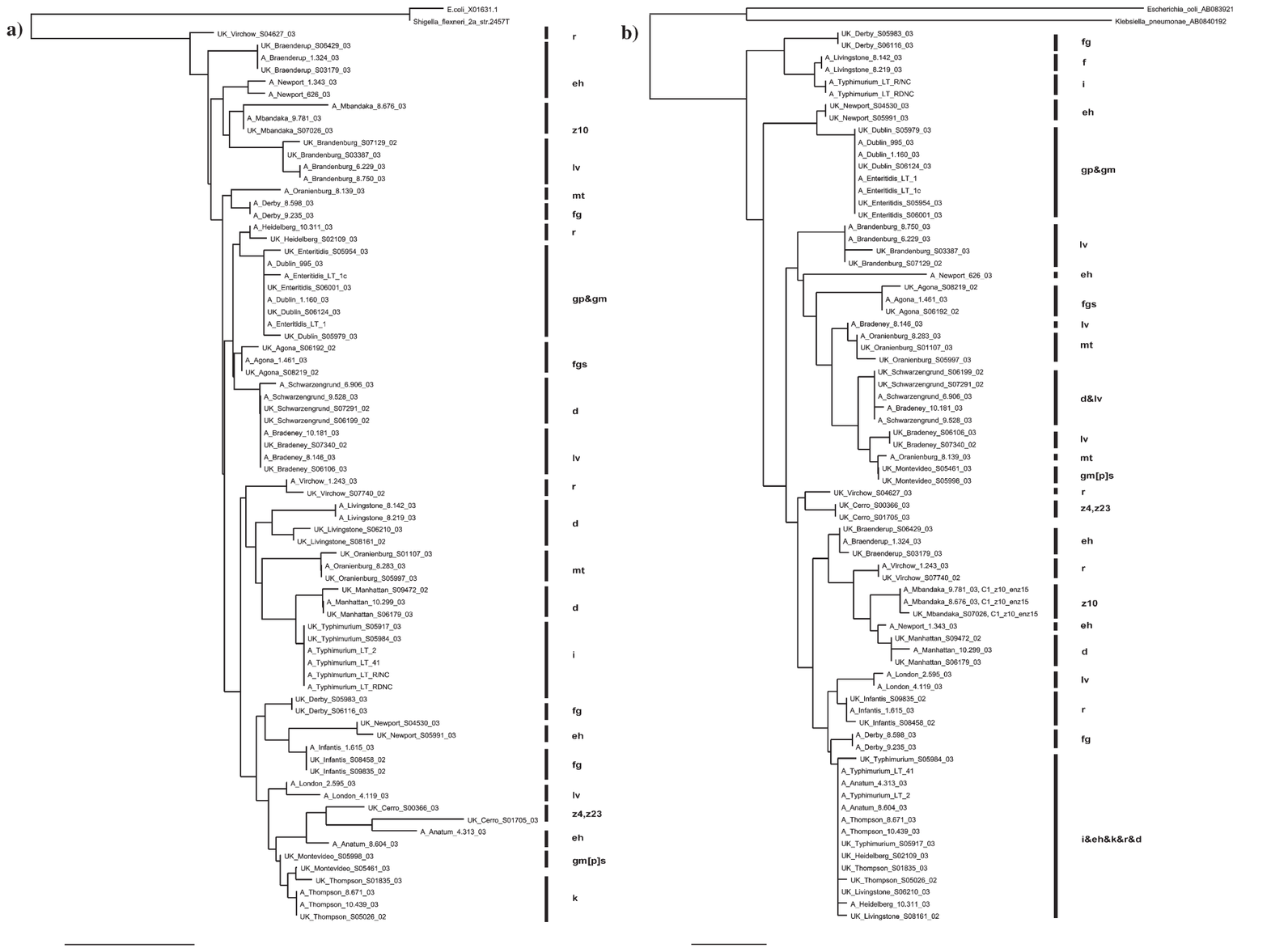

Fig. 1. Neighbor-joining dendrogramms based on nucleotide sequences of the *atpD* (a) and *gyrB* (b) genes, respectively. The positions considered to generate the tree are shown in Table 3. H1 antigen categories are indicated. The bar represents 10% evolutionary sequence divergence.

Table 4
MLST differentiation of *S. enterica* subsp. *enterica* serovars

Strains	Allelic profile for:							ST	Assortative recombination
	gyrB	*atpD*	fliC=C1′+V+C2′				*fljB*		
			C1′	V′	C2′	C1′+V+C2′			
Agona_1.461_03	1	1	1	1	1	1	N.A.	1	
Agona_S06192_02	2	2	1	1	1	1	N.A.	2	
Agona_S08219_02	2	1	1	1	1	1	N.A.	3	
Anatum_4.313_03	3	3	2	2	2	2	1	4	
Anatum_8.604_03	3	4	2	2	2	2	1	5	
Braenderup_1.324_03	4	5	3	3	2	3	2	6	
Braenderup_S03179_03	5	5	3	3	2	3	2	7	
Braenderup_S06429_03	4	5	3	3	2	3	3	8	
Brandenburg_6.229_03	6	6	4	5	4	4	4	9	
Brandenburg_8.750_03	6	6	4	5	4	4	4	9	
Brandenburg_S03387_03	7	7	4	5	4	4	4	10	
Brandenburg_S07129_02	6	8	5	5	4	5	5	11	
Bredeney_10.181_03	8	9	6	6	4	6	6	12	+
Bredeney_8.146_03	9	9	4	5	4	7	7	13	+
Bredeney_S06106_03	10	9	4	5	4	7	6	14	+
Bredeney_S07340_02	10	9	4	5	4	7	6	14	+
Cerro_S00366_03	11	10	7	7	5	8	N.A.	15	
Cerro_S01705_03	11	11	7	7	5	8	N.A.	16	
Derby_8.598_03	12	12	8	8	6	9	N.A.	17	+
Derby_9.235_03	12	12	8	8	6	9	N.A.	17	+
Derby_S05983_03	13	13	9	9	6	10	N.A.	18	+
Derby_S06116_03	13	13	9	9	6	10	N.A.	18	+
Dublin_1.160_03	14	14	10	10	7	11	N.A.	19	
Dublin_995_03	14	14	10	10	7	11	N.A.	19	
Dublin_S06124_03	14	14	10	10	7	11	N.A.	19	
Dublin_S05979_03	14	15	10	10	8	12	N.A.	20	
Enteritidis_LT_1c	14	16	10	11	9	13	N.A.	21	
Enteritidis_LT_1	14	14	10	11	7	14	N.A.	22	
Enteritidis_S06001_03	14	14	10	11	7	14	N.A.	22	
Enteritidis_S05954_03	14	17	10	11	7	14	N.A.	23	
Heidelberg_10.311_03	15	18	11	12	10	15	8	24	
Heidelberg_S02109_03	3	19	11	13	10	16	8	25	
Infantis_1.615_03	16	20	12	14	10	17	9	26	
Infantis_S09835_02	16	20	12	14	10	17	9	26	
Infantis_S08458_02	17	20	12	14	10	17	9	27	
Livingstone_8.142_03	18	21	13	15	11	18	10	28	+
Livingstone_8.219_03	18	21	13	15	11	18	11	29	+
Livingstone_S06210_03	19	22	14	16	12	19	12	30	+
Livingstone_S08161_02	20	23	14	16	12	19	12	31	+
London_2.595_03	21	24	15	17	13	20	13	32	
London_4.119_03	22	25	15	17	13	20	13	33	
Manhattan_10.299_03	23	26	16	18	14	21	14	34	
Manhattan_S06179_03	24	26	16	18	14	21	14	35	
Manhattan_S09472_02	24	27	16	18	14	21	14	36	
Mbandaka_8.676_03	25	28	17	19	15	22	15	37	
Mbandaka_9.781_03	25	29	17	19	15	22	15	38	
Mbandaka_S07026_03	26	29	17	19	15	22	15	39	
Montevideo_S05461_03	27	30	18	20	16	23	N.A.	40	
Montevideo_S05998_03	27	31	18	20	16	23	N.A.	41	
Newport_1.343_03	28	32	19	4	3	24	16	42	+
Newport_626_03	29	33	19	3	2	25	16	43	+
Newport_S04530_03	30	34	19	3	2	25	16	44	+
Newport_S05991_03	30	35	20	3	2	26	16	45	+
Oranienburg_8.139_03	31	36	18	21	17	27	N.A.	46	+
Oranienburg_8.283_03	32	37	18	22	17	28	N.A.	47	+
Oranienburg_S05997_03	33	37	18	22	17	28	N.A.	47	+
Oranienburg_S01107_03	32	38	18	23	17	29	N.A.	48	+
Schwarzengrund_6.906_03	34	39	21	24	14	30	6	49	
Schwarzengrund_9.528_03	34	9	21	24	14	30	6	50	
Schwarzengrund_S06199_02	34	9	21	24	14	30	6	50	

(continued on next page)

Table 4 (*continued*)

Strains	Allelic profile for:							ST	Assortative recombination
	gyrB	*atpD*	fliC=C1′+V+C2′				*fljB*		
			C1′	V′	C2′	C1′+V+C2′			
Schwarzengrund_S07291_02	34	9	21	24	14	30	6	50	
Thompson_10.439_03	19	40	22	25	15	31	14	51	
Thompson_8.671_03	19	40	22	26	15	32	14	51	
Thompson_S01835_03	19	41	22	25	15	31	14	52	
Thompson_S05026_02	35	40	22	27	15	33	14	53	
Typhimurium_LT_2	19	42	23	28	10	34	8	54	+
Typhimurium_LT_41	19	42	23	28	10	34	8	54	+
Typhimurium_LT_R/NC	36	42	23	28	10	34	8	55	+
Typhimurium_LT_RDNC	36	42	23	28	10	34	8	55	+
Typhimurium_S05917_03	19	42	23	29	10	35	8	56	+
Typhimurium_S05984_03	37	42	23	29	10	35	8	57	+
Virchow_1.243_03	38	43	24	30	10	36	16	58	+
Virchow_S04627_03	39	44	25	30	10	37	16	59	+
Virchow_S07740_02	38	45	24	30	10	36	16	60	+

Occurrence of gene transfer and recombination of the entire *fliC* gene (assortative recombination) is also indicated.

and Harayama, 1995). Primers for amplification and sequencing of *atpD*, *gyrB*, *fliC* and *fljB* genes are listed in Table 2. Sequencing was performed by the dideoxy chain termination method using an ABI 373A automated DNA sequencer and the ABI PRISM Big Dye terminator Cycle Sequencing Kit (PE Applied Biosystems Inc., Foster City, CA, USA).

2.5. DNA sequence analysis

The resulting sequences and additionally those downloaded from GenBank were imported into the freely available ARB software package (Ludwig et al., 2004) and properly aligned according to the alignment of amino acid sequences originating from translated DNA sequences. Alignments were trimmed to a uniform length (corresponding to nucleotide positions 105 to 867 for *atpD*, 500 to 1308 for *gyrB*, 90 to 1371 for *fliC* and 95 to 630 for *fljB* of the genome of *S.* Typhimurium_LT2 (Acc.No. NC_003197)). The trimmed alignments were used to construct a concatenated alignment too. Trees based on the neighbour-joining method were built from the individual alignments as well as from the concatenated alignment. Reliability of the tree structure was assessed via bootstrap analysis, using the PHYLIP phylogenetic software package (functions SEQBOOT, DNADIST, NEIGHBOR and CONSENSE) (Felsenstein, 1989). All sequences obtained in this study were deposited with GenBank under accession numbers DQ095313–DQ095611.

3. Results and discussion

3.1. Selection and characteristics of MLST marker genes

In order to stay compatible with the traditional serotyping approach, an ideal MLST scheme would be based on the genetic determinants of the serotype itself. While this is achievable for the flagellin genes (*fliC* and *fljB*) encoding for the H antigen, there is not a single gene which could be applied as marker gene for MLST (it is not possible to design consensus PCR primers to amplify a gene of these clusters from all or nearly all *Salmonella* serotypes). Our strategy was to use selected housekeeping genes, making up for the lack of sequence information directly related to the O antigen, and to use them for MLST in combination with the flagellin genes. This strategy was based on the assumption, that the O antigen, being the result of a complex interaction of a set of genes, was less prone to lateral gene transfer, than the H1 and H2 flagellar antigens — consequently it would be more related to phylogeny.

A preliminary analysis of several housekeeping genes (*atpD*, *gapA*, *gyrA*, *gyrB*, *rpoB*, the 16S rRNA, the 23S rRNA and the tmRNA) indicated that *atpD* and *gyrB* displayed considerable sequence diversity amongst *Salmonellae*, providing the best phyologenetic resolution amongst the genes tested. The combination of these two housekeeping genes conveniently grouped *Salmonella* strains into groups corresponding to recognised species and subspecies (Suppl. Fig. 1). It should be noted, that the most widely used phylogenetic markers, the 16S and 23S rRNA genes displayed hardly any sequence variability within the genus *Salmonella*.

The characteristics of the marker genes are summarised in Table 3. Similar to earlier findings (Sukhnanand et al., 2005), sequence variability did not show positive correlation with the ability to discriminate amongst serovars. Individual marker genes allowed for the discrimination of 16 to 45 allelic types, differentiated by at least one nucleotide difference. The ratio of polymorphic sites in genes employed by previously published MLST schemes was 3.6–6.4% (Torpdahl et al., 2005), 0.6–5.6% (Kotetishvili et al., 2002) and 1.60–5.78% (Sukhnanand et al., 2005). Sequence variation in our scheme ranged from 6.9 to 69.9%, with the *fliC* gene displaying by far the highest variability. It should be noted, that only the 3′ part of the *fljB* gene was included in the current analysis due to technical difficulties associated with the sequencing of the 5′ region. Still, the analysed *fljB* fragment contained sufficient sequence information for successful MLST of *Salmonella* strains.

Previous MLST schemes for *Salmonella* reported isolated difficulties in amplifying the employed genes from all the isolates investigated (Sukhnanand et al., 2005; Kotetishvili

et al., 2002). PCR amplification of the four genes applied in this scheme was straightforward with all isolates tested.

3.2. Comparative sequence analyses

Even though *atpD* and *gyrB* are widely accepted as reliable phylogenetic marker genes (Dauga, 2002; Fukushima et al., 2002; Christensen and Olsen, 1998), the phylogenetic relationships indicated at the serotype level (Fig. 1) were different from those obtained with MLEE analysis (Boyd et al., 1996) (data not shown). Also, *atpD* and *gyrB* based tree topologies differed, several serovars clustering very differently on them (*i.e.* Braenderup, Mbandaka or Montevideo). These indicate that results should be interpreted with care regarding the phylogenetic relationships amongst strains and serovars of *Salmonella enterica* subsp. *enterica*. However, these genes were found to be suitable markers for *Salmonella* serotype level differentiation, providing effective resolution of subspecies *enterica* into clusters of 1–6 serovars (Fig. 1).

Sequence analysis of *fliC* and *fljB*, the genes encoding for the phase 1 and phase 2 flagellar genes, respectively, followed clearly the H1 and H2 classical serotype classification. As suggested by the analysis of *atpD* and *gyrB* tree topologies vs. phase 1 flagellar data (Fig. 1), sequence analyses of flagellar genes did not follow phylogenetic relationships, but reflected clearly and only the phase 1 and phase 2 antigenic formulae (Suppl. Figs. 2–3).

3.3. MLST differentiation of Salmonella serotypes

Based on the *atpD*, *gyrB*, *fliC* and *fljB*, 60 sequence types (ST) could be identified amongst the 74 *Salmonella enterica* subsp. *enterica* strains analysed (Table 4), corresponding to a discriminatory ability of 1 ST for 1.23 isolates. This compares favourably to the discriminatory ability reported before (25ST/66 isolates, 1 ST for 2,64 isolates, Sukhanand et al. 2005; 50ST/110 isolates, 1 ST for 2,2 isolates, Torpdahl et al., 2005). Reported resolution values for PFGE are in the range of 1 PFGE type for 1.36–2.96 isolates (Fakhr et al. 2005; Torpdahl et al., 2005; Kotetishvili et al., 2002), and 1 AFLP type for 1.4 isolates resolution was found for AFLP (Torpdahl et al., 2005).

While none of the individual marker genes allowed for the identification of at least a majority of the serovars analysed, their combination provided a resolution well beyond that level. Strains belonging to the same serovar conveniently branched together on a phylogenetic tree (Fig. 2). Strains of the same serovar from multiple origin were also clearly indicated by the branching pattern of the tree (Bredeney, Derby, Livingstone, Newport, Oranienburg, Typhimurium, Virchow). A comprehensive sequence database is required in order to maximise the full potential of molecular methods for *Salmonella* serotyping. The presented *atpD*, *gyrB*, *fliC* and *fljB* sequences from multiple strains representing 28 *Salmonella* serovars and subspecies, showed substantial intra-serovar phylogenetic variation, indicating a strong subtyping potential. While the isolates selected can be regarded as typical for the European continent, sequence data from other countries, especially from other geographical regions, would most likely further broaden

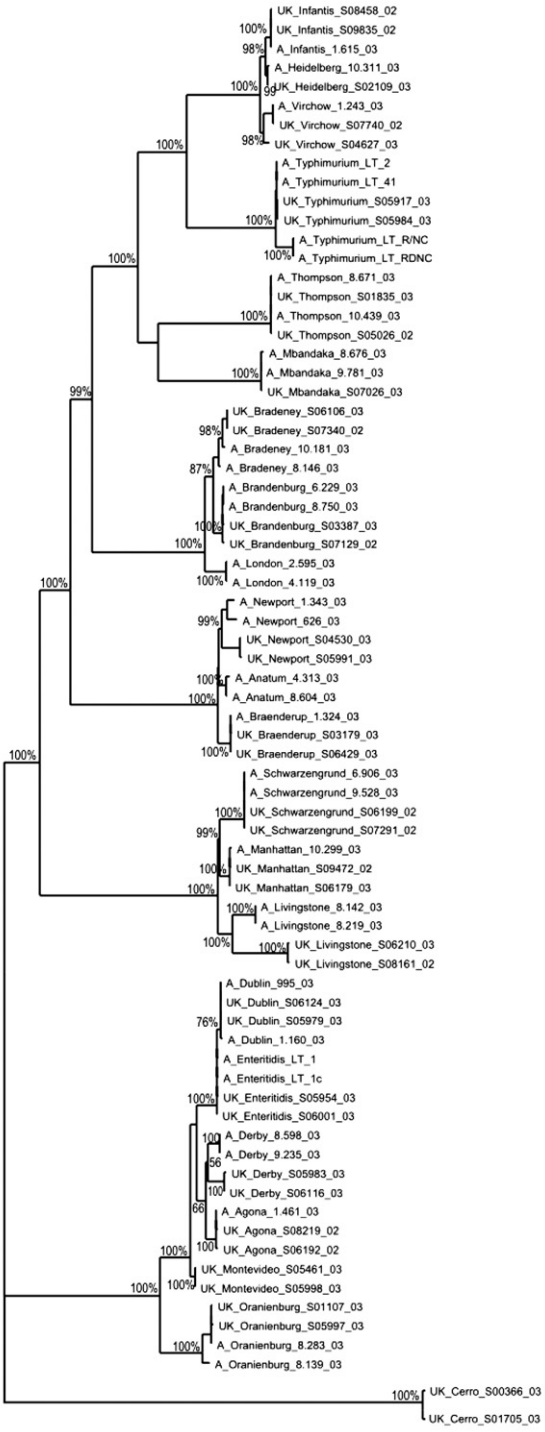

Fig. 2. Neighbor-joining dendrogramm based on concatenated nucleotide sequences of the *atpD*, *gyrB*, *fliC* and *fljB* genes. The positions considered to generate the tree are shown in Table 3. Bootstrap values higher, than 50%, from 100 replicates are shown. The bar represents 1% evolutionary sequence divergence.

the known diversity. Besides being a major step towards molecular serotyping, the establishment of such a comprehensive sequence database would also shed light onto the evolution and large-scale epidemiology of *Salmonella*.

The combination of all four marker genes is required for a reliable serotyping of all the serovars, taking into consideration potential minor sequence variations. A limited sequence

information based on *atpD* and the 5′ region of *fliC* provides, however, sufficient information for the correct diagnosis of most strains investigated so far. Further extension of the database is needed in order to assess the viability of a simplified typing scheme, initially checking only *atpD* and 5′ *fliC* sequence information, for routine serotyping purposes.

3.4. Evidence for gene transfer and recombination of the entire fliC gene (assortative recombination)

Comparative sequence analysis of the phylogenetic marker genes *gyrB*, *atpD* and of the flagellin genes *fliC* and *fljB* was performed. Sequence analysis revealed that 7 of the 22 investigated serovars were of multiple phylogenetic origin (represented by 29 of the 60 strains sequenced), despite having mostly identical or very similar *fliC* sequences (Fig. 1). This indicated the occurrence of an assortative recombination of the *fliC* gene in these serovars (Table 1). Our results support the findings that the same O (polysaccharide) and H (flagellin) antigenic structures may occur in distantly related strains. This is in agreement with previous studies suggesting that horizontal gene transfer and recombination events involving the *fliC*, *fljB* and *rfb* genes are relatively frequent (Beltran et al., 1988; Li et al., 1994). In a previous study *S.* Derby and *S.* Newport were explicitly identified as serovars of multiple phylogenetic origin (Beltran et al., 1988). Our results confirmed these findings, extending this list with another 5 serovars (Bredeney, Livingstone, Oranienburg, Typhimurium, Virchow). Lateral transfer of serotype determining

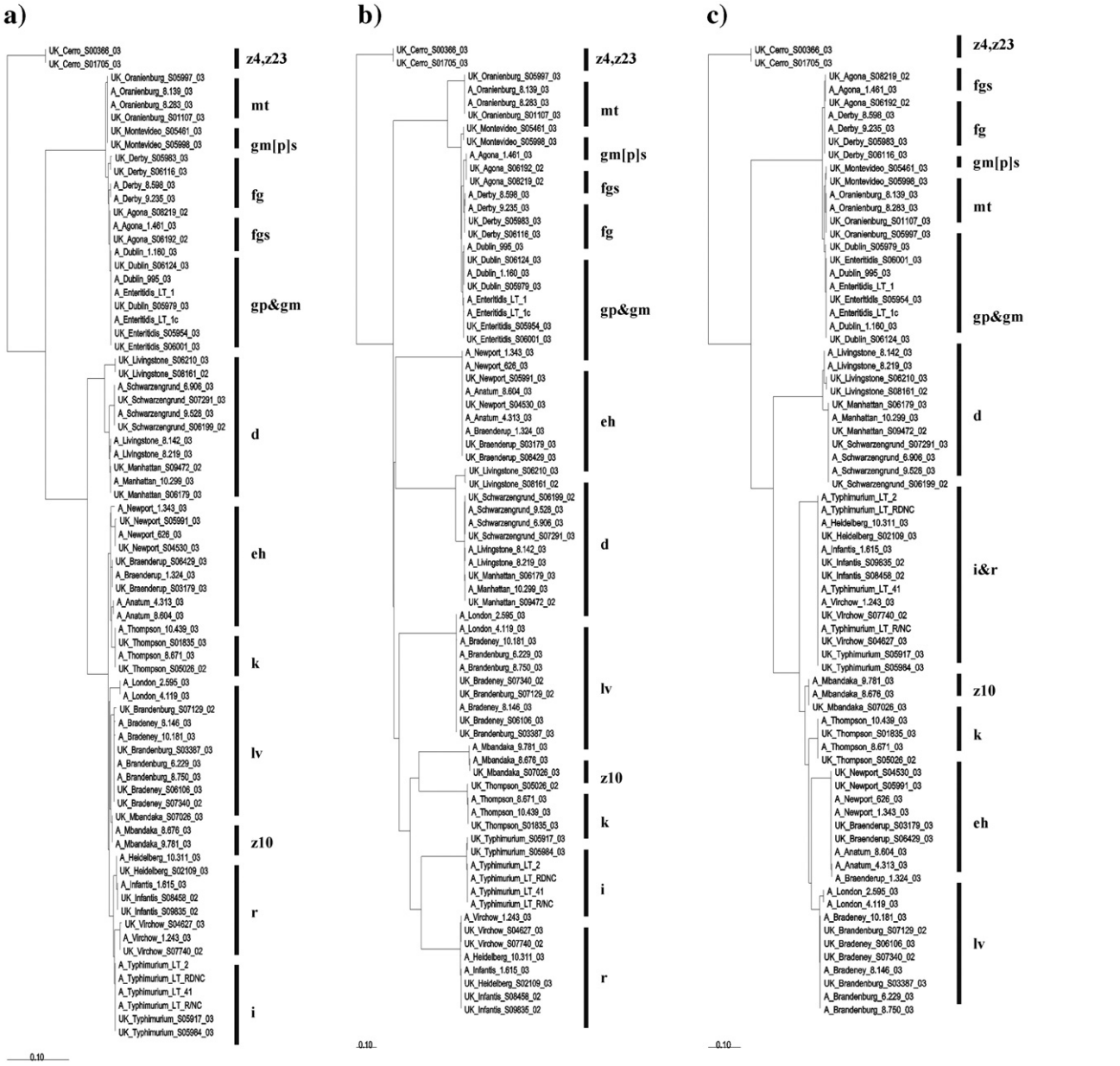

Fig. 3. Neighbor-joining trees based on (a) 5′ end nucleotide sequences of *fliC* gene (C1′: codon 29–181), (b) middle antigen determining region nucleotide sequences of *fliC* gene (V: codon 182–358) and (c) 3′ end nucleotide sequences of *fliC* gene (C2′: codon 359–457). H1 antigen categories are indicated. The bar represents 10% evolutionary sequence divergence.

genes may result in cases where conventional serotyping looses the track. The consideration of housekeeping genes (not prone to lateral gene transfer) in a molecular typing scheme offers a way to overcome this problem.

3.5. Sequence variation in relation to flagellin structure

The *fliC* gene consists of approximately 1600 nucleotides. In order to be used as a marker gene for *Salmonella* serotyping, and considering the high number of different serovars, extensive sequencing efforts would be required. However, identifying the region of *fliC* most suitable for serovar-level *Salmonella* differentiation would be a significant help in this task.

Three regions of *fliC* have been distinguished in previous studies: the C1 (conserved region 1) including codons 1–181; the V (variable antigen determining region) which is the central segment composed of codons 182–370; and the C2 (conserved region 2), consisting of codons 371–505 (Li et al., 1994). For the purposes of the present analysis we used slightly shorter fragments of the C1 (C1′: codons 29–181) and C2 (C2′: codons 371–466) regions, and the entire V region (codons 182–370) (Fig. 3).

The variability of *fliC* sequences among serovars with different phase 1 structure was found to increase in the order C1′<C2′<V. However, the variability of *fliC* sequences among serovars with similar phase 1 structure increases in the reverse order: V<C2′<C1′. For example, serovars Virchow, Heidelberg and Infantis, which have similar phase 1 structures (r), can still be differentiated by using the C1′ region; however in regions V and C2′ the sequences of these three serovars are highly similar and thus a differentiation is not possible. Similarly, serovars Newport, Braenderup and Anatum which have similar phase 1 structures (eh) can still be differentiated by using the C1′ region; however in regions V and C2′ the sequences of these three serovars are similar and thus differentiation is not possible. On the other hand, the sequence of the V region correlates best with the phase I flagellar antigenic structure. The differentiation of Oranienburg (phase 1: mt) from Montevideo (phase 1: gm[p]s) is possible only based on the V region, but not on the C1, while there is only a single nucleotide mismatch in their C2 region (Fig. 3).

Based on this comparative analysis, C1′ is the most suitable region of the *fliC* gene for DNA based differentiation of *Salmonella enterica* subsp. *enterica* serovars within the same phase 1 group, while the V region is the most suitable for the differentiation and identification of the phase 1 structure groups themselves.

4. Conclusions

The combination of housekeeping genes *gyrB*, *atpD* with the flagellin genes *fliC* and *fljB*, showed good discrimination amongst the *Salmonella* serovars investigated. These results indicate, that the approach enables efficient molecular serotyping for at least the more common serovars of *Salmonella*. Various molecular methods, taking advantage of the sequence differences, may be utilised for molecular serotyping, including MLST, pyrosequencing, microarray based hybridisation, serotype specific PCR, etc.

Sequence analysis indicated several instances of lateral gene transfer of the *fliC* gene (assortative recombination), transforming one serovar into another one. When such an event occurs, it results in the "appearance" of a new serovar. The relationship between such a new strain and its immediate predecessor is not recognised by serotyping, which, in isolated cases, may potentially lead to erroneous conclusions regarding *Salmonella* epidemiology. The combination of sequence information from housekeeping genes (phylogenetic markers) and flagellar genes may be a solution, resolving these cases.

Acknowledgements

Work at ARC was supported by the Fonds zur Förderung der wissenschaftlichen Forschung, Austria (Project number L70). Funding from the EU towards COST Action 853 (Agricultural Biomarkers for Array Technology) contributed to the presented results by enhancing exchange of ideas. LB is indebted to Fodor Szilvia for her support and fresh, independent views and ideas. Comments and suggestions from two anonymous reviewers are gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mimet.2006.11.013.

References

Beltran, P., Musser, J.M., Helmuth, R., Farmer III, J.J., Frerichs, W.M., Wachsmuth, I.K., Ferris, K., McWhorter, A.C., Wells, J.G., Cravioto, A., Selander, R.K., 1988. Toward a population genetic analysis of Salmonella: genetic diversity and relationships among strains of serotypes *S. choleraesuis, S. derby, S. dublin, S. enteritidis, S. heidelberg, S. infantis, S. newport*, and *S. typhimurium*. Proc. Natl. Acad. Sci. U. S. A. 85, 7753–7757.

Boyd, E.F., Wang, F.S., Whittam, T.S., Selander, R.K., 1996. Molecular genetic relationships of the *Salmonellae*. Appl. Environ. Microbiol. 62, 804–808.

Brenner, F.W., Villar, R.G., Angulo, F.J., Tauxe, R., Swaminathan, B., 2000. *Salmonella* nomenclature. J. Clin. Microbiol. 38, 2465–2467.

Christensen, H., Olsen, J.E., 1998. Phylogenetic relationships of *Salmonella* based on DNA sequence comparison of *atpD* encoding the beta subunit of ATP synthase. FEMS Microbiol. Lett. 161, 89–96.

Dauga, C., 2002. Evolution of the *gyrB* gene and the molecular phylogeny of *Enterobacteriaceae*: a model molecule for molecular systematic studies. Int. J. Syst. Evol. Microbiol. 52, 531–547.

Dauga, C., Zabrovskaia, A., Grimont, P.A., 1998. Restriction fragment length polymorphism analysis of some flagellin genes of *Salmonella enterica*. J. Clin. Microbiol. 36, 2835–2843.

Echeit, M.A., Herrera, S., Garaizar, J., Usera, M.A., 2002. Multiplex PCR-based detection and identification of the most common *Salmonella* second-phase flagellar antigens. Res. Microbiol. 153, 107–113.

Felsenstein, J., 1989. PHYLIP — Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.

Fukushima, M., Kakinuma, K., Kawaguchi, R., 2002. Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. J. Clin. Microbiol. 40, 2779–2785.

Herrera-Leon, S., McQuiston, J.R., Usera, M.A., Fields, P.I., Garaizar, J., Echeita, M.A., 2004. Multiplex PCR for distinguishing the most common phase-1 flagellar antigens of *Salmonella* spp. J. Clin. Microbiol. 42, 2581–2586.

Kotetishvili, M., Stine, O.C., Kreger, A., Morris Jr., J.G., Sulakvelidze, A., 2002. Multilocus sequence typing for characterization of clinical and environmental salmonella strains. J. Clin. Microbiol. 40, 1626–1635.

Li, J., Nelson, K., McWhorter, A.C., Whittam, T.S., Selander, R.K., 1994. Recombinational basis of serovar diversity in *Salmonella enterica*. Proc. Natl. Acad. Sci. U. S. A. 91, 2552–2556.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., Schleifer, K.H., 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32, 1363–1371.

Macnab, R.M., 1992. Genetics and biogenesis of bacterial flagella. Annu. Rev. Genet. 26, 131–158.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95, 3140–3145.

McQuiston, J.R., Parrenas, R., Ortiz-Rivera, M., Gheesling, L., Brenner, F., Fields, P.I., 2004. Sequencing and comparative analysis of flagellin genes *fliC*, *fljB*, and *flpA* from *Salmonella*. J. Clin. Microbiol. 42, 1923–1932.

Paradis, S., Boissinot, M., Paquette, N., Belanger, S.D., Martel, E.A., Boudreau, D.K., Picard, F.J., Ouellette, M., Roy, P.H., Bergeron, M.G., 2005. Phylogeny of the *Enterobacteriaceae* based on genes encoding elongation factor Tu and F-ATPase beta-subunit. Int. J. Syst. Evol. Microbiol. 55, 2013–2025.

Popoff, M.Y., Bockemuhl, J., Gheesling, L.L., 2003. Supplement 2001 (no. 45) to the Kauffmann–White scheme. Res. Microbiol. 154, 173–174.

Porwollik, S., Boyd, E.F., Choy, C., Cheng, P., Florea, L., Proctor, E., McClelland, M., 2004. Characterization of *Salmonella* enterica subspecies I genovars by use of microarrays. J. Bacteriol. 186, 5883–5898.

Reeves, P.R., Hobbs, M., Valvano, M.A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C.R., Rick, P.D., 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol. 4, 495–503.

Sukhnanand, S., Alcaine, S., Warnick, L.D., Su, W.L., Hof, J., Craver, M.P., McDonough, P., Boor, K.J., Wiedmann, M., 2005. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella enterica* serotypes. J. Clin. Microbiol. 43, 3688–3698.

Torpdahl, M., Skov, M.N., Sandvang, D., Baggesen, D.L., 2005. Genotypic characterization of *Salmonella* by multilocus sequence typing, pulsed-field gel electrophoresis and amplified fragment length polymorphism. J. Microbiol. Methods 63, 173–184.

Wei, L.N., Joys, T.M., 1985. Covalent structure of three phase 1 flagellar filament proteins of *Salmonella*. J. Mol. Biol. 186, 791–803.

Yamamoto, S., Harayama, S., 1995. PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. Appl. Environ. Microbiol. 61, 1104–1109.