Clonal relationships determined by multilocus sequence typing among enteropathogenic *Escherichia coli* isolated in Brazil

André Pitondo-Silva, Luciene A.R. Minarini, Ilana L.B.C. Camargo, and Ana Lúcia C. Darini

Abstract: Enteropathogenic *Escherichia coli* (EPEC) infections are a leading cause of infantile diarrhea in developing nations. Multilocus sequence typing (MLST) characterizes bacterial strains based on the sequences of internal fragments in housekeeping genes. Little is known about strains of EPEC analyzed by MLST from Brazil. In this study, a diverse collection of 29 EPEC strains isolated from patients with diarrhea, admitted to the University Hospital of Ribeirao Preto, was characterized by MLST. Strain analysis demonstrated 22 different sequence types (STs), of which almost half (48%) were new, indicating a high genotype diversity. The 22 STs were divided by eBURST into 12 clonal complexes. It was not possible to correlate typical and atypical EPEC with other strains in the MLST database. This is the first study that analyzed EPEC strains from South America that are included in the *E. coli* MLST database. Nine (31%) out of 29 strains are part of the CC10 clonal complex, the major clonal complex in the database, which comprises 174 strains and 86 different STs, suggesting that these strains might be the most important intestinal pathogenic *E. coli* worldwide. Genetic relationships between typical and atypical EPEC, enterohemorrhagic *E. coli*, and enteroaggregative *E. coli* strains were not established by MLST.

Key words: EPEC, enteropathogenic *Escherichia coli*, MLST.

Résumé : Les infections à *Escherichia coli* entéropathogène (EPEC) sont les principales causes de la diarrhée infantile dans les pays en voie de développement. Le typage de séquences multilocus (*multilocus sequence typing* ou MLST) permet de caractériser les souches bactériennes en fonction des séquences de fragments internes de gènes domestiques. On connaît peu de choses des souches EPEC analysées par MLST au Brésil. Dans cette étude, une collection variée de 29 souches d'EPEC isolées de patients souffrant de diarrhée, admis à l'hôpital universitaire de Ribeirao Preto, a été caractérisée par MLST. L'analyse des souches a révélé la présence de 22 types différents de séquences, parmi lesquels près de la moitié (48 %) étaient nouveaux, indiquant une haute diversité génotypique. Les 22 types de séquences ont été divisés par eBURST en 12 complexes clonaux. Il n'a pas été possible de corréler les EPEC typiques et atypiques avec les autres souches dans la base de données du MLST. Cette étude est la première à analyser les souches d'EPEC d'Amérique du sud comprises dans le MLST de *E. coli*. Neuf (31 %) des 29 souches font partie du complexe clonal CC10, qui est le complexe clonal principal de la base de données qui comprend 174 souches et 86 types de séquences différents, ce qui suggère que ces souches peuvent constituer le groupe d'*E. coli* pathogène intestinal le plus important dans le monde. Les relations génétiques entre les souches EPEC, *E. coli* entéro-hemorragique et *E. coli* entéro-agrégant typiques et atypiques n'ont pas été établies par le MLST.

Mots-clés : EPEC, *Escherichia coli* entéropathogène, MLST.

[Traduit par la Rédaction]

Introduction

Six categories of *Escherichia coli* cause diarrhea: enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), and diffusely adhering *E. coli* (DAEC) (Nataro and Kaper 1998). EPEC is an important category of diarrheagenic *E. coli*, which has

Received 29 October 2008. Accepted 16 December 2008. Published on the NRC Research Press Web site at cjm.nrc.ca on 23 June 2009.

A. Pitondo-Silva. Departamento de Biologia Celular e Molecular e Bioagentes Patogênicos, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (USP), Ribeirão Preto, SP, Brazil; Av. Bandeirantes, 3900, Monte Alegre, CEP: 14049-900; Departamento de Análises Clínicas, Toxicológicas e Bromatológicas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo (USP), Via do Café s/n°, 14040-903, Ribeirão Preto, SP, Brazil.
L.A.R. Minarini, I.L.B.C. Camargo,[1] and A.L.C. Darini.[2] Departamento de Análises Clínicas, Toxicológicas e Bromatológicas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo (USP), Via do Café s/n°, 14040-903, Ribeirão Preto, SP, Brazil.

[1]Present address: Departamento de Física e Informática, Instituto de Física de São Carlos, Universidade de São Paulo (USP), Av. Trabalhador São Carlense, 400, Centro - CEP 13566-590, São Carlos, SP, Brazil.
[2]Corresponding author (e-mail: aldarini@fcfrp.usp.br).

been linked to infant diarrhea in the developing world. In industrialized countries, the frequency of these organisms has decreased, but they continue to be an important cause of diarrhea (Nataro and Kaper 1998). The ability of EPEC to form attaching and effacing intestinal lesions is a major characteristic of its pathogenicity, which is characterized by microvilli destruction, intimate adherence of bacteria to the intestinal epithelium, pedestal formation, and aggregation of polarized actin and other elements of the cytoskeleton at sites of bacterial attachment (Jerse et al. 1990; Knutton et al. 1989). The genetic determinants for the production of attaching and effacing lesions are on the locus of enterocyte effacement, a pathogenicity island located in the bacterial chromosome (Elliott et al. 1998; McDaniel et al. 1995). The locus comprises 41 genes that include a type III secretion system, an adhesin denominated intimin (*eae*), its receptor (*tir*), several secreted proteins (*espA*, *espD*, *espB*, and *espF*), and their chaperones (Jarvis et al. 1995; Jerse et al. 1990; Kenny et al. 1997; Creasey et al. 2003). To date, more than 25 major allelic variants of intimin have been described (Lacher et al. 2006).

In 1996, EPEC strains were classified as typical or atypical (Kaper 1996). Both categories produce a characteristic attaching and effacing histopathology on intestinal cells, and do not produce Shiga, Shiga-like, or verocytotoxins. Typical EPEC has a virulence plasmid EPEC adherence factor (EAF) that encodes localized adherence on cultured epithelial cells mediated by genes for the bundle-forming pilus; atypical EPEC does not posses this plasmid (Kaper 1996; Trabulsi et al. 2002). Characteristically, typical EPEC produces localized adherence with compact microcolonies in HEp-2 cells by means of the bundle-forming pilus, whereas most atypical EPEC form looser microcolonies (Hernandes et al. 2008). Some reports suggest that genetic relationships among atypical EPEC strains are closer to EHEC (Trabulsi et al. 2002) and EAEC strains (Dulguer et al. 2003).

Several molecular genetic methods have been applied to many bacterial strains for species identification, discrimination, and epidemiological studies. PCR, PCR–RFLP (restriction fragment length polymorphism), and PFGE (pulsed-field gel electrophoresis) are the most widely and successfully used. Furthermore, whole-genome sequence data have opened up new insights into epidemiological surveillance (Lukinmaa et al. 2004).

Multilocus sequence typing (MLST) is a relatively new technique in which multiple genes (loci) are sequenced to measure genetic relatedness, and the sequence variation among alleles from many strains is analyzed (Maiden et al. 1998). Wirth et al. (2006) recently described a MLST scheme for *E. coli* (Wirth et al. 2006), along with a public database (available from http://mlst.ucc.ie/mlst/dbs/Ecoli) containing information about more than 2000 strains analyzed so far. Among these, ~163 (8.2%), including the 29 strains in this study, are classified as EPEC. However, the majority of the available database contents refers to *E. coli* strains from Europe, North America, Asia, and Africa, with only a few strains that were isolated in South America.

MLST data are a rich source of inference that can be applied to a broad range of studies of population biology and pathogenicity. A particularly exciting prospect is the integration of MLST with epidemiological information, potentially producing global real-time epidemiological surveillance of bacterial pathogens (Maiden 2006).

Little is known about sequence types (ST) of EPEC strains from Brazil. In this study, a diverse collection of 29 EPEC strains isolated from different patients with diarrhea, admitted to the University Hospital of Ribeirão Preto in Brazil, was characterized by MLST. In addition to describing new alleles and STs, the MLST database was used to determine the genetic relationships between atypical EPEC, EHEC, and EAEC strains.

Materials and methods

Bacterial strains

Twenty-nine *E. coli* strains were selected from a collection of 63 strains, previously studied and identified by phenotypic and genotypic assays as EPEC, which include different PFGE profiles to cover all the different clones of the collection (A. Pitondo-Silva, J.P. Falcão, G. Nakazato, K. Irino, A.F. Pestana de Castro, W.D. Silveira, A.L.C. Darini, J.L. Proença-Módena, R. Martinez, M.P. Sircili, and M. Brocchi, unpublished data). These strains were isolated from different patients with diarrhea from 1999 to 2001. Most of the patients were newborns and children under 5 years old with diarrheic symptoms who were admitted to the University Hospital, Faculty of Medicine of Ribeirão Preto, University of São Paulo, SP, Brazil. All the strains were isolated at the same hospital; however, they are representative of the circulating EPEC strains in an important metropolitan region in Brazil, since they were isolated from patients from different towns located within a radius of 100 km.

Typical and atypical EPEC strains were distinguished by the presence or absence of EAF plasmid and *bfpA* gene, respectively, and HEp-2 adherence patterns (Kaper 1996; Trabulsi et al. 2002; Scaletsky et al. 1996). The strains were stored at –80 °C in Luria–Bertani medium (Oxoid) containing 50% glycerol. Relevant characteristics of all strains used in this study are listed in Table 1.

PFGE

The PFGE of the 29 strains in this study was revaluated and compared with the results of MLST. PFGE was performed using the restriction enzyme *Xba*I (New England Biolabs, Beverly, Massachusetts), according to methods described elsewhere (Falcão et al. 2006). Macrorestriction fragments were resolved by counter-clamped homogeneous electric field electrophoresis in a CHEF-DRIII apparatus (Bio-Rad Laboratories, Hercules, Califfornia), under conditions described elsewhere (Yatsuyanagi et al. 2002). The band patterns were compared by means of the Dice coefficient, using the unweighted pair group method to determine band similarity. The isolate clonalities were judged using criteria described elsewhere (Tenover et al. 1995; Morrison et al. 1999). EPEC strains with the same clonal profile were designated as a single pattern type. A standard molecular-weight ladder (Lambda Ladder (50–1000 kb) PFG Marker; New England Biolabs) was included twice in each gel to compare the fingerprints on different gels. Data analysis was performed with NTSYS-pc (Numerical Taxonomy and Multivariate Analysis System), version 1.7 (Exeter Software,

Table 1. Phenotypic and genotypic data of the 29 enteropathogenic *Escherichia coli* (EPEC) strains studied.

					Allelic profile							
EPEC strain	EPEC classification	PFGE/ PT	CC/S	ST	*adk*	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*	Serotype
HC35	Atypical	PT09	CC10	570*	10	04	04	08	43	08	02	O26:H11
HC66	Atypical	PT10	CC10	571*	112	11	04	01	08	08	02	O111:H–
HC68	Atypical	PT11	CC10	10	10	11	04	08	08	08	02	O55:H–
HC10	Atypical	PT12	CC17	20	06	04	03	18	07	07	06	O128:HNT
HC40	Atypical	PT28	CC17	560*	06	04	03	77	07	07	06	O119:H2
HC41	Atypical	PT15	CC17	561*	06	04	03	18	07	07	23	O25:H1
HC45	Atypical	PT22	CC17	20	06	04	03	18	07	07	06	O119:H2
HC75	Atypical	PT08	CC17	20	06	04	03	18	07	07	06	O119:H2
HC04	Typical	PT05	CC28	551*	13	21	13	22	17	14	25	O86:H–
HC39	Atypical	PT14	CC28	552*	13	21	13	22	17	20	25	O119:H6
HC43	Typical	PT19	CC29	29	06	04	12	16	09	07	07	O26H:11
HC15	Atypical	PT06	CC29	29	06	04	12	16	09	07	07	O26:H–
HC87	Typical	PT18	CC58	572*	56	04	04	16	24	08	14	O114:H40
HC77	Typical	PT24	CC101	101	43	41	15	18	11	07	06	O125:H21
HC16	Atypical	PT13	CC127	127	13	14	19	36	23	11	10	O142:HNT
HC69	Atypical	PT26	CC642	296	09	23	64	77	11	08	06	O126:H7
HC33	Atypical	PT07	CC88	295	06	04	12	01	09	02	07	O125:H9
HC34	Atypical	PT16	CC88	295	06	04	12	01	09	02	07	O125:H9
HC73	Atypical	PT27	CC349	565*	34	140*	39	87	67	16	04	O128:H–
HC95	Typical	PT20	CC349	564*	34	142*	39	87	67	16	04	O153:H2
HC91	Atypical	PT17	CC46	563*	64	07	109*	01	08	08	06	O128:H21
HC17	Atypical	PT04	CC335	595*	29	12	113*	12	15	02	02	O55:H7
HC82	Typical	PT29	S337	337	06	19	15	18	09	08	06	O128:H21
HC06	Typical	PT02	S562	562*	15	14	13	13	17	11	13	O55:H–
HC36	Typical	PT21	S562	562*	15	14	13	13	17	11	13	O125:H9
HC86	Typical	PT03	S562	562*	15	14	13	13	17	11	13	O55:H–
HC63	Typical	PT25	S566	566*	15	15	10	15	88*	11	11	O111:H6
HC09	Typical	PT01	S578	578*	76	11	04	56	90*	01	02	ONT:H49

Note: CC/S, clonal complex/singleton; HNT, nontypable with antisera H1–H56; ONT, nontypable with antisera O1a O1181; PFGE, pulsed-field gel electrophoresis; PT, pattern type; ST, sequence type.

*New alleles and STs.

Setauked, New York). The similarity dendrogram was constructed with the unweighted pair group method with arithmetic means (UPGMA), using the DICE similarity index.

Amplification of housekeeping genes for MLST

Genomic DNA was extracted as described elsewhere (Covone et al. 1998). Amplifications were carried out in a total volume of 50 μL, with 50 ng of template DNA, 20 pmol of primers (Invitrogen, Carlsbad, Califfornia), 5 μL 10× buffer, 2 mmol/L $MgSO_4$ (Invitrogen), 0.2 mmol/L of each deoxynucleoside triphosphate (Eppendorf, Westbury, New York), and 1 U *Taq* DNA polymerase High Fidelity (Invitrogen). PCR reactions were performed in a Mastercycler Gradient (Eppendorf), using the cycling conditions recommended by the MLST Web site (http://mlst.ucc.ie/mlst/dbs/Ecoli/documents/primersColi_html). *Escherichia coli* ATCC 25922 was used as positive control, and reactions without DNA were used as negative control for all PCR assays.

Purification of PCR products and DNA sequencing

PCR products were purified for sequencing with GFX PCR (GE Healthcare, Buckinghamshire, UK).

Automated DNA sequencing was performed, using MegaBACE 1000 DNA sequencers (GE Healthcare), with the same PCR primer sets. Forward and reverse sequence readings from the PCR products were assembled automatically, using Phrap (Ewing et al. 1998). Each forward and reverse strand was sequenced at least 3 times. Raw sequences were reviewed by visual inspection, using ChromasPro version 1.33 software (Technelysium Pty. Ltd).

Multilocus sequence typing

MLST was performed based on the DNA sequences of 7 conserved housekeeping genes — adenylate kinase (*adk*), fumarate hydratase (*fumC*), isocitrate/isopropylmalate dehydrogenase (*icd*), adenylosuccinate dehydrogenase (*purA*), DNA gyrase (*gyrB*), ATP/GTP binding motif (*recA*), and malate dehydrogenase (*mdh*) — that were amplified with specific primers, as described elsewhere (Wirth et al. 2006). The detailed protocol of the MLST procedure, including allelic type and sequence type assignment methods, can be found on the MLST Web site (http://mlst.ucc.ie/mlst/dbs/Ecoli/documents/primersColi_html), hosted at University College Cork, Ireland.

Sequence analysis

Sequences of each allele were trimmed and compared

with all alleles in the database. A unique allele number was assigned to each unique nucleotide sequence. The allelic profile for each strain was determined, and consisted of numbers, one for each allele. A ST was assigned to each isolate, according to its allelic profile. Strains presenting the same allelic profile were considered genetically identical and, thus, received the same ST assignment. Alleles presenting new variations were submitted to the curator of the MLST *E. coli* submission site for assignment of a new allele number and, consequently, a new ST number.

eBURST algorithm

The algorithm eBURST V3 (available from http://eburst.mlst.net/) was used to determine single-locus variation within groups of related strains (clonal complexes) (Feil et al. 2004), to obtain a population snapshot, and to get the most likely patterns of evolutionary descent in our collection. The statistical confidence in the assigned primary founders was determined by a bootstrap resampling procedure (1000 samples). An ST was considered to be a subgroup founder if it had at least 3 single locus variants. Clonal complexes are represented by the abbreviation CC, followed by the number of the clonal complex founder or by the number of the ST present in the highest number of strains inside that clonal complex. For example, CC10 is the clonal complex for which ST10 is the founder. Singletons are represented by the abbreviation S, followed by the corresponding ST number.

Phylogenetic analysis

An MLST dendrogram was constructed from the data matrix of allelic mismatches of the 29 strains in this study with sequence type analysis and recombinational tests (START) (available from http://outbreak.ceid.ox.ac.uk/software.htm), applying UPGMA (Nei 1987). Strains were grouped by BURST, implemented in the START program, when they differed at no more than 2 loci (Nemoy et al. 2005; Jolley et al. 2001).

The discriminatory power of PFGE and MLST dendrograms was compared using Simpson's index of diversity (*D*), as presented by Hunter and Gaston (1988).

Phylogenetic comparison among strains of atypical EPEC with EHEC and EAEC

One ST of each strain classified as EHEC and EAEC in the MLST *E. coli* database was selected. In this case, repeated STs were not considered. A total of 90 different STs for EAEC and 29 for EHEC covered all strains of these categories available on database. The strains were renamed only as EHEC and (or) EAEC, and the number of ST and respective alleles of each strain was documented. Similarly, all strains in this study were renamed as typical EPEC or atypical EPEC. The full data set was used to construct the dendrogram compiled by START, using UPGMA.

Results and discussion

In this study, we analyzed 29 *E. coli* strains from a collection of 63 strains identified as EPEC by phenotypic and genotypic assays, which include different PFGE profiles. We analyzed the PFGE profiles according to the generally accepted criteria proposed by Tenover et al. (1995), although they were established to analyze discrete sets of isolates obtained during epidemiologic studies of potential outbreaks in hospitals or communities spanning relatively short periods (1–3 months). The PFGE profiles were also analyzed by using criteria proposed by Morrison et al. (1999), which consider that single strains may differ from each other in 7 bands in the PFGE typing. The clonality of isolates proposed by both criteria showed the same results. The study of genetic diversity, carried out by PFGE in 29 representative EPEC strains, revealed distinct pattern types (PT1–PT29) (Table 1).

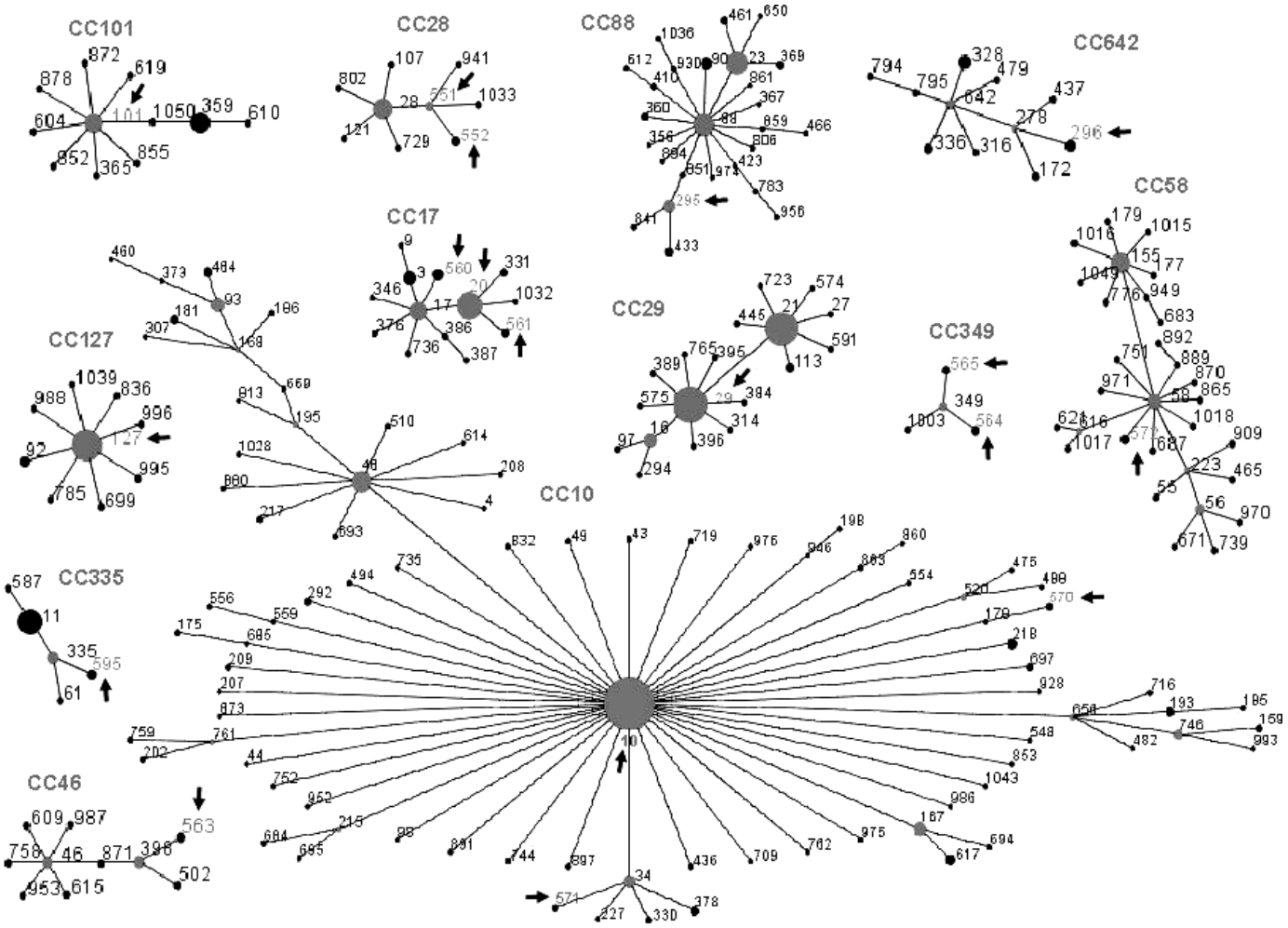
All selected strains were initially identified using the MicroScan Walkaway (DADE BEHRING) automated system. Identification of both O and H antigens (serotypes) was performed with standard tube agglutination tests (Ewing 1986). Only EPEC strains that presented a segment of the *eae* gene amplified by PCR, according to Batchelor et al. (1999), were used in this study. Also, PCR reactions, previously described to detect virulence-related genes and intimin types, were used to determine *bfpA* gene, EAF plasmid, and intimin types (Adu-Bobie et al. 1998; Oswald et al. 2000; Ramachandran et al. 2003). Based on the presence or absence of EAF plasmid, *bfpA* gene, and HEp-2 adherence patterns, 12 strains were classified as typical EPEC and 17 as atypical EPEC (A. Pitondo-Silva, J.P. Falcão, G. Nakazato, K. Irino, A.F. Pestana de Castro, W.D. Silveira, A.L.C. Darini, J.L. Proença-Módena, R. Martinez, M.P. Sircili, and M. Brocchi, unpublished data).

All 29 EPEC strains analyzed by MLST generated sequence tracings acceptable for a ST number assignment and were included in the *E. coli* MLST database. Six new STs were generated by the 6 different alleles found. Furthermore, the combination of previously known alleles generated 8 new STs. The total of 14 new STs and all new alleles were registered in the MLST *E. coli* database (Table 1).

Analysis by MLST demonstrated that 22 of the 29 EPEC strains in this study represent different STs, divided by eBURST into 12 clonal complexes; almost half of the strains (48%) represent new STs. This result is in contrast to a recent one that revealed only 21 clonal types in 129 typical EPEC strains isolated over several years and from many countries (Lacher et al. 2007). Therefore, our study indicated a higher genetic variability in the EPEC strains we analyzed. These discrepancies could be assigned to PFGE patterns that were different than the ones presented in the strains selected for this study, and may also be a result of the fact that Lacher et al. (2007) studied only typical EPEC; our study evaluated both typical and atypical EPEC. In addition, in the report by Lacher et al. (2007), which also used 7 different alleles for MLST, only 2 were shared with the MLST typing method of this study. This may explain some of the differences between the analysis of the EPEC strains in this study and the typical EPEC strains studied in their report.

Seventeen STs corresponded to single strains, whereas 5 STs included 2 and (or) 3 strains. The most representative STs were ST20 and ST562, each comprising ~10% of all strains (3 out of 29 strains). The 22 STs were divided by eBURST into 12 clonal complexes. The major clonal complex in the database was CC10, which comprised 86 differ-

Fig. 1. Application of eBURST algorithm to multilocus sequence typing (MLST) data showing the 12 clonal complexes and the respective sequence types (STs) referring to enteropathogenic *Escherichia coli* (EPEC) strains in this study. The clonal complexes and respective STs found are CC10 (ST10, ST570, ST571), CC17 (ST20, ST560, ST561), CC28 (ST551, ST552), CC29 (ST29), CC46 (ST563), CC58 (ST572), CC88 (ST295), CC101 (ST101), CC127 (ST127), CC335 (ST595), CC349 (ST564, ST565), and CC642 (ST296). Each ST is represented by a dot. STs found in this study are indicated by arrows.



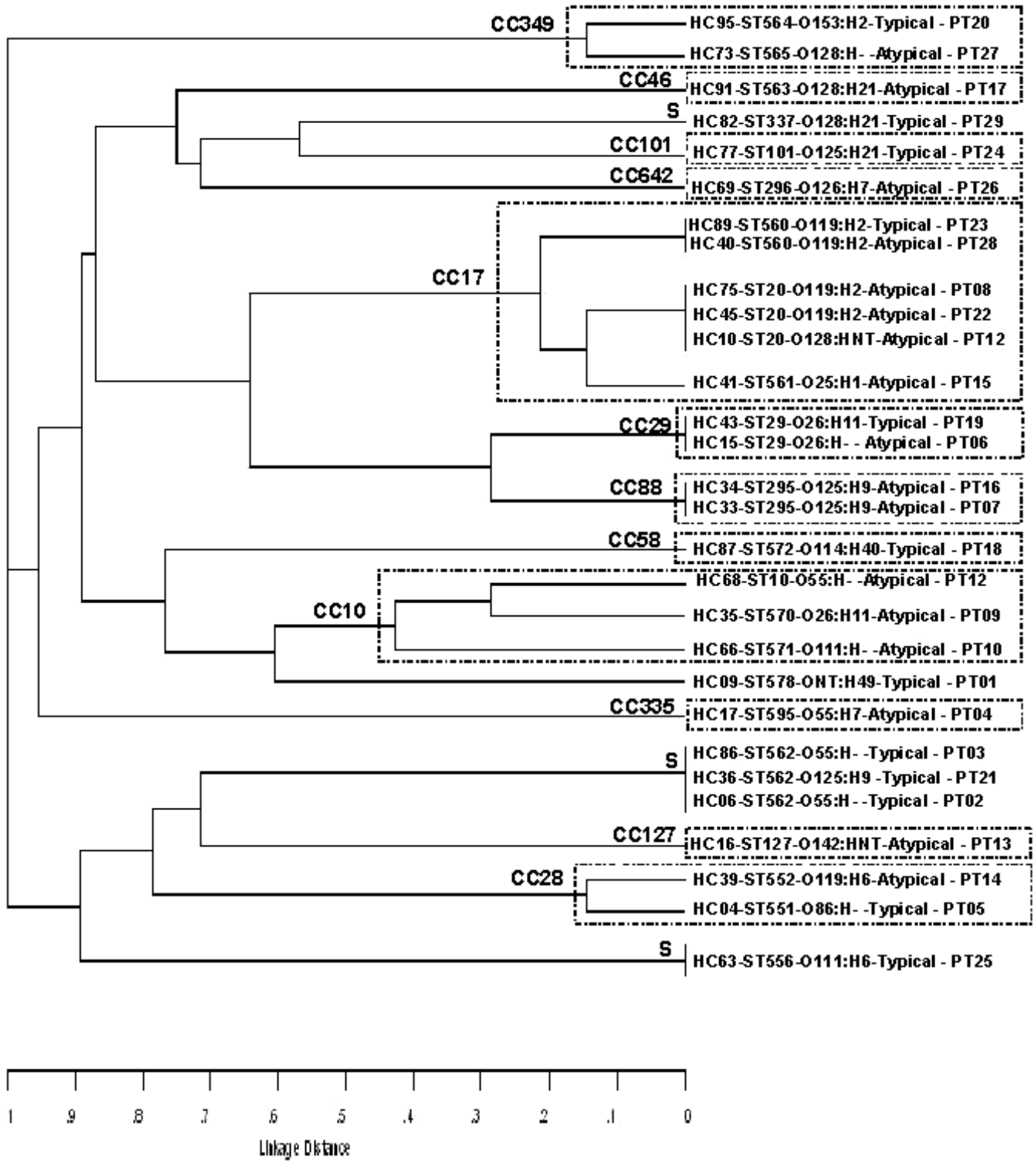
ent STs. The 12 clonal complexes and the respective STs referring to EPEC strains in this study are represented in Fig. 1. The CC10 clonal complex possesses a total of 197 *E. coli* strains, including HC35 (ST570), HC66 (ST571), and HC68 (ST10), the 3 strains in this study (Fig. 1 and Table 1). Four STs described in this study were singletons (S337, S562, S566, and S578). A greater number of strains from this study are comprised by CC17, which harbors 6 strains: HC10, HC45, HC75 (ST20), HC40, HC89 (ST560), and HC41 (ST561) (Fig. 1).

In this study, no correlations among STs and intimin types, geographic source, pathogenicity and pathogen types, serotypes, and simple diseases were found with MLST. Identical serotypes were grouped separately or associated with other serotypes in clonal complexes in which identical STs presented different serotypes (Fig. 2). It was not possible to detect the predominance of any ST among the strains in this study in relation to geographic origin. This is probably due to a lack of data derived from South America, including Brazil. These findings correspond to those of a recent study by Wirth et al. (2006), who concluded that, despite claims that *E. coli* is largely clonal, it is not possible to convincingly deduce the ancestral relationships among the major modern groups with phylogenetic reconstructions. This phenomenon may be due to the fast radiation of the groups after the population bottlenecks, and (or) to frequent recombination within *E. coli* (Wirth et al. 2006). However, future studies should be performed to conclude whether or not the involved regional strains are geographically associated.

The UPGMA dendrogram, based on strain-shared alleles, separated all clonal complexes and singletons presented in the study (Fig. 2). This topology is largely congruent with the eBURST algorithm. Twelve of the studied strains (41%) are typical EPEC, of which 9 presented new STs, whereas 17 strains (59%) are atypical EPEC, of which 8 showed new STs. There was no significant predominance of STs between categories of typical or atypical EPEC (Table 1).

Previous studies, based on the RAPD technique, have suggested that EHEC and EAEC may be genetically related to

Fig. 2. MLST unweighted pair group method with arithmetic means (UPGMA) dendrogram of the 29 EPEC strains are shown and labeled by individual name, sequence type, serotypes, typical or atypical categories, and pulsed-field gel electrophoresis (PFGE) pattern types. Strains with identical STs are shown on the same vertical branch. The clonal complexes (broken lines) were determined with the eBURST program. STs that were presented as isolated (singleton) by the eBURST program are represented by S. The figure shows only STs referring to strains in this study.



atypical EPEC (Trabulsi et al. 2002; Dulguer et al. 2003). When the dendrogram was analyzed for genetic similarity, including all the different STs of EHEC and EAEC categories in the MLST *E. coli* database, it was observed that some categories of atypical EPEC had been grouped with EHEC and EAEC (data not shown). However, since the clusters

were not uniform and showed some exceptions, it is not possible to affirm, at least with MLST data, the existence of genetic similarities between atypical EPEC and EHEC or EAEC. These results are in contrast to studies by Trabulsi et al. (2002) and Dulguer et al. (2003); however, further analyses are necessary to clarify this question. As the housekeeping genes are highly conserved, it is possible that the actual degree of divergence is not enough to allow the analysis of more recent phylogenetic relationships, for example, genetic relations among atypical EPEC and other categories of *E. coli*.

In a similar way, it was not determined by MLST if typical and atypical EPEC form defined genetic groups. In the 29 strains in the study, typical and atypical EPEC significant clusters grouping these 2 categories separately were not found (Fig. 2), suggesting that both categories evolved randomly a number of times from *E. coli* of distinct genetic backgrounds, probably by losing their EAF plasmid from typical EPEC and being linked to the acquisition of pathogenicity islands (locus of enterocyte effacement). These findings corroborate previous studies (Whittam and McGraw 1996).

The discriminatory abilities of PFGE and MLST were compared using the number of unique STs or patterns determined by each method and by Simpson's diversity index (Hunter and Gaston 1988). Using Simpson's *D* value, we determined that the PFGE technique possesses greater discriminatory power than MLST, with *D* values of 0.988 and 0.977, respectively. Although both of these methods have high discriminatory abilities, PFGE distinguished more types (29 vs. 22 different allelic profiles distinguished by MLST). These results were expected, since PFGE covers the bacterial genome as a whole, whereas MLST analyzes the diversity of a restricted number (generally 7) of housekeeping genes. However, because MLST is based on sequence analysis, this technique is more appropriate for evolutionary studies. Although our study showed that PFGE had a higher discriminatory ability, MLST has the advantage of giving information about the clonal relationships of isolates that PFGE does not give.

Neither method grouped the strains according to serotypes, reaffirming that serotyping is still the gold standard typing tool for EPEC strains. These findings also suggest that EPEC strains with an identical phylotype or even clone reflect different serotypes.

Summarizing, 6 new alleles were found in 29 strains of EPEC, and the number of new STs found was high, indicating the existence of considerable genetic variability among the strains. Nine (31%) of the 29 strains are part of the CC10 clonal complex, which is the major clonal complex of the database, comprising 197 strains in 86 different STs, suggesting that these strains might be the most important intestinal pathogenic *E. coli* worldwide. It was not possible to make a comprehensive analysis relating strains in this study to strains from Brazil, previously inserted in the database, because few were EPEC. In the same way, genetic similarity between atypical EPEC and EHEC or EAEC was not observed with MLST data. Therefore, our study is a pioneer in making available a larger number of typical and atypical EPEC strains from Brazil to the MLST *E. coli* database, which will contribute to further studies.

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant No. 06/02320-9) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant No. 680209/2000-0). The authors are grateful to M. Brocchi for generous and helpful discussions and suggestions in this work. We also thank W.A. Silva Jr., coordinator of the Laboratory of Molecular Genetics and Bioinformatics Department of Genetics, Faculty of Medicine of Ribeirao Preto, University of São Paulo, for providing part of the sequencing. Also, we are grateful to I.C.V. Palazzo for critical reading of the manuscript.

References

Adu-Bobie, J., Frankel, G., Bain, C., Gonçalves, A.G., Trabulsi, L.R., Douce, G., et al. 1998. Detection of intimins α, β, γ, and δ, four intimin derivates expressed by attaching and efacing microbial pathogens. J. Clin. Microbiol. **36**: 662–668. PMID: 9508292.

Batchelor, M., Knutton, S., Capriolli, A., Hutler, V., Zanial, M., Dougan, G., and Frankel, G. 1999. Development of a universal intimin antiserum and PCR primers. J. Clin. Microbiol. **37**: 3822–3827. PMID:10565891.

Covone, M.G., Brocchi, M., Palla, E., Silveira, W.D., Rappuoli, R., and Galeotti, C.L. 1998. Levels of expression and immunogenicity of attenuated *Salmonella anterica* serovar Typhimurium strains expressing *Escherichia coli* mutant heat-labile enterotoxin. Infect. Immun. **66**: 224–231. PMID:9423862.

Creasey, E.A., Delahay, R.M., Bishop, A.A., Shaw, R.K., Kenny, B., Knutton, S., and Frankel, G. 2003. CesT is a bivalent enteropathogenic *Escherichia coli* chaperone required for translocation of both Tir and Map. Mol. Microbiol. **47**: 209–221. doi:10.1046/j.1365-2958.2003.03290.x. PMID:12492865.

Dulguer, M.V., Fabbricotti, S.H., Bando, S.Y., Moreira-Filho, C.A., Fagundes-Neto, U., and Scaletsky, I.C. 2003. Atypical enteropathogenic *Escherichia coli* strains: phenotypic and genetic profiling reveals a strong association among enteroaggregative *E. coli* heat-stable enterotoxin and diarrhea. J. Infect. Dis. **188**: 1685–1694. doi:10.1086/379666. PMID:14639540.

Elliott, S.J., Wainwright, L.A., McDaniel, T.K., Jarvis, K.G., Deng, Y.K., Lai, L.C., et al. 1998. The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. Mol. Microbiol. **28**: 1–4. PMID:9593291.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8**: 175–185. doi:10.1101/gr.8.3.175. PMID: 9521921.

Ewing, E.H. 1986. Genus *Escherichia*. *In* Edwards and Ewing's identification of *Enterobacteriaceae*. 4th ed. *Edited by* W.H. Ewing. Elsevier Science, New York. pp.96–134.

Falcão, J.P., Falcao, D.P., Pitondo-Silva, A., Malaspina, A.C., and Brocchi, M. 2006. Molecular typing and virulence markers of *Yersinia enterocolitica* strains from human, animal, and food origins isolated between 1968 and 2000 in Brazil. J. Med. Microbiol. **55**: 1539–1548. doi:10.1099/jmm.0.46733-0. PMID:17030914.

Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., and Spratt, B.G. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186**: 1518–1530. doi:10.1128/JB.186.5.1518-1530.2004. PMID:14973027.

Hernandes, R.T., Silva, R.M., Carneiro, S.M., Salvador, F.A., Fernandes, M.C., Padovan, A.C., et al. 2008. The localized adherence pattern of an atypical enteropathogenic *Escherichia coli*

is mediated by intimin omicron and unexpectedly promotes HeLa cell invasion. Cell. Microbiol. **10**: 415–425. doi:10.1111/j.1462-5822.2007.01054.x. PMID:17910741.

Hunter, P.R., and Gaston, M.A. 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26**: 2465–2466. PMID:3069867.

Jarvis, K.G., Girón, J.A., Jerse, A.E., McDaniel, T.K., Donnenberg, M.S., and Kaper, J.B. 1995. Enteropathogenic *Escherichia coli* contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation. Proc. Natl. Acad. Sci. U.S.A. **92**: 7996–8000. PMID: 7644527.

Jerse, A.E., Yu, J., Tall, B.D., and Kaper, J.B. 1990. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. Proc. Natl. Acad. Sci. U.S.A. **87**: 7839–7843. PMID:2172966.

Jolley, K.A., Feil, E.J., Chan, M.S., and Maiden, M.C. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics, **17**: 1230–1231. PMID:11751234.

Kaper, J.B. 1996. Defining EPEC. Rev. Microbiol. **27**: 130–133.

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D.J., Frey, E.A., and Finlay, B.B. 1997. Enteropathogenic *Escherichia coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. Cell, **91**: 511–520. doi:10.1016/S0092-8674(00)80437-7. PMID:9390560.

Knutton, S., Baldwin, T., Williams, P.H., and McNeish, A.S. 1989. Actin accumulation at sites of bacterial adhesion to tissue culture cells: basis of a new diagnostic test for enteropathogenic and enterohemorrhagic *Escherichia coli.* Infect. Immun. **57**: 1290–1298. PMID:2647635.

Lacher, D.W., Steinsland, H., Blank, T.E., Donnenberg, M.S., and Whittam, T.S. 2007. Molecular evolution of typical enteropathogenic Escherichia coli: clonal analysis by multilocus sequence typing and virulence gene allelic profiling. J. Bacteriol. **189**: 342–350. doi:10.1128/JB.01472-06. PMID:17098897.

Lacher, D.W., Steinsland, H., and Whittam, T.S. 2006. Allelic subtyping of the intimin locus (*eae*) of pathogenic *Escherichia coli* by fluorescent RFLP. FEMS Microbiol. Lett. **261**: 80–87. doi:10.1111/j.1574-6968.2006.00328.x. PMID:16842363.

Lukinmaa, S., Nakari, U.M., Eklund, M., and Siitonen, A. 2004. Application of molecular genetic methods in diagnostics and epidemiology of food-borne bacterial pathogens. APMIS, **112**: 908–929. doi:10.1111/j.1600-0463.2004.apm11211-1213.x. PMID: 15638843.

Maiden, M.C. 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. **60**: 561–588. doi:10.1146/annurev.micro.59.030804.121325. PMID:16774461.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., et al. 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. PMID:9501229.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S., and Kaper, J.B. 1995. A genetic locus of enterocyte conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. U.S.A. **92**: 1664–1668. PMID:7878036.

Morrison, D., Woodford, N., Barrett, S.P., Sisson, P., and Cookson, B.D. 1999. DNA banding pattern polimorphism in vancomycin-resistant Enterococcus faecium and criteria for defining strains. J. Clin. Microbiol. **37**: 1084–1091. PMID:10074530.

Nataro, J.P., and Kaper, J.B. 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11**: 142–201. PMID:9457432.

Nei, M. 1987. Phylogenetic trees. *In* Molecular evolutionary genetics. *Edited by* M. Nei. Columbia. University Press, New York. pp. 287–326.

Nemoy, L.L., Kotetishvili, M., Tigno, J., Keefer-Norris, A., Harris, A.D., Perencevich, E.N., et al. 2005. Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum β-lactamase-producing *Escherichia coli* isolates. J. Clin. Microbiol. **43**: 1776–1781. doi:10.1128/JCM.43.4.1776-1781.2005. PMID:15814998.

Oswald, E., Schmidt, H., Morabito, S., Karch, H., Marches, O., and Caprioli, A. 2000. Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli*: characterization of a new intimin variant. Infect. Immun. **68**: 64–71. PMID:10603369.

Ramachandran, V., Brett, K., Hornitzky, M.A., Dowton, M., Bettelheim, K.A., Walker, M.J., and Djordjevic, S.P. 2003. Distribution of intimin subtypes among *Escherichia coli* isolates from ruminant and human sources. J. Clin. Microbiol. **41**: 5022–5032. doi:10.1128/JCM.41.11.5022-5032.2003. PMID:14605134.

Scaletsky, I.C.A., Pelayo, J.S., Girard, R., Rodrigues, J., Pedroso, M.Z., and Trabulsi, L.R. 1996. EPEC adherence to HEp-2 cells. Rev. Microbiol. **27**: 58–62.

Tenover, F.C., Arbeit, R.D., Goering, R.V., Mickelsen, P.A., Murray, B.E., Persing, D.H., and Swaminathan, B. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33**: 2233–2239. PMID:7494007.

Trabulsi, L.R., Keller, R., and Gomes, T.A.T. 2002. Typical and atypical enteropathogenic *Escherichia coli*. Emerg. Infect. Dis. **8**: 508–513. PMID:11996687.

Whittam, T.S., and McGraw, E.A. 1996. Clonal analysis of EPEC serogroups. Rev. Microbiol. **27**: 7–16.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., et al. 2006. Sex and virulence in *E. coli:* an evolutionary perspective. Mol. Microbiol. **60**: 1136–1151. doi:10.1111/j.1365-2958.2006.05172.x. PMID:16689791.

Yatsuyanagi, J., Saito, S., Sato, H., Miyajima, Y., Amano, K., and Enomoto, K. 2002. Characterization of enteropathogenic and enteroaggregative *Escherichia coli* isolated from diarrheal outbreaks. J. Clin. Microbiol. **40**: 294–297. doi:10.1128/JCM.40.1.294-297.2002. PMID:11773137.