GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Matrix-assisted laser desorption ionization-time of flight mass spectrometry analysis of *Escherichia coli* categories

C.B. Dallagassa[1], L.F. Huergo[2], M.I. Stets[2], F.O. Pedrosa[2], E.M. Souza[2], L.M. Cruz[2], F.E.A. Assis[1], S. Wolf[1], W. Volanski[1], G. Picheth[1], C.P. Pigatto-Denardi[3], S.M.S.S. Farah[4] and C.M.T. Fadel-Picheth[1]

[1]Departamento de Patologia Médica, Universidade Federal do Paraná, Curitiba, PR, Brasil
[2]Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba, PR, Brasil
[3]Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Matão, SP, Brasil
[4]Laboratório Central do Estado do Paraná, Curitiba, PR, Brasil

Corresponding author: C.M.T. Fadel-Picheth
E-mail: fpicheth@ufpr.br

Genet. Mol. Res. 13 (1): 716-722 (2014)
Received March 15, 2013
Accepted October 20, 2013
Published January 29, 2014
DOI http://dx.doi.org/10.4238/2014.January.29.2

ABSTRACT. The mass profiles of cell-free extracts of 180 commensal and pathogenic strains of *Escherichia coli* were determined by MALDI-TOF mass spectrometry (MS). While some peaks were highly conserved in all *E. coli*, several peaks occurred only in some strains, showing heterogeneity among them. We did not detect strain-specific peaks for any of the *E. coli* categories tested. However, review of the fully conserved and the variable peaks suggested that MALDI-TOF MS has the potential to distinguish commensal and uropathogenic *E. coli* strains. Additionally, eight *Shigella sonnei* isolates were tested and found to be indistinguishable from *E. coli* by MALDI-TOF MS under the test conditions.

Key words: MALDI-TOF; *Escherichia coli*; Commensal; Pathotypes; Mass fingerprint

INTRODUCTION

Escherichia coli is the most common species of facultative anaerobe found in the human gastrointestinal tract (Donnenberg, 2010). It is a remarkable and diverse species (Croxen and Finlay, 2010), exhibiting a significant degree of intraspecific heterogeneity and substantial genetic and metabolic diversity (Maharjan and Ferenci, 2005; Rasko et al., 2008; Touchon et al., 2009). Although most *E. coli* strains reside harmlessly in the lumen of the human intestine and seem poorly adapted to causing disease in healthy individuals, there are several pathotypes that can cause infections in humans (Donnenberg, 2010). At least eight pathotypes have been described, of which six, called diarrheagenic *E. coli* (DEC), are associated with intestinal infections: enteropathogenic *E. coli* (EPEC, typical and atypical), Shiga toxin-producing *E. coli* (STEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), and diffusely adherent *E. coli* (DAEC). Molecular methods for the detection of the virulence markers as well as immunological and cellular assays are required to distinguish these pathotypes. In addition to these pathotypes, extraintestinal infections are caused by neonatal meningitis *E. coli* (NMEC) and uropathogenic *E. coli* (UPEC), the major cause of urinary infection in humans. Other pathotypes have also been identified, but their mechanisms of pathogenesis are not well defined (Croxen and Finlay, 2010).

Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) is being increasingly applied for bacterial identification. In contrast to classical microbiological techniques based on culture, which may require several days for microbial identification, MALDI-TOF MS relies on mass fingerprinting of bacterial colonies, generates results in a few minutes at low cost, and provides high throughput bacterial identification up to the subspecies level (Donohue et al., 2006; Dieckmann et al., 2008; Seibold et al., 2010; Dieckmann and Malorny, 2011). Here, we studied the ability of MALDI-TOF MS to discriminate *E. coli* groups.

MATERIAL AND METHODS

Strains

A total of 180 strains were analyzed using MALDI-TOF mass spectrometry. These included commensals with no DEC virulence genes; *E. coli* pathotypes isolated from humans and cattle that had been identified by standard biochemical methods and shown to carry virulence genes (Farah et al., 2007; Pigatto et al., 2008; De Toni et al., 2009; Oliveira et al., 2011; Fialho, O.B., and Dallagassa, C.B., unpublished data); and *E. coli* strains ATCC 25922, ATCC 35218, DH10B, IAL 307 (O124:K72, EIEC), IAL 2391 (EAEC), C1845 (DAEC), E2348/69 (EPEC), and EDL 933 (STEC). *Aeromonas caviae* ATCC 14486 and *Salmonella enterica* serotype Braenderup ATCC BAA 664 were used as examples of external groups.

MALDI-TOF MS

Sample preparation

For extract preparation, bacteria were grown on MacConkey agar for 18 h at 36 ± 1°C.

One loopful (~10 colonies, 2 mm) of each bacterial culture was washed twice in sterile water, resuspended in 1 mL 70% ethanol, and centrifuged at 14,000 rpm for 2 min. The pellet was dried at 37°C for 30 min, and the cells were lysed in 100 µL 35% formic acid and 50% acetonitrile. The mixture was centrifuged and the supernatant used for MS analysis (Sauer et al., 2008).

Analysis

One microliter of each bacterial extract was spotted onto a MALDI target plate, air dried, and covered with 1 µL of the matrix (10 mg/mL α-cyano-4-hydroxycinnamic acid in 50% acetonitrile/2.5% trifluoroacetic acid). MALDI-TOF analyses were performed on a Bruker Autoflex II MALDI-TOF spectrometer (Bruker, Bremen, Germany) in linear positive mode with delayed ion extraction (20 kV as accelerating voltage). Spectra were obtained with an average of 1000 laser shots (10 data sets of 100 shots at different spot positions). The mass-to-charge ratio (m/z) window, 3000-20,000, was analyzed. Data were acquired using the AutoXecute tool of the FlexControl software 3.0 (Bruker-Daltonics, Bremen, Germany). The peak resolution was in the range of 500-600. External calibration was performed using insulin (5734.51 m/z), ubiquitin I (8565.76 m/z), and cytochrome *c* (12,360.97 m/z). Raw data were converted into peak lists using FlexAnalysis 3.0 (Bruker-Daltonics). Minimal signal/noise was set at 3, and the peak m/z was defined as the centroid of the peak at 80% height. Cell-free extracts were spotted on three different positions on the MALDI plate, generating three spectra per strain. The peak lists of each strain were analyzed with SPECLUST (Alm et al., 2006), available at http://bioinfo.thep.lu.se/speclust.html to generate a consensus peak list containing only the m/z values detected in at least two of the replicates, using an error window of ± 5 m/z. The consensus lists were then compared to determine common peaks among the strains.

RESULTS AND DISCUSSION

For *E. coli*, the mass peaks ranged from approximately 3000 to 15,000 m/z; those most frequently found are shown in Table 1. The peaks characteristic of *E. coli* (Figure 1, Table 1) are in close agreement with those already described (Lynn et al., 1999; Mazzeo et al., 2006; Liu et al., 2007), confirming that they are detectable under distinct conditions of analysis and can be used as markers for *E. coli* identification. In addition, several potential mass spectra markers for the Enterobacteriaceae family have been described (Lynn et al., 1999), seven of which were found under our experimental conditions. The peaks at 4364, 5380, and 6253 m/z were found in 82-100% of the *E. coli* strains analyzed, as well as in *Salmonella*, but were absent in *Aeromonas*. Peaks at 6314, 7156, 7271, and 9532 m/z were found only in *E. coli* strains (35-100%) (Table 1). Furthermore, the signal at 8897 m/z previously described in the *Escherichia* genus (Mazzeo et al., 2006) probably corresponds to 8892 m/z detected in 15-63% of the strains analyzed in the present work. The data indicate that these masses are frequently found in *E. coli* strains; some of them may represent family markers. As expected, only a few peaks (5069, 5149, 8989, and 10,294 m/z) were common between *Aeromonas* and *E. coli*, and between *Salmonella* and *E. coli* (4364, 4613, 5380, 6253, and 8365 m/z).

Some m/z peaks were highly conserved among the *E. coli* categories, while others were present at variable frequencies (0-100%) (Table 1), showing heterogeneity among the strains.

The potential of the MALDI-TOF MS for *E. coli* strain-level discrimination has been

proposed (Muroi et al., 2011). We did not detect strain-specific peaks for any of the *E. coli* strains tested. However, combining the fully conserved and the variable peaks suggests that there may be minor differences and that MALDI-TOF methodology has the potential to distinguish the commensal and UPEC categories.

Table 1. Frequencies of major mass peaks (m/z) in *Escherichia coli*, *Salmonella enterica* Braenderup and *Aeromonas caviae* strains.

Peak detected (m/z)	N	EPEC (N = 2)[a]	EIEC (N = 2)[a]	DAEC (N = 4)[a]	COM (N = 34)	EAEC (N = 17)	STEC (N = 23)	aEPEC (N = 38)	UPEC (N = 57)	ATCC 25922	ATCC 35218	DH10B	Salm	Aer
3128 (3125-31)	50	50%	0%	0%	3%	6%	9%	8%	68%	+	+	+	-	-
3158 (3156-60)	47	50%	0%	0%	3%	23%	4%	16%	54%	+	+	+	-	-
3637 (3635-40)	96	100%	0%	50%	26%	59%	39%	42%	79%	+	+	+	-	-
3792 (3788-96)	71	100%	50%	50%	62%	12%	43%	21%	40%	+	+	-	-	-
3820 (3814-25)	47	50%	50%	25%	56%	47%	9%	5%	21%	+	-	-	-	-
3849 (3843-54)	24	0%	0%	0%	53%	12%	0%	10%	0%	-	-	-	-	-
4184 (4182-89)	57	50%	50%	25%	12%	18%	9%	18%	63%	+	-	+	-	-
4364 (4360-67)	174	100%	100%	100%	100%	82%	100%	89%	100%	+	+	+	+	-
4532 (4529-35)	71	50%	0%	50%	20%	53%	17%	39%	58%	-	-	-	-	-
4613 (4610-18)	35	0%	50%	0%	0%	0%	4%	10%	44%	+	+	+	+	-
4768 (4765-71)	83	50%	50%	50%	26%	59%	26%	58%	53%	+	-	+	-	-
4776 (4773-80)	67	50%	0%	25%	26%	41%	17%	16%	65%	+	+	-	-	-
4869 (4866-72)	91	50%	50%	0%	18%	76%	61%	68%	51%	-	-	+	-	-
5069 (5065-74)	60	100%	0%	0%	9%	29%	4%	21%	67%	+	-	+	-	+
5095 (5093-99)	178	100%	100%	100%	100%	100%	100%	95%	100%	+	+	+	-	-
5149 (5146-53)	73	50%	50%	25%	15%	41%	9%	29%	72%	+	+	+	-	+
5342 (5338-51)	108	100%	100%	50%	88%	47%	52%	34%	63%	+	+	+	-	-
5380 (5377-84)	181	100%	100%	100%	100%	100%	100%	100%	100%	+	+	+	+	-
5752 (5748-55)	121	100%	50%	75%	100%	65%	48%	53%	65%	-	+	+	-	-
6253 (6250-58)	178	100%	100%	100%	100%	94%	100%	95%	100%	+	+	+	+	-
6314 (6310-18)	169	100%	100%	75%	100%	88%	100%	79%	100%	+	+	+	-	-
6409 (6405-14)	139	50%	50%	75%	65%	76%	74%	84%	82%	+	+	+	-	-
7156 (7153-61)	103	0%	100%	50%	35%	41%	87%	37%	77%	+	-	+	-	-
7271 (7266-77)	180	100%	100%	100%	100%	100%	100%	100%	100%	+	+	+	-	-
8365 (8360-73)	114	100%	100%	25%	35%	53%	65%	39%	95%	+	+	+	+	-
8871 (8866-80)	89	50%	50%	50%	12%	35%	65%	47%	68%	+	+	+	-	-
8892 (8887-8902)	82	50%	0%	50%	15%	29%	61%	45%	63%	+	-	+	-	-
8989 (8984-98)	97	50%	50%	25%	15%	47%	56%	47%	81%	+	+	+	-	+
9060 (9052-68)	179	100%	100%	100%	100%	100%	96%[b]	100%	100%	+	+	+	-	-
9184 (9180-90)	93	50%	50%	50%	3%	53%	61%	76%	63%	+	-	-	-	-
9220 (9214-31)	124	50%	50%	75%	18%	82%	69%	84%	84%	+	+	+	-	-
9532 (9525-40)	176	100%	100%	100%	88%	100%	100%	100%	100%	+	+	+	-	-
9550 (9544-55)	138	100%	50%	25%	41%	76%	87%	74%	98%	+	+	+	-	-
9735 (9729-45)	126	50%	50%	0%	68%	88%	91%	79%	60%	-	-	+	-	-
10131 (10127-38)	42	0%	0%	0%	0%	18%	17%	16%	49%	+	-	-	-	-
10294 (10288-304)	95	0%	50%	25%	18%	41%	52%	40%	86%	+	+	+	-	+

In bold, the m/z values reported for *E. coli* in other studies are indicated (Lynn et al., 1999; Mazzeo et al., 2006; Liu et al., 2007); number in parenthesis indicate the range of variation of the m/z values; + or - indicate presence or absence of peak; N = number of strains showing the mass peak; COM = commensal strains; aEPEC = atypical enteropathogenic *E. coli*; DAEC = diffusely adherent *E. coli*; EAEC = enteroaggregative *E. coli*; EIEC = enteroinvasive *E. coli*; STEC = Shiga toxin-producing *E. coli*; UPEC = uropathogenic *E. coli*; EPEC = typical enteropathogenic *E. coli*; Salm = *Salmonella enterica* serotype Braenderup ATCC BAA 664; Aer = Aeromonas caviae ATCC 14486; N = number of strains belonging to an *E. coli* group; [a]the low sample size does not allow reliable analysis; [b]except EDL 933 (O157:H7).

The UPEC strains (57) contained eight fully conserved peaks (4364, 5095, 5380, 6253, 6314, 7271, 9060, and 9532 m/z), which are also shared with most of the strains. In contrast, the peaks at 8365 and 9550 m/z appeared to be highly conserved only among UPEC strains. In addition, UPEC presented seven other mass peaks with frequencies ranging from 72

to 86% (Table 1), some of which (3637, 5149, 8989, and 10,294 m/z) can help to distinguish this pathotype. Moreover, the presence of a peak at 3128, which is conserved only in this pathotype, appears to be a good indicator of UPEC. In addition, all UPEC strains lacked the peak at 3849 m/z. Together, these characteristics may help distinguish UPEC strains.

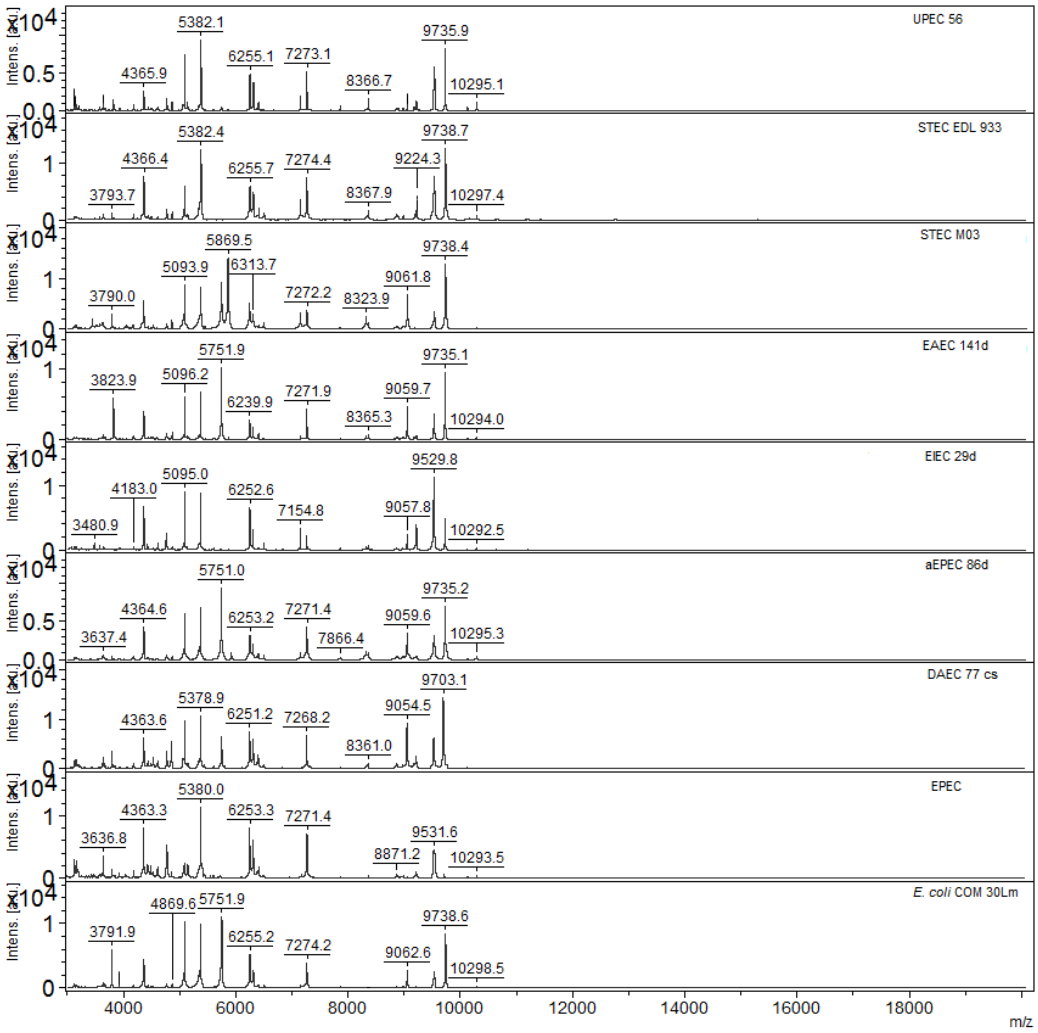

Figure 1. Mass spectrometric profiles from *Escherichia coli* strains. Masses are indicated in Daltons. Intensity is indicated in arbitrary units (a.u), COM = commensal strains; aEPEC = atypical enteropathogenic *E. coli*; DAEC, diffusely adherent *E. coli*; EAEC = enteroaggregative *E. coli*; EIEC = enteroinvasive *E. coli*; STEC = Shiga toxin-producing *E. coli*; UPEC = uropathogenic *E. coli*; EPEC = typical enteropathogenic *E. coli*.

All the commensal strains (34) gave peaks at 4364, 5095, 5380, 5752, 6253, 6314, 7271, and 9060 m/z. Furthermore, the presence of peaks at 3849 and 5342 m/z, which was more common among commensal strains than other *E. coli* categories, and the low frequency

of peaks at 9184, 9220, and 10131 m/z among the commensals, may help to identify these strains. These characteristics suggest that commensal strains of *E. coli* may be distinguishable from pathotypes by using this technology.

Furthermore, the peak at 9060 m/z was present in all the *E. coli* strains analyzed, except in the STEC strain EDL 933, belonging to the O157:H7 serotype. Since all the other STECs analyzed belonged to other serotypes (Farah et al., 2007; Pigatto et al., 2008; De Toni et al., 2009), our data confirm previous reports that only strains of the O157:H7 serotype do not have a peak at 9060 m/z (Mazzeo et al., 2006; Fagerquist et al., 2010).

There were no clear differences between the MALDI-TOF mass profiles of the other *E. coli* categories. Moreover, since only 2 typical EPEC and EIEC and 4 DAEC strains were analyzed, further studies are necessary to characterize these pathotypes.

Since it is generally accepted that EIEC and *Shigella* should form a single pathotype (Croxen and Finlay, 2010), we tested eight *Shigella sonnei* isolates. All of them presented the peaks that were the most conserved among *E. coli* strains, namely those at 4364, 5095, 5380, 6253, 6314, 7271, 9060, and 9532 m/z (Table 1). Additionally, mass peaks at 7156 and 9220 m/z were also fully conserved. We did not identify peaks characteristic of *S. sonnei.* Therefore, the *S. sonnei* isolates were indistinguishable from *E. coli* by MALDI-TOF MS under the test conditions.

To our knowledge, this study is the largest of its kind in terms of the number and range of *E. coli* pathotypes examined. The data show that there is great diversity in the MALDI-TOF mass profiles among *E. coli* strains and suggest that comprehensive studies with larger numbers of strains belonging to the several pathotypes may allow identification of markers for accurate discrimination of *E. coli* categories.

ACKNOWLEDGEMENTS

We thank the Brazilian Program of the Instituto Nacional de Ciência e Tecnologia-INCT/Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq/MCT and the Fundação Araucária for financial support, and CAPES/REUNI for scholarships.

REFERENCES

Alm R, Johansson P, Hjerno K, Emanuelsson C, et al. (2006). Detection and identification of protein isoforms using cluster analysis of MALDI-MS mass spectra. *J. Proteome. Res.* 5: 785-792.

Croxen MA and Finlay BB (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat. Rev. Microbiol.* 8: 26-38.

De Toni F, de Souza EM, Pedrosa FO, Klassen G, et al. (2009). A prospective study on Shiga toxin-producing *Escherichia coli* in children with diarrhea in Parana State, Brazil. *Lett. Appl. Microbiol.* 48: 645-647.

Dieckmann R and Malorny B (2011). Rapid screening of epidemiologically important *Salmonella enterica* subsp. enterica serovars by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* 77: 4136-4146.

Dieckmann R, Helmuth R, Erhard M and Malorny B (2008). Rapid classification and identification of salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* 74: 7767-7778.

Donnenberg MS (2010). Enterobacteriaceae. In: Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Mandell GL, Bennett JE and Dolin R, eds.). 7th edn., Churchill Livingstone, 2815-2833.

Donohue MJ, Smallwood AW, Pfaller S, Rodgers M, et al. (2006). The development of a matrix-assisted laser desorption/

ionization mass spectrometry-based method for the protein fingerprinting and identification of *Aeromonas* species using whole cells. *J. Microbiol. Methods* 65: 380-389.

Fagerquist CK, Garbus BR, Miller WG, Williams KE, et al. (2010). Rapid identification of protein biomarkers of *Escherichia coli* O157:H7 by matrix-assisted laser desorption ionization-time-of-flight-time-of-flight mass spectrometry and top-down proteomics. *Anal Chem.* 82: 2717-2725.

Farah SM, de Souza EM, Pedrosa FO, Irino K, et al. (2007). Phenotypic and genotypic traits of Shiga toxin-producing *Escherichia coli* strains isolated from beef cattle from Parana State, southern Brazil. *Lett. Appl. Microbiol.* 44: 607-612.

Liu H, Du Z, Wang J and Yang R (2007). Universal sample preparation method for characterization of bacteria by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* 73: 1899-1907.

Lynn EC, Chung MC, Tsai WC and Han CC (1999). Identification of *Enterobacteriaceae* bacteria by direct matrix-assisted laser desorptiom/ionization mass spectrometric analysis of whole cells. *Rapid. Commun. Mass Spectrom.* 13: 2022-2027.

Maharjan RP and Ferenci T (2005). Metabolomic diversity in the species *Escherichia coli* and its relationship to genetic population structure. *Metabolomics* 1: 235-242.

Mazzeo MF, Sorrentino A, Gaita M, Cacace G, et al. (2006). Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the discrimination of food-borne microorganisms. *Appl. Environ. Microbiol.* 72: 1180-1189.

Muroi M, Shima K, Nakagawa Y and Tanamoto K (2011). Application of matrix-assisted laser desorption ionization-time of flight mass spectrometry for discrimination of *Escherichia* strains possessing highly conserved ribosomal RNA gene sequences. *Biol. Pharm. Bull.* 34: 430-432.

Oliveira FA, Paludo KS, Arend LN, Farah SM, et al. (2011). Virulence characteristics and antimicrobial susceptibility of uropathogenic *Escherichia coli* strains. *Genet. Mol. Res.* 10: 4114-4125.

Pigatto CP, Schocken-Iturrino RP, Souza EM, Pedrosa FO, et al. (2008). Virulence properties and antimicrobial susceptibility of Shiga toxin-producing *Escherichia coli* strains isolated from healthy cattle from Parana State, Brazil. *Can. J. Microbiol.* 54: 588-593.

Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, et al. (2008). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190: 6881-6893.

Sauer S, Freiwald A, Maier T, Kube M, et al. (2008). Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS One* 3: e2843.

Seibold E, Maier T, Kostrzewa M, Zeman E, et al. (2010). Identification of *Francisella tularensis* by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry: fast, reliable, robust, and cost-effective differentiation on species and subspecies levels. *J. Clin. Microbiol.* 48: 1061-1069.

Touchon M, Hoede C, Tenaillon O, Barbe V, et al. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* 5: e1000344.