

Phylogenetic group distribution among *Escherichia coli* isolated from rivers in São Paulo State, Brazil

Renato H. Orsi · Nancy C. Stoppe ·
Maria Inês Z. Sato · Paulo Inácio Prado ·
Laura M. M. Ottoboni

Received: 31 July 2007 / Accepted: 17 December 2007 / Published online: 6 January 2008
© Springer Science+Business Media B.V. 2008

Abstract The phylogenetic group distribution of *Escherichia coli* strains isolated from the Sorocaba and Jaguari Rivers located in the State of São Paulo, Brazil, is described. *E. coli* strains from group D were found in both rivers while one strain from group B2 was isolated from the Sorocaba river. These two groups often include strains that can cause extraintestinal diseases. Most of the strains analyzed were allocated into the phylogenetic groups A and B1, supporting the hypothesis that strains from these phylogenetic groups are more abundant in tropical areas. Though both rivers are located in urbanized and industrialized areas where the main source of water pollution is considered to derive from domestic sewage, our results suggest that the major sources of contamination in the sampling sites of both rivers might have originated from animals and not humans.

Keywords Brazil · *Escherichia coli* ·
Phylogenetic groups · River · Water

R. H. Orsi · L. M. M. Ottoboni (✉)
Centro de Biologia Molecular e Engenharia Genética,
Universidade Estadual de Campinas, CP 6010, 13083-875
Campinas, SP, Brazil
e-mail: ottoboni@unicamp.br

N. C. Stoppe · M. I. Z. Sato
Companhia de Tecnologia de Saneamento Ambiental, Av. Prof.
Frederico Hermann Jr. 345, 05489-900 São Paulo, SP, Brazil

P. I. Prado
Departamento de Ecologia Geral, Instituto de Biociências,
Universidade de São Paulo, Rua do Matão, travessa 14, 321,
Cidade Universitária, 05508-900 São Paulo, SP, Brazil

Introduction

Escherichia coli strains can be separated into four main phylogenetic groups: A, B1, B2 and D (Selander et al. 1986; Herzer et al. 1990). Groups A and B1 often include commensal strains (Johnson et al. 2001) and group B2, and to a lesser extent group D, usually allocate extraintestinal pathogenic strains (Picard et al. 1999; Johnson and Stell 2000). Among the *E. coli* pathotypes responsible for extraintestinal infections are UPEC (uropathogenic *E. coli*), EHEC (enterohaemorrhagic *E. coli*) and MNEC (meningitis-associated *E. coli*) (Kaper et al. 2004). *E. coli* from these pathotypes can cause haemolytic uremic syndrome, urinary tract infection, newborn meningitis, sepsis, and others (Dobrindt et al. 2003; Kaper et al. 2004). The intestinal pathogenic *E. coli* strains belong to the pathotypes: ETEC (enterotoxigenic *E. coli*), EPEC (enteropathogenic *E. coli*), EIEC (enteroinvasive *E. coli*), EHEC (enterohaemorrhagic *E. coli*), EAEC (enteroaggregative *E. coli*) and DAEC (diffusely adherent *E. coli*). These pathotypes have been associated with cases of mild and severe diarrhea in adults and children, mostly in developing countries (Kaper et al. 2004). The intestinal pathogenic strains are usually assigned to groups A, B1 and D (Pupo et al. 1997).

Clermont et al. (2000) described a simple PCR-based method that uses a combination of the *chuA* and *yjaA* genes and the DNA fragment TSPE4.C2 to assign *E. coli* strains to the phylogenetic groups A, B1, B2 and D. This methodology has been used, with different purposes, by authors interested in assigning *E. coli* strains into the phylogenetic groups. In this way, Gordon and Cowling (2003) reported, after analyzing non-domesticated vertebrates in Australia, that climate, host diet and body mass can influence the distribution of *E. coli* into the phylogenetic groups A, B1,

B2 and D, in mammals. Dixit et al. (2004) observed that *E. coli* strains isolated from different regions of the gut of pigs belonged to the phylogenetic groups A and B1. Nowrouzian et al. (2005) isolated *E. coli* strains from the commensal intestinal flora of 70 Swedish infants and suggested that strains from the phylogenetic group B2 have evolved to survive in the human intestine.

The contamination of surface water by fecal pollution is a serious problem since it represents a risk to both animal and human health. Fecal pollution can be introduced from multiple sources. Surface runoff and field drainage water from fields containing grazing animals, slurry spreading, farmyard runoff, direct fecal inputs and others can contribute to riverine fecal coliform loads (Vinten et al. 2004). Hence, surface waters are constantly monitored by the competent agencies such as the organization responsible for the control of environmental pollution, sewage and water quality in the State of São Paulo (CETESB), Brazil.

The aim of this work was to allocate *E. coli* strains isolated from the Jaguari and Sorocaba Rivers into the phylogenetic groups A, B1, B2 and D as well as to find the host source and compare their relative abundance, in each phylogenetic group, among the two rivers and with samples from other areas of the world that have been previously published.

Materials and methods

Escherichia coli strains

One hundred and twenty eight strains of *E. coli* were isolated by CETESB from water samples of the rivers Jaguari (60) and Sorocaba (68). The number of strains isolated from each river between January and November are shown in Table 1.

Phylogenetic groups

Phylogenetic group determination was accomplished as described by Clermont et al. (2000). PCR amplifications

were carried out using bacterial lysates, to identify the *chuA* and *yjaA* genes and the DNA fragment TSPE4.C2. The amplification products were separated in a 2% agarose gel containing ethidium bromide. After electrophoresis, the gel was photographed under U.V. light and strains were assigned to phylogenetic group B2 (*chuA*+, *yjaA*+, D (*chuA*+, *yjaA*-), B1 (*chuA*-, TSPE4.C2+) or A (*chuA*-, TSPE4.C2-).

Statistical analysis

The differences in the frequencies of each phylogenetic group among the rivers and periods of the year were tested through log-linear models (Everitt 1977; Fienberg 1978) using the function “loglm” of the package MASS (Venables and Ripley 2002). The frequencies of the phylogenetic groups in our samples and those found for isolates from different human populations (Escobar-Parámo et al. 2004; Nowrouzian et al. 2005) were also compared by using Correspondence Analysis (CA, implemented in the package “vegan” from Oksanen et al. 2005). CA calculates sets of scores (the ordination axes) that order samples and sampled taxonomic entities reciprocally (Gauch 1982; ter Braak 1995). Samples with similar taxa will have close scores at each axis, as well as taxons that occurred in the same samples. Hence, structured populations can be identified as clusters of samples and their associated phylogenetic groups that share similar standard scores. The frequencies of the phylogenetic groups among clusters identified through CA were tested using a Chi-square test, but with P-values estimated from Monte Carlo randomizations, and not from Chi-square Probability Distribution (hence degrees of freedom were not reported). The frequency of strains with *chuA* among these clusters was also tested in the same way. All the statistical calculations were done under the R environment version 2.1.0 for LINUX (R Core Team 2005).

Results and discussion

The Jaguari and Sorocaba Rivers are part of the São Paulo State water bodies monitoring program. This program evaluates the water quality using two indexes, one for water supply and another for aquatic life protection. These rivers are located in urbanized and industrialized areas and the pressure for hydro resources is strong. The levels of chemical and biological parameters indicate that the main source of pollution in these rivers derives from domestic sewage. Both rivers also receive discharge of treated industrial sewage (CETESB 2004).

In this work, sixty *E. coli* strains isolated from the Jaguari River and 68 strains isolated from the Sorocaba

Table 1 *Escherichia coli* strains isolated from rivers Jaguari and Sorocaba and their distribution into phylogenetic groups A, B1, B2 and D

Months	Jaguari				Sorocaba				Total
	A	B1	B2	D	A	B1	B2	D	
Jan–Mar	–	–	–	–	20	4	1	2	27
May–Jul	15	3	0	1	8	6	0	6	39
Sep–Nov	27	10	0	5	17	4	0	0	63
Total	42	13	0	6	45	14	1	8	129

River were allocated into four phylogenetic groups (i.e. A, B1, B2 and D) according to the methodology described by Clermont et al. (2000). Among the strains isolated from the Jaguari River, 42 (70%) were allocated into phylogenetic group A, 13 (22%) into B1 and five (8%) into D (Table 1). Strains isolated from the Sorocaba River were allocated into group A (45 strains, 66%), group B1 (14 strains, 21%), group D (8 strains, 12%) and B2 (1 strain, 1%) (Table 1).

The presence of strains from group D in both rivers and from group B2 in the Sorocaba River deserves attention since the strains from these groups are usually pathogenic. The strains from group B2 are usually responsible for extraintestinal infections and exhibit several virulence factors such as adhesins and toxins (Picard et al. 1999; Johnson and Stell 2000). These strains can cause meningitis, intra-abdominal infections and pneumonia (Russo and Johnson 2003). The phylogenetic group D includes pathogenic strains such as O157:H7, which is highly virulent and can cause diarrhea, hemolytic uremic syndrome and hemorrhagic colitis (Parry and Palmer 2000).

The presence of pathogenic strains in water was already mentioned by others. Müller et al. (2001) found genes for the virulence factors Stx1, Stx2 and enterohaemolysin among *E. coli* strains isolated from water samples in South Africa. Ohno et al. (1997) reported a high bacterial contamination, which included ETEC, EPEC and EIEC strains, in the La Paz River in Bolivia. In Quenia, Simiyu et al. (1998) reported that 22.5% of the strains isolated from the River Nairobi harbored the heat-stable toxin and 17.5% harbored the heat-labile toxin, both produced by ETEC.

Besides the detection of pathogenic *E. coli* in water, several authors have been investigating methods to differentiate the origin (human or animal) of the strains (Turner et al. 1997; Parveen et al. 1999; Dombek et al. 2000; Carson et al. 2003). Goulet and Picard (1986) reported different percentages of strains from group B2 among *E. coli* isolates from humans and animals. These authors observed that only 1.6% of the strains isolated from animals belong to group B2. Among the strains isolated from humans, 9% belong to this group. The percentage of B2 strains isolated from the Sorocaba River (1.47%) is very close to the one described by Goulet and Picard (1986) for fecal isolates from animals, and did not differ statistically from it ($\chi^2 = 0.01$, $P = 1.0$). Despite no strains from group B2 were found in the Jaguari River, this is not statistically different from the expected 1.6% ($\chi^2 = 0.99$, $P = 0.63$). For both rivers, the frequency of B2 strains was statistically different from 9% (Sorocaba River: $\chi^2 = 4.80$, $P = 0.03$; Jaguari River: $\chi^2 = 6.03$, $P = 0.02$). Based on the results, we can speculate that the major contamination sources in the sample collection sites of these rivers originated from animals and not humans. This is in agreement with the fact that the sample collection site in the Jaguari

River is located in a pig feedlots area and the one from the Sorocaba River is located near a cattle slaughtering facility.

Log-linear models showed that the frequency of strains in each phylogenetic group did not differ among the rivers ($\chi^2 = 1.43$, 3 D.F., $P = 0.70$) and among sampling periods ($\chi^2 = 6.84$, 6 D.F., $P = 0.34$). Also, no significant differences were found when the groups were aggregated according to the presence of *chuA* (A + B1 and B2 + D, among rivers $\chi^2 = 0.36$, 1 D.F., $P = 0.55$; among periods $\chi^2 = 6.03$, 2 D.F., $P = 0.32$). In February–March only the Sorocaba River was sampled (Table 1), introducing in this way three structural zeros in the models. However, the results mentioned above held even when this period was excluded.

Escobar-Páramo et al. (2004) showed that isolates from the phylogenetic groups A and B1 were prevalent in the tropical populations analyzed by them. This pattern was confirmed by the Correspondence Analysis of the dataset of these authors, to which we have added the data obtained from a sample of infant intestinal isolates in Sweden (Nowrouzian et al. 2005), and the strains isolated from the Sorocaba and Jaguari Rivers. The first CA axis separated the populations that presented a prevalence of strains from groups A and B1 from those where the prevalent strains belonged to groups B2 and D (Fig. 1). The former cluster included all populations with 50% or more of prevalence of group A strains, namely the three samples from tropical regions, Bogota in Colombia, Cotonou in Benin and Amerindians from French Guiana, analyzed by Escobar-Páramo et al. (2004), and the samples from the rivers Jaguari and Sorocaba (Fig. 1). Populations from the Northern hemisphere were clustered at the opposite side of the axis, and all had less than 35% of prevalence of strains from group A, and at least 19.3% of prevalence of group D strains. The only exception for this clear-cut pattern was a sample obtained from pig farmers from Brittany (Escobar-Páramo et al. 2004) that exhibited the highest prevalence of strains from group A and B1 (32 and 28%, respectively) among the Northern populations, which resulted in an intermediary score (Fig. 1).

The first CA axis explained 81% of the total inertia in the data, which means that the main pattern of the variation among populations is the lower prevalence of *chuA* in the tropical areas. Additional samples from selected latitudes can assert the validity of this apparent geographical gradient. The second CA axis explained more 11% of the total inertia, and separated some populations that exhibited a higher prevalence of strains from group B1, but this pattern showed no correlation with the regions (Fig. 1).

The Northern populations were more clustered in the CA ordination space. In fact, the frequencies of strains in each phylogenetic group did not differ if the Tokyo population was excluded ($\chi^2 = 23.0$, $P = 0.18$). In contrast,

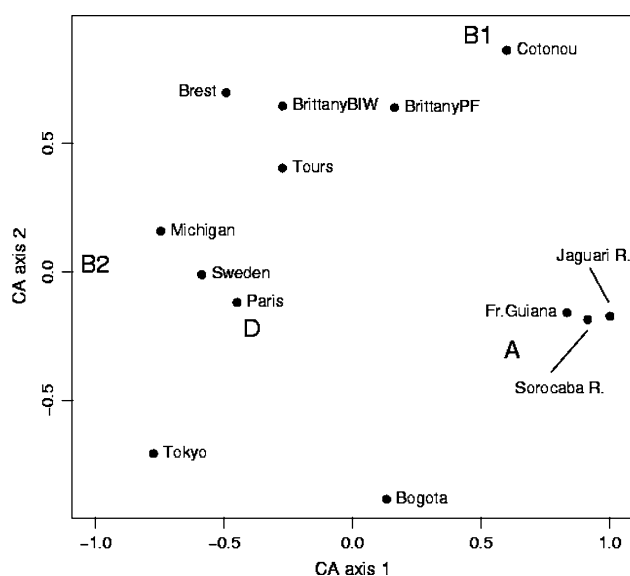


Fig. 1 Reciprocal ordination of *E. coli* strains and their phylogenetic groups. The figure shows the scores on the first two CA axis of samples of *E. coli* from sites worldwide, as well the scores of the phylogenetic groups found in these samples. Clusters of samples in this ordination space indicate samples with similar proportions of strains of each group, and clusters of groups indicate those that tend to occur in the same samples. Finally, groups and samples that are associated fall close. Eigenvalues are $\lambda_1 = 0.241$ for axis 1 and $\lambda_2 = 0.033$ for axis 2, from a total inertia of 0.297. Samples are (a) fecal isolates from human populations living in Brest, Brittany (PF = pig farmers, BIW = Bank workers), and Tours (all in France); in Michigan (USA), Tokyo (Japan), Bogota (Colombia), Cotonou (Benin), and French Guyana (Escobar-Páramo et al. 2004), and in Sweden Nowrouzian et al. (2005); (b) isolates from superficial water from Rivers Sorocaba and Jaguari, São Paulo State, Brazil

samples from the tropics were more dispersed in the ordination space, and their group frequencies were significantly different ($\chi^2 = 46.6$, $P < 0.001$), a result that did not change with the exclusion of any of the samples. However, the proximity in CA ordination of the samples from the rivers Jaguari and Sorocaba and those from the French Guiana Amerindians indicates that they have similar proportions of each phylogenetic group. For the French Guiana Amerindians these proportions were 63.4% of strains from group A, 20.4% from group B1, 3.2% from group B2 and 12.9% from group D, and for samples from the rivers Jaguari and Sorocaba the proportions were 68.9–66.2%, 21.3–20.6%, 0–1.5%, and 9.8–11.75%, respectively. As expected, the frequencies of the phylogenetic groups in the rivers Jaguari and Sorocaba did not differ from those in the French Guiana sample ($\chi^2 = 2.67$, $P = 0.88$), but differed from the Bogota ($\chi^2 = 31.3$, $P < 0.001$) and the Cotonou ($\chi^2 = 27.1$, $P < 0.001$) samples.

The large number of strains from group A and to a lesser extent from group B1 observed in the Jaguari and the Sorocaba Rivers is a matter of concern since according to

Escobar-Páramo et al. (2004), *E. coli* from groups A and B1 can emerge as intestinal pathogenic strains. Taken all together, our data emphasize that the contamination of surface water by fecal pollution is always a potential threat to animal and human health.

Acknowledgments This work was supported by grant 2000/05721-8 from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). R.H.O. had fellowship from CAPES. L.M.M.O. had research fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

- Carson CA, Shear BL, Ellersieck MR, Schnell JD (2003) Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. *Appl Environ Microbiol* 69(3):1836–1839
- CETESB (2004) Relatório de qualidade de águas interiores do Estado de São Paulo—2003. CETESB, São Paulo <http://www.cetesb.sp.gov.br>
- Clermont O, Bonacorsi S, Bingen E (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* 66(10):4555–4558
- Dixit SM, Gordon DM, Wu XY, Chapman T, Kailasapathy K, Chin JJC (2004) Diversity analysis of commensal porcine *Escherichia coli*-associations between genotypes and habitat in the porcine gastrointestinal tract. *Microbiology* 150:1735–1740
- Dobrindt U, Agerer F, Michaelis K, Janka A, Buchrieser C, Samuelson M, Svanborg C, Gottschalk G, Karch H, Hacker J (2003) Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J Bacteriol* 185:1831–1840
- Dombek PE, Johnson LK, Zimmerley ST, Sadowsky MJ (2000) Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Appl Environ Microbiol* 66:2572–2577
- Escobar-Páramo P, Grenet K, Le Menac'h A, Rode L, Salgado E, Amorin C, Gouriou S, Picard B, Rahimy MC, Andreumont A, Denamur E, Ruimy R (2004) Large-scale population structure of human commensal *Escherichia coli* isolates. *Appl Environ Microbiol* 70(9):5698–5700
- Everitt BS (1977) The analysis of contingency tables. Chapman & Hall, London
- Fienberg SE (1978) The analysis of cross-classified categorical data. MIT Press, Massachusetts
- Gauch HGJ (1982) Multivariate analysis in community ecology. Cambridge University Press, Cambridge
- Gordon DM, Cowling A (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 149:3575–3586
- Goulet PH, Picard B (1986) Comparative esterase electrophoretic polymorphism of *Escherichia coli* isolates obtained from animal and human sources. *J Gen Microbiol* 132:1843–1851
- Herzer PJ, Inouye S, Inouye M, Whittam TS (1990) Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol* 172:6175–6181
- Johnson JR, O'Bryan TT, Kuskowski MA, Maslow JN (2001) Ongoing horizontal and vertical transmission of virulence genes and *papA* alleles among *Escherichia coli* blood isolates from patients with diverse source bacteremia. *Infect Immun* 69:5363–5374

- Johnson JR, Stell AL (2000) Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J Infect Dis* 181:261–272
- Kaper JB, Nataro JP, Mobley HLT (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2:123–140
- Müller EE, Ehlers MM, Grabow WOK (2001) The occurrence of *E. coli* O157:H7 in South African water sources intended for direct and indirect human consumption. *Water Res* 35:3085–3088
- Nowrouzian FL, Wold AE, Adlerberth I (2005) *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J Infect Diseases* 191:1078–1083
- Ohno A, Marui A, Castro ES, Reyes AAB, Elio-Calvo D, Kasitani H, Ishii Y, Yamaguchi K (1997) Enteropathogenic bacteria in the La Paz River of Bolivia. *Am J Trop Med Hyg* 57:438–444
- Oksanen J, Kindt R, O'Hara RB (2005) Vegan: community ecology package version 1.6-9. <http://cc.oulu.fi/~jarioksa/>
- Parry SM, Palmer SR (2000) The public health significance of VTEC O157. *Symp Ser Soc Appl Microbiol* 88:1S–9S
- Parveen S, Portier KM, Robinson K, Edmiston L, Tamplim MI (1999) Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. *Appl Environ Microbiol* 65:3142–3147
- Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N, Bingen E, Elion J, Denamur E (1999) The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* 67:546–553
- Pupo GM, Karaolis DKR, Lan R, Reeves PR (1997) Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. *Infect Immun* 65(7):2685–2692
- R Development Core Team (2005) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- Russo TA, Johnson JR (2003) Medical and economic impact of extraintestinal infections due to *Escherichia coli*: an overlooked epidemic. *Microbes Infect* 5:449–456
- Selander RK, Korhonen TK, Vaisanen-Rhen V, Williams PH, Pattison PE, Caugant DA (1986) Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. *Infect Immun* 52:213–222
- Simiyu KW, Gathura PB, Kyule MN, Kanja LW, Ombui JN (1998) Toxin production and antimicrobial resistance of *Escherichia coli* river water isolates. *East Afr Med J* 75:699–702
- Ter Braak CJF (1995) Ordination. In: Jongman RHG, Ter Braak CJF, van Tongeren OFR (eds) *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge, pp 91–173
- Turner SJ, Lewis GD, Bellamy AR (1997) A genomic polymorphism located downstream of the *gcvP* gene of *Escherichia coli* that correlates with ecological niche. *Mol Ecol* 6:1019–1032
- Venables WN, Ripley BD (2002) *Modern applied statistics with S* 4 edn. Springer, New York
- Vinten AJA, Lewis DR, McGeachan M, Duncan A, Aitken M, Hill C, Crawford C (2004) Predicting the effect of livestock inputs of *E. coli* on microbiological compliance of bathing waters. *Water Res* 38:3215–3224