

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



CAMILA CARLOS

**“Identificação de marcadores moleculares hospedeiro-específicos de
Escherichia coli de águas superficiais do Estado de São Paulo”**

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
CAMILA CARLOS
e aprovada pela Comissão Julgadora.

Laura m. m. Ottoboni

Dissertação apresentada ao Instituto de
Biologia para obtenção do Título de
Mestre em Genética e Biologia
Molecular, na área de Genética de
Microrganismos.

Orientadora: Profa. Dra. Laura Maria Mariscal Ottoboni

Campinas, 2010

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

C195i	<p>Carlos, Camila Identificação de marcadores moleculares hospedeiro-específicos de <i>Escherichia coli</i> de águas superficiais do Estado de São Paulo / Camila Carlos. – Campinas, SP: [s.n.], 2010.</p> <p>Orientadora: Laura Maria Mariscal Ottoboni. Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.</p> <p>1. <i>Escherichia coli</i>. 2. Rep-PCR. 3. Espectroscopia de infravermelho. 4. Filogenia. I. Ottoboni, Laura Maria Mariscal. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.</p> <p style="text-align: right;">(rcdt/ib)</p>
--------------	--

Título em inglês: Identification of host-specific molecular markers of *Escherichia coli* from State of São Paulo.

Palavras-chave em inglês: *Escherichia coli*; Rep-PCR; Infrared spectroscopy; Phylogeny.

Área de concentração: Genética de Microorganismos.

Titulação: Mestre em Genética e Biologia Molecular.

Banca examinadora: Laura Maria Mariscal Ottoboni, Luciana Campos Paulino, Valéria Maria de Oliveira.

Data da defesa: 24/02/2010.

Programa de Pós-Graduação: Genética de Microorganismos.

Campinas, 24 de fevereiro de 2010 .

BANCA EXAMINADORA

Profa. Dra . Laura Maria Mariscal Ottoboni (Orientadora)

Laura m. m. Ottoboni
Assinatura

Profa. Dra. Valéria Maria de Oliveira

Valéria Maria de Oliveira
Assinatura

Profa. Dra . Luciana Campos Paulino

Luciana Campos Paulino
Assinatura

Prof. Dr. Flávio Henrique da Silva

Assinatura

Profa. Dra. Edi Lúcia Sartorato

Assinatura

AGRADECIMENTOS

Agradeço a Deus por ter me guiado e abençoado nessa caminhada.

À minha mãe pelo seu apoio e amor incondicional e por tudo o que sou hoje.

À minha irmã Beatriz pelo seu carinho e admiração.

Ao Glauber por estar sempre ao meu lado, por seu amor, por sua paciência e por tudo que tem feito por mim.

A toda minha família pelo apoio e carinho.

À minha orientadora Laura Ottoboni por acreditar em mim e por sua fundamental contribuição para a minha formação, desde a graduação até hoje.

A todos os meus amigos que, além dos momentos de descontração, contribuíram de alguma forma para a profissional que estou me formando, Gustavo, Laura, Victor, Mathias e Maurício.

A todos os colegas de laboratório por sempre me ajudarem, me ensinarem, e pela amizade, especialmente à Daniela, Fabiana e Viviane.

Agradeço aos membros e suplentes da banca, Profa. Valéria M. de Oliveira, Profa. Luciana Paulino, Profa. Edi Sartorato e Prof. Flávio Henrique da Silva, pela contribuição dada ao trabalho.

Agradeço aos colaboradores deste projeto, Dra. Inês Sato, Nancy Stoppe, Dra. Elayse Hachich, Prof. Luiz Amaral e Profa. Tânia Gomes pela contribuição dada ao trabalho.

A todos os meus amigos da faculdade pela paciência e apoio, especialmente à Carol, Camila, Talita, Giovanna, Sassaki.

Às minhas amigas, Daniele, Natália, Evelyn, Camila, Débora, Juliana e Carol, por estarem sempre comigo apesar de todas as mudanças pelas quais passamos.

Às minhas amigas da Casa das bonecas por nunca se esquecerem de mim, Vanessa, Priscila, Juliana e Cíntia.

Aos colegas e funcionários do CBMEG e da Secretaria de Pós-Graduação.

À FAPESP pelo apoio financeiro e ao CBMEG e Unicamp pelo apoio institucional.

A todos que contribuíram de alguma forma para este trabalho.

ÍNDICE

RESUMO	1
SUMMARY.....	3
INTRODUÇÃO	5
QUALIDADE MICROBIOLÓGICA DA ÁGUA E INDICADORES DE CONTAMINAÇÃO FECAL	5
ESCHERICHIA COLI	7
RASTREAMENTO DE FONTES DE CONTAMINAÇÃO MICROBIANA	9
OBJETIVOS	12
CAPÍTULO 1: ESTRUTURA POPULACIONAL DE <i>ESCHERICHIA COLI</i> DE ORIGEM AVIÁRIA, BOVINA, CAPRINA, HUMANA, SUÍNA E OVINA DO ESTADO DE SÃO PAULO	13
INTRODUÇÃO	14
MATERIAIS E MÉTODOS	15
RESULTADOS E DISCUSSÃO	19
REFERÊNCIAS	29
CAPÍTULO 2: UTILIZAÇÃO DE BOX E (GTG)₅-PCR NA DISCRIMINAÇÃO DE <i>ESCHERICHIA COLI</i> DE DIFERENTES ANIMAIS	34
INTRODUÇÃO	35
MATERIAIS E MÉTODOS	36
RESULTADOS E DISCUSSÃO	39
REFERÊNCIAS	52
CAPÍTULO 3: UTILIZAÇÃO DE ESPECTROSCOPIA NO INFRAVERMELHO COM TRANSFORMADA DE FOURIER (FT-IR) NA DISCRIMINAÇÃO DE <i>ESCHERICHIA COLI</i> DE DIFERENTES ANIMAIS	56
INTRODUÇÃO	57
MATERIAIS E MÉTODOS	58
RESULTADOS E DISCUSSÃO	60
REFERÊNCIAS	70
CONCLUSÕES	74
REFERÊNCIAS	76

RESUMO

Coliformes e enterococos fecais de origem humana ou animal na água podem indicar a presença de patógenos de veiculação hídrica como, por exemplo, *Salmonella* e *Giardia*. A identificação da fonte de contaminação fecal (lançamento de esgoto doméstico, escoamento de fezes animais de criação no solo, de animais silvestres, aves e outros) é importante para a implantação de medidas efetivas de gerenciamento e remediação de águas superficiais. Dessa forma, o desenvolvimento de métodos para identificação da fonte de contaminação fecal é de fundamental importância para as ações de controle, para preservar a integridade dos corpos d'água e para proteger a saúde da população. Até o momento, não existe um método único e universal para este tipo de análise e, no Brasil, as pesquisas nessa área são praticamente nulas. Assim sendo, este trabalho teve como objetivo a obtenção de marcadores moleculares hospedeiro-específicos em *Escherichia coli*, que permitam a identificação da fonte animal de contaminação fecal em águas superficiais. Neste trabalho foram utilizadas 174 linhagens de origem humana, 50 de origem bovina, 39 de origem suína, 16 de origem aviária, 29 de origem ovina, 16 de origem caprina, 44 de esgoto, 36 de reservatórios com contaminação esperada de origem humana e 30 de rios e reservatórios com contaminação esperada de origem animal. A determinação do grupo filogenético de todas as linhagens foi realizada pela detecção dos genes *chuA* e *yjaA* e do fragmento Tspe4.C2 por PCR. Os resultados mostraram que a distribuição dos grupos filogenéticos principais A, B1, B2 e D foi diferente entre os hospedeiros analisados, o que permitiu a predição da fonte de contaminação fecal da maioria dos pontos de amostragem. Cem linhagens de humanos e todas as linhagens de origem animal foram analisadas por BOX- e (GTG)₅-PCR. O BOX-PCR apresentou uma taxa global de classificação correta de 63,70%, o

(GTG)₅-PCR de 49,10% e quando os dois métodos foram utilizados a taxa foi de 57,61%. Outras 32 linhagens de humanos e 28 de esgoto também foram analisadas por BOX-PCR, sendo que, 59,4% das linhagens de humanos foram corretamente classificadas e 85,2% das linhagens de esgoto foram classificadas como de humanos. Vinte linhagens de humanos, 15 de bois e 16 de galinhas foram analisadas por espectroscopia no infravermelho com transformada de Fourier (FT-IR). Utilizando-se um modelo PLS-DA com a segunda derivada do espectro na região 2816 e 3026 cm⁻¹ foi possível separar completamente as linhagens segundo sua origem animal. Assim sendo, a espectroscopia FT-IR foi considerada a técnica mais promissora para futuros estudos de rastreamento de fonte de contaminação fecal.

SUMMARY

The detection of fecal coliforms and enterococci in water indicates the presence of waterborne pathogens such as *Salmonella* and *Giardia*. The identification of the source of fecal contamination is important for the effective management of superficial water pollution. The development of methods for the identification of the source of fecal contamination is essential to preserve the quality of the water systems and to protect the public health. Until now, there is no universal method for this analysis and, in Brazil, there is no research in this area. In this way, the aim of this work was to obtain host-specific molecular markers in *Escherichia coli* for the identification of the source of fecal contamination in superficial water. For this work it was used, 174 strains from humans, 50 from cows, 39 from pigs, 29 from sheep, 16 from goat, 16 from chickens, 44 from sewage, 36 from water reservoirs whose the expected contamination source is human and 30 from water reservoirs and rivers whose the expected contamination source is animal. The determination of the phylogenetic groups of all strains was performed by the detection of the genes *chuA* and *yjaA* and the fragment Tspe4.C2 by PCR. The results showed that the distribution of the phylogenetic groups A, B1, B2 and D differs among the hosts analyzed, which allowed the prediction of the contamination source of most of the environmental samples. One hundred strains from humans, 50 from cows, 39 from pigs, 29 from sheep, 16 from goat and 16 from chickens were analyzed by BOX- and (GTG)₅-PCR. The BOX-PCR presented an overall rate of correct assignment of 63.70%, the (GTG)₅-PCR of 49.10% and, when the two methods were used, the rate was 57.61%. Thirty two other strains from humans and 28 strains from sewage were analyzed by BOX-PCR and compared with the profiles previously obtained. 59.4% of the human strains were correctly assigned as human and 85.2% of the sewage strains

were assigned as human. Twenty strains from humans, 15 from cows and 16 from chickens were analyzed by Fourier transform infrared spectroscopy (FT-IR). The use of a PLS-DA model with the second derivative of spectra at the region 2816 and 3026 cm^{-1} made it possible to completely discriminate the strains according to the animal source. Therefore, the FT-IR spectroscopy was considered the most promising method for the identification of fecal contamination.

INTRODUÇÃO

Qualidade Microbiológica da Água e Indicadores de Contaminação Fecal

A contaminação fecal dos corpos hídricos é uma das principais causas de doenças entéricas veiculadas pela água no mundo, as quais tem sido responsáveis pela morte de cerca de 2 milhões de crianças por ano (WHO, 2007). Essa situação tem se agravado nos últimos anos devido à contaminação frequente das águas de consumo humano, águas recreacionais e águas de irrigação por patógenos emergentes como *Giardia*, *Cryptosporidium*, *Escherichia coli* O157:H7, entre outros.

Para garantir a qualidade da água do ponto de vista microbiológico é importante estabelecer programas de controle de poluição que limitem o lançamento de despejos de origem doméstica e animal nos recursos hídricos. A adoção de padrões bacteriológicos de qualidade de água é um instrumento essencial nesse processo para redução da contaminação fecal, principalmente nas águas superficiais. No Estado de São Paulo, uma porcentagem significativa das águas superficiais não atende aos padrões bacteriológicos de uso estabelecidos pela Resolução CONAMA 357 (Brasil, 2005) devido às altas densidades de indicadores de contaminação fecal. Isso representa uma preocupação constante dos órgãos de saúde e meio ambiente devido aos riscos da utilização da água contaminada para fins recreacionais, irrigação e aquicultura. Alguns desses corpos d'água são mananciais utilizados na captação para consumo humano, e se não forem adequadamente tratados, representam um risco significativo de ocorrência dessas doenças.

A contaminação fecal dos recursos hídricos, além de representar um risco à saúde humana, pode resultar em perda econômica significativa devido aos impactos causados em áreas recreacionais e aquiculturas (Rabinovici *et al.*, 2004).

Os indicadores bacteriológicos de contaminação fecal usualmente empregados na avaliação da qualidade da água e estabelecidos em legislações são os coliformes termotolerantes, *E. coli* e enterococos fecais. Esses microrganismos normalmente estão presentes em densidades elevadas nas fezes de animais de sangue quente, incluindo animais silvestres, bovinos e humanos e, portanto podem indicar a presença de patógenos entéricos, cuja detecção e quantificação são muito mais complexas, o que dificulta sua determinação em trabalhos de rotina. Atualmente, existem muitos debates sobre quais os melhores indicadores de contaminação fecal, uma vez que muitas vezes não existe uma associação direta entre o microrganismo indicador e o patógeno cuja presença ele deveria indicar (Meays *et al.*, 2004). Nesse sentido, existe questionamento se as bactérias seriam bons indicadores da presença de vírus e protozoários, que são organismos mais resistentes aos processos de tratamento e as condições adversas do meio ambiente. Entretanto é a presença de *E. coli* ou coliformes termotolerantes na água ainda é usada como indicador da ocorrência de contaminação fecal no Brasil (Brasil, 2004).

Um indicador ideal de contaminação segundo Scott *et al.* (2002) seria um microrganismo não patogênico, que pode ser detectado e enumerado por técnicas simples e rápidas e que apresenta sobrevivência no meio ambiente similar aos patógenos. Apesar de se dispor hoje de uma série de indicadores microbianos para detectar contaminação fecal em águas costeiras, interiores e subterrâneas, esses métodos não permitem distinguir se essa contaminação é de origem humana ou animal. Um indicador ideal para rastrear a fonte de contaminação microbiana deve apresentar todas as características do indicador clássico acima e ter poder de discriminar a

origem do hospedeiro (USEPA, 2005). O uso desses indicadores para identificar a fonte de contaminação fecal é imperativo para preservar a integridade dos corpos hídricos e proteger a saúde da população.

Escherichia coli

Bacterium coli foi descrita no século XIX por Theodore Escherich (1855) em fezes humanas, onde ocorria em altas densidades e estava frequentemente associada ao bacilo da febre tifóide. Em 1912, já era utilizada como indicador da qualidade sanitária da água. Na década de 50, passou a ser denominada *Escherichia coli* e foi escolhida pelos cientistas como modelo para estudos de processos biológicos básicos, tais como vias metabólicas, regulação gênica, transdução de sinais, estrutura de parede celular e conjugação (Kuhnert *et al.*, 2000). Apesar de sua importância como agente etiológico de doenças animais e humanas poucos laboratórios estudavam *E. coli* como um patógeno, mesmo esta sendo responsável pela infecção de milhões de pessoas em todo o mundo anualmente (Kuhnert *et al.*, 2000).

E. coli é um dos principais componentes, dentre as Enterobacteriaceas, da microbiota intestinal de humanos e animais (Kuhnert *et al.*, 2000). Estes microrganismos são bacilos Gram negativos, anaeróbios facultativos e se movimentam usualmente através de flagelos peritríquios. O genoma da bactéria varia de 2,3 a 3,0 x 10⁹ daltons e o conteúdo de G + C fica entre 49 e 52% (Brenner & Falkow, 1971; Brenner, 1983; Hartl & Dykhuizen, 1984). Esta bactéria está presente em concentrações elevadas na flora intestinal de humanos e animais, em relação de comensalismo. Em outras partes do corpo pode entretanto causar doenças graves, como por exemplo, infecções urinárias, bacteremia e meningite. Algumas linhagens enteropatogênicas

podem causar diarreia aguda, devido à presença de fatores de virulência, como por exemplo, as *E. coli* enterohemorrágicas (EHEC), as *E. coli* enterotoxigênicas (ETEC), as *E. coli* enteropatogênicas (EPEC), as *E. coli* enteroinvasivas (EIEC), as *E. coli* enteroagregativas (EAEC) e as *E. coli* difusamente aderente (DAEC) (WHO, 2004). Essa versatilidade tem sido associada ao fato de diferentes linhagens de *E. coli* terem adquirido diferentes genes de virulência (Kuhnert *et al.*, 2000). Estima-se que 10-20% da informação genética encontrada em uma *E. coli* patogênica não está presente na linhagem K-12, não patogênica (Kuhnert *et al.*, 2000).

O epitélio intestinal dos animais de sangue quente é o habitat primário da *E. coli*, o que faz com que ela seja abundante em fezes de origem humana e animal, onde está presente em concentrações de 10^9 células por grama de fezes frescas. Esta bactéria é encontrada em esgotos, efluentes tratados, em águas, sedimentos e solos sujeitos a contaminação fecal recente de origem humana, ou de outros mamíferos e aves (WHO, 2004).

A *E. coli* é considerada o principal indicador da qualidade da água e de contaminação fecal (Edberg *et al.*, 2000). A presença dessa bactéria em água indica contaminação fecal potencial, uma vez que uma série de patógenos pode ser veiculada através das fezes e requer ação imediata no que diz respeito ao controle de qualidade das águas nos seus mais diversos usos: consumo humano, recreação, irrigação, aquicultura, e outros (WHO, 2004).

A legislação brasileira sobre a qualidade de águas destinadas ao consumo humano, águas minerais e águas naturais determina que seja analisada preferencialmente a *E. coli*, que deve estar ausente nessas águas, aceitando como alternativa os coliformes termotolerantes (Brasil, 2004; Brasil, 2005). A Organização Mundial da Saúde também dá preferência à *E. coli* para verificação da qualidade microbiológica de águas de consumo humano (WHO, 2004). Essa bactéria é considerada, juntamente com o grupo dos enterococos fecais, o melhor indicador de

balneabilidade (USEPA, 2002). As legislações canadense e européia adotam os enterococos e *E. coli* para avaliação da qualidade de águas recreacionais (Health and Welfare Canadá, 1992, Directive 2006/7/EC). A Resolução CONAMA 357/2005, que dispõe sobre a classificação de corpos de água e diretrizes ambientais para seu enquadramento, utiliza os coliformes termotolerantes como padrão de qualidade microbiológica, mas permite sua substituição pela *E. coli*, de acordo com critérios estabelecidos pelo órgão ambiental competente (BRASIL, 2005).

Rastreamento de Fontes de Contaminação Microbiana

Os métodos microbiológicos de rastreamento de fontes de contaminação, também chamados de RFM (Rastreamento de Fontes Microbianas), se baseiam na suposição de que diferentes sistemas intestinais selecionam diferentes populações de microrganismos, principalmente, devido a diferenças na dieta e no sistema digestivo de diferentes hospedeiros. Os métodos desenvolvidos para RFM também são relevantes para outros campos de pesquisa, como segurança alimentar e microbiologia agrícola e veterinária (Santo Domingo *et al.*, 2007).

Em países desenvolvidos, as fontes de contaminação não caracterizadas são as maiores responsáveis pela poluição fecal dos recursos hídricos (Santo Domingo *et al.*, 2007). Nos Estados Unidos da América, recentemente, a necessidade do RFM, se tornou uma prioridade dos estados e territórios, em virtude do requerimento federal de desenvolvimento e execução da quantidade máxima de poluente (TMDL - *total maximum daily load*) que um corpo hídrico pode receber mantendo os padrões exigidos de qualidade (Santo Domingo *et al.*, 2007). A TMDL é também utilizada para verificar as contribuições de fontes pontuais e não pontuais de poluição. Em geral, a contribuição das fontes pontuais, como estações de tratamento de esgoto ou efluentes

industriais, é fixada por legislação específica, sendo de mais fácil controle. Por outro lado, as fontes não pontuais, que incluem agricultura, reflorestamento, animais silvestres e escoamento do solo urbano, são de difícil controle e na maior parte dos casos responsáveis por alterações na qualidade da água. Portanto, é necessária a distinção entre fontes pontuais e não-pontuais, a determinação da categoria da fonte não-pontual e igualmente, da origem animal da contaminação, para eliminá-la de forma efetiva. A contaminação animal pode ter origem em animais de criação, animais de estimação ou animais silvestres, sendo de maior relevância a contaminação proveniente de criações de alto impacto comercial, como por exemplo bovinos, caprinos, suínos e as avícolas, devido ao grande número de animais envolvidos (Simpson *et al.*, 2002). Pouco se sabe sobre os riscos que a contaminação da água por fezes animais oferece à saúde humana, portanto é importante a prevenção e remediação dessas fontes poluidoras (Field & Samadpour, 2007).

Os métodos de RFM podem ser divididos em dependentes de cultivo e independentes de cultivo. Dentre os métodos independentes de cultivo e de biblioteca estão o DGGE e PCR de sequências específicas de vírus e bactérias fecais. Estes métodos são rápidos, porém normalmente são limitados por identificarem poucas espécies de hospedeiro além dos humanos. Os métodos dependentes de cultivo podem ser divididos em independentes de biblioteca e dependentes de biblioteca. Os métodos independentes de biblioteca consistem no enriquecimento do organismo alvo para posterior identificação, como por exemplo, a hibridização com genes específicos. Os métodos dependentes de biblioteca podem consistir em análises fenotípicas, como por exemplo, análise de resistência à antibióticos (ARA) (Jiang *et al.*, 2007; Price *et al.*, 2007), perfil de utilização de fontes de carbono (Harwood *et al.*, 2003; Moussa and Massengale, 2008) e MALDI-TOF/MS (Siegrist *et al.*, 2007), ou em análises genotípicas, como ribotipagem (Parveen *et al.*,

1999; Nelson *et al.*, 2008), PFGE (*Pulse Field Gel Electrophoresis*) (Casarez *et al.*, 2007) e rep-PCR (*Repetitive Element Sequence PCR*) (Dombek *et al.*, 2000; Mohapatra *et al.*, 2007). Os métodos dependentes de biblioteca permitem a identificação pela comparação entre os perfis encontrados na amostra teste e os padrões obtidos dos isolados das bibliotecas (Santo Domingo *et al.*, 2007; Field & Samadpour, 2007). Os métodos de RFM devem ser confiáveis, apresentar relevância para o gerenciamento de recursos hídricos, apresentar baixo custo e ser de fácil interpretação.

A comparação de duas ou mais metodologias é bastante enriquecedora para auxiliar na escolha de métodos para o gerenciamento de águas. A comparação entre ARA e PFGE mostrou que apesar da análise por PFGE ser mais informativa para o RFM, a ARA apresenta menor custo e pode ser aplicada para TMDL (Price *et al.*, 2007). Recentemente, Casarez *et al.* (2007) compararam quatro métodos de RFM: ERIC-PCR, ribotipagem, PFGE e análise de resistência a antibióticos Kirby-Bauer. Os autores concluíram que o uso de quatro métodos torna a identificação de fonte da contaminação mais confiável, sendo que o uso conjunto de dois métodos já é mais informativo do que a análise individual de cada método.

Até o momento, não existe um método único e universal para este tipo de análise e, no Brasil, as pesquisas nessa área são praticamente nulas. Portanto a obtenção de marcadores moleculares que permitam a identificação de fontes de contaminação fecal humana e animal poderá ajudar na implantação de técnicas que possibilitem rastrear essas fontes de contaminação em águas superficiais do Estado de São Paulo.

OBJETIVOS

Objetivo geral:

Este trabalho teve como objetivo a obtenção de métodos e marcadores moleculares para a identificação de fontes de contaminação fecal em águas superficiais do Estado de São Paulo.

Objetivos específicos:

CAPÍTULO 1: Estrutura populacional de *Escherichia coli* de origem aviária, bovina, caprina, humana, suína e ovina do estado de São Paulo

- Verificação do padrão de distribuição dos grupos filogenéticos A, B1, B2 e D nos diferentes hospedeiros animais.

CAPÍTULO 2: Utilização de BOX e (GTG)₅-PCR na discriminação de *Escherichia coli* de diferentes animais

- Discriminação de linhagens de *Escherichia coli* de diferentes hospedeiros empregando rep-PCR (BOX- e (GTG)₅-PCR).

CAPÍTULO 3: Utilização de espectroscopia no infravermelho com transformada de Fourier (FT-IR) na discriminação de *Escherichia coli* de diferentes animais

- Utilização de linhagens de *Escherichia coli* de diferentes hospedeiros para identificação de marcadores moleculares empregando FT-IR.

CAPÍTULO 1

**Estrutura populacional de *Escherichia coli* de origem aviária,
bovina, caprina, humana, suína e ovina do Estado de São Paulo**

INTRODUÇÃO

Com base nas árvores filogenéticas geradas pelas técnicas de MLEE (*Multilocus Enzyme Electrophoresis*) e ribotipagem foram definidos quatro grupos filogenéticos principais de *E. coli*: A, B1, B2 e D (Selander *et al.*, 1986; Herzer *et al.*, 1990). Esses grupos diferem em diversas características como virulência, nicho ecológico e história de vida. As linhagens virulentas extra-intestinais pertencem em sua maioria ao grupo B2 e, em menor proporção ao grupo D (Picard *et al.*, 1999; Johnson & Stell, 2000). As linhagens comensais geralmente pertencem aos grupos A e B1 (Bingen *et al.*, 1998) e as patogênicas intestinais, geralmente se enquadram nos grupos A, B1 e D (Pupo *et al.*, 1997).

Bonacorsi *et al.* (2000) propuseram que alguns genes e fragmentos de DNA poderiam ser utilizados como marcadores para a determinação do grupo filogenético de linhagens de *E. coli*. Assim sendo, foi proposto a utilização dos genes *chuA* (que codifica para um heme receptor) e *yjaA* (produto com função desconhecida) e do fragmento de DNA TSPE4.C2 para rápida identificação do grupo filogenético de linhagens de *E. coli* por PCR multiplex (Clermont *et al.*, 2000).

Alguns autores analisaram a distribuição dos grupos filogenéticos de linhagens de *E. coli* isoladas de fezes animais e humanas. Escobar-Páramo *et al.* (2006) observaram que a prevalência dos grupos filogenéticos de *E. coli* em aves, mamíferos não-humanos e humanos é diferente, com uma predominância dos grupos D/B1, A/B1 e A/B2, respectivamente. Além disso, esses autores identificaram uma linhagem pertencente ao grupo filogenético B1 e positiva para o gene *hly* (que codifica pra uma hemolisina), que aparentemente é específica de animais. Baldy-Chudzik *et al.* (2008) encontraram diferentes distribuições dos grupos filogenéticos entre animais de zoológico,

encontrando uma maior proporção do grupo B1 nos herbívoros, e em carnívoros e onívoros uma maior proporção de linhagens do grupo A. Clermont *et al.* (2008) encontraram uma linhagem de *E. coli* que pertence ao grupo filogenético B2 e que, aparentemente, só ocorre em humanos. Essa linhagem foi detectada em pessoas que vivem na África, Europa e América do Sul. Os autores sugerem que esta linhagem pode vir a ser um indicador de contaminação fecal de origem humana.

Assim sendo, o objetivo deste trabalho foi analisar a distribuição dos grupos filogenéticos de *E. coli* em linhagens de origem aviária, bovina, caprina, humana, suína e ovina e verificar o potencial desta análise como uma ferramenta de identificação da fonte animal de contaminação fecal.

MATERIAIS E MÉTODOS

Linhagens utilizadas e condições de isolamento e cultivo

Foram utilizadas na construção do banco de linhagens, 174 linhagens de origem humana, 50 de origem bovina, 39 de origem suína, 16 de origem aviária, 29 de origem ovina e 16 de origem caprina (Tabela 1). Além disso, 110 linhagens ambientais foram utilizadas como amostras de validação externa; Destas linhagens, 44 são das estações de tratamento de esgoto (ETE) Jesus Netto e Barueri, 36 de reservatórios com contaminação esperada pela CETESB (Companhia Ambiental do Estado de São Paulo) como de origem humana (Billings e Guarapiranga) e 30 de rios e reservatórios com contaminação esperada pela CETESB como de origem animal (Tietê, Reservatório Tanque Grande e Aguapeí) (Tabela 2). Todas as linhagens de origem animal foram

isoladas de indivíduos saudáveis, que não apresentavam diarreia e cada linhagem foi isolada de um único indivíduo. As linhagens de rios e reservatórios foram isoladas entre julho e agosto de 2009 e as linhagens de esgoto entre março e maio de 2008.

As linhagens de *E. coli* foram isoladas a partir de fezes, por plaqueamento direto em agar Mac Conkey. As amostras de esgoto foram diluídas e filtradas em agar mTEC modificado (USEPA, 2002). As colônias típicas foram re-isoladas quando necessário nos meios seletivos: ágar Mac Conkey e triadas no meio de IAL (Instituto Adolfo Lutz), ou EPM (Escola Paulista de Medicina) e MILi, ágar citrato de Simmon's, EC MUG e reação de oxidase. As linhagens com características típicas de *E.coli* foram inoculadas em 1 mL de meio LB líquido (Sambrook *et al.*, 1989). Aproximadamente 10 µL das culturas obtidas foram plaqueadas separadamente em meio LB sólido (Sambrook *et al.*, 1989). Uma colônia isolada foi escolhida em cada uma das placas e inoculada em 2 mL de meio LB líquido para a preparação de cultura permanente com 25% de glicerol (Sambrook *et al.*, 1989). As culturas permanentes foram armazenadas a -80°C.

Tabela 1. Origem e número de linhagens de *E. coli* utilizadas neste trabalho.

Origem	Número de linhagens	Nomenclatura utilizada	Instituição doadora
Humana (crianças)	100	Hc	UNIFESP
Humana (adultos)	74	HUM	CETESB
Bovina	50	Bs	UNESP - Jaboticabal
Aviária	16	As	Unicamp
Suína	39	Ss	UNESP - Jaboticabal
Caprina	16	Cs	UNESP - Jaboticabal
Ovina	29	Os	UNESP - Jaboticabal
Esgoto	44	AMB	CETESB
Rios e reservatórios	66	Rio	CETESB
Total	434		

Tabela 2. Pontos de coleta das amostras de *E. coli* de origem ambiental.

Ponto	Descrição do ponto	Município	Fonte poluição esperada	Número de linhagens
Ani-1	Rio Tietê, a jusante da ETE de Suzano	Suzano	Animal	8
Ani-2	Reservatório do Tanque Grande, junto a barragem no município de Guarulhos	Guarulhos	Animal	12
Ani-3	Rio Aguapeí, Ponte que liga Tupi Paulista a Andradina	Junqueirópolis	Animal	10
Hum-1	Reservatório Guarapiranga – restaurante do Odair	São Paulo	Humano	8
Hum-2	Reservatório Guarapiranga – clube de campo Castelo	São Paulo	Humano	9
Hum-3	Reservatório Billings – Prainha do Jardim Los Angeles	São Bernardo do Campo	Humano	11
Hum-4	Reservatório Billings – Prainha em frente a ETE	São Bernardo do Campo	Humano	8
Esg-1	ETE Jesus Netto	São Paulo	Humana	24
Esg-2	ETE Barueri	São Paulo	Humana	20

Determinação dos grupos filogenéticos das linhagens

A determinação dos grupos filogenéticos das linhagens foi realizada de acordo com Clermont *et al.* (2000). Para tal, foram utilizados os *primers* ChuA.1 (5'-GACGAACCAACGGTCAGGAT-3') e ChuA.2 (5'-TGCCGCCAGTACCAAAGACA-3'), para amplificar um fragmento de 279 pb do gene *chuA*; YjaA.1 (5'-TGAAGTGTCAGGAGACGCTG-3') e YjaA.2 (5'-ATGGAGAATGCGTTCCTCAAC-3'), para amplificar um fragmento de 211 pb do gene *yjaA*; TspE4C2.1 (5'-GAGTAATGTCGGGGCATTC-3') e TspE4C2.2 (5'-CGCGCCAACAAAGTATTACG-3'), para amplificar o fragmento TSPE4.C2 de 152 pb. Os três marcadores foram amplificados na mesma reação. As reações de amplificação foram realizadas em um volume final de 20 µL

contendo: 200 ng de DNA, 4 pmoles de cada *primer* e pré-mix GoTaq Green Master Mix 1X concentrado (Promega). As condições de amplificação foram: desnaturação inicial a 94°C por 5 min, seguida de 30 ciclos de amplificação (94°C – 30 s, 55°C – 30 s, 72°C – 30 s) e extensão final a 72°C por 7 min. Os produtos de amplificação foram visualizados por eletroforese em gel de 2,0% de agarose, contendo brometo de etídio, em tampão TAE 1X (Sambrook *et al.*, 1989). Como marcador de peso molecular foi utilizado no gel o *Ladder* 100 pb (Invitrogen). Após a eletroforese, o gel foi fotografado sob luz U.V. e analisado. Linhagens com o perfil *chuA*+, *yjaA*+ foram classificadas no grupo B2; linhagens com o perfil *chuA*+, *yjaA*- foram classificadas como D; linhagens *chuA*-, Tspe4.C2+ foram classificadas como B1 e linhagens *chuA*-, Tspe4.C2- foram classificadas como A.

Os grupos filogenéticos A, B1, B2 e D podem ainda ser divididos em sete sub-grupos ou filo-tipos, de acordo com a combinação dos marcadores genéticos: filo-tipo A₀ (grupo A) *chuA*-, *yjaA*-, Tspe4.C2-; filo-tipo A₁ (grupo A) *chuA*-, *yjaA*+, Tspe4.C2-; grupo B1 *chuA*-, *yjaA*-, Tspe4.C2+; filo-tipo B2₂ (grupo B2) *chuA*+, *yjaA*+, Tspe4.C2-; filo-tipo B2₃ (grupo B2) *chuA*+, *yjaA*+, Tspe4.C2+; filo-tipo D₁ (grupo D) *chuA*+, *yjaA*-, Tspe4.C2-; filo-tipo D₂ (grupo D) *chuA*+, *yjaA*-, Tspe4.C2+ (Escobar-Páramo *et al.*, 2006).

Análises estatísticas

O índice de prevalência (P) foi calculado pela divisão do número de hospedeiros que apresentaram um determinado filo-tipo pelo número total de hospedeiros analisados (Bush *et al.*, 1997). Os resultados foram expressos em porcentagem.

O índice de Pianka foi calculado para avaliar a sobreposição de filo-tipos entre dois hospedeiros, utilizando-se a fórmula: $O = \sum p_j p_k / \sqrt{\sum p_j^2 \sum p_k^2}$, na qual p_j e p_k são as proporções de um filo-tipo nos hospedeiros j e k , respectivamente. Os resultados foram expressos em porcentagem (Pianka, 1973).

O teste G e os índices de diversidade de Shannon e Simpson foram feitos no software BioEstat v. 5.0 (Ayres *et al.*, 1998), usando os dados dos filo-tipos.

A frequência dos filo-tipos foram comparadas entre os hospedeiros utilizando-se a análise de correspondência (CA), que foi realizada no software STATISTICA 6.0 (Statsoft Inc., Tulsa, OK). As amostras de esgoto e ambientais foram utilizadas como amostras de validação externa dos modelos de CA.

RESULTADOS E DISCUSSÃO

Neste trabalho, 324 linhagens de *E. coli* isoladas de diferente animais, 44 linhagens isoladas de duas estações de tratamento de esgoto e 66 linhagens isoladas de rios e reservatórios foram classificadas em um dos quatro grupos filogenéticos principais A, B1, B2 e D utilizando o método PCR triplex descrito por Clermont *et al.* (2000). Com esta técnica, as linhagens puderam ser classificadas de acordo com a presença/ausência dos genes *chuA* e *yjaA* e do fragmento Tspe4.C2 (ver exemplo na Figura 1).

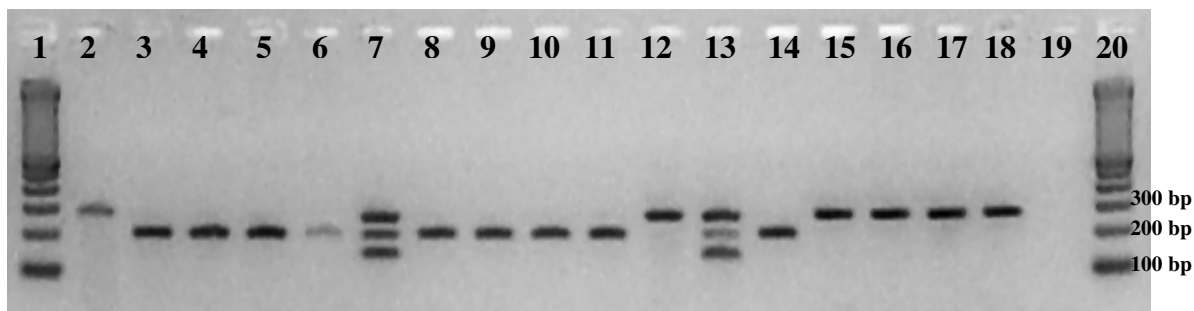


Figura 1. Gel de agarose 2% contendo os produtos de amplificação por PCR *triplex* para classificação das linhagens nos grupos filogenéticos A, B1, B2 e D. Um e 20 correspondem ao *Ladder* 100 pb (Invitrogen). As amostras de 2 a 18 correspondem às linhagens Hc53 a Hc70 e a amostra 19 corresponde ao controle negativo.

A Tabela 3 e a figura 2 mostram a distribuição dos filo-tipos entre os hospedeiros analisados. A prevalência de cada filo-tipo foi: $A_0 = 100\%$, $A_1 = 83.33\%$, $B1 = 100\%$, $B2_2 = 50\%$, $B2_3 = 16.67\%$, $D_1 = 66.67\%$, $D_2 = 66.67\%$. O grupo B1 e o filo-tipo A_0 foram encontrados em todos os animais, enquanto que o filo-tipo $B2_3$ só foi encontrado em humanos. O valor do teste G foi 207,66 (30 graus de liberdade) com uma probabilidade $p < 0,0001$ da distribuição dos filo-tipos ser ao acaso. Pode-se observar a predominância dos grupos A e D nas linhagens de origem humana e suína (mamíferos onívoros), a predominância dos grupos B1 nas linhagens de origem bovina, ovina e caprina (mamíferos herbívoros) e a predominância do grupo A nas linhagens de origem aviária. Segundo Gordon & Cowling (2003), a abundância relativa dos grupos filogenéticos nos mamíferos depende da dieta, da massa corpórea e do clima ao qual o hospedeiro está submetido. Os autores observaram nos mamíferos uma predominância dos grupos B1 e B2. É interessante notar que, no referido trabalho, nenhum dos animais estudados (peixes, sapos, tartarugas, cobras, lagartos, crocodilos, aves e mamíferos) apresentou

predominância do grupo D e que aves e mamíferos apresentaram números similares de linhagens do grupo D.

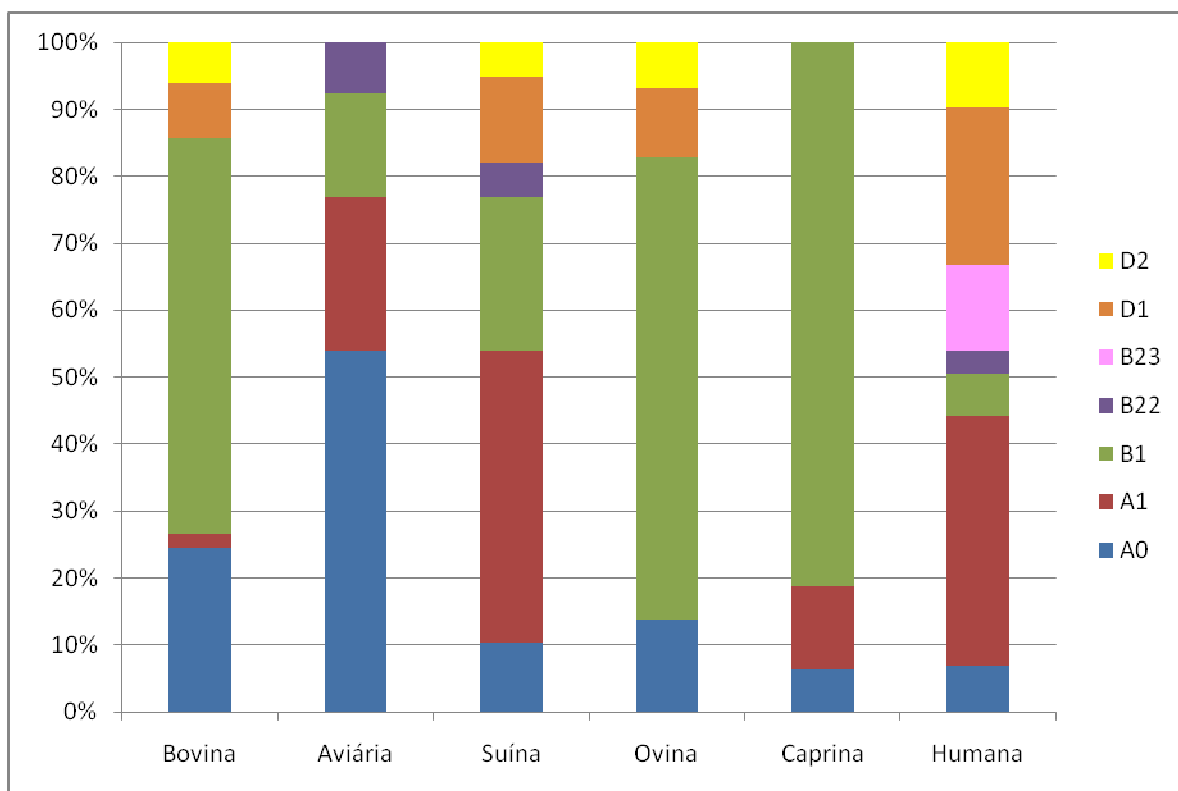


Figura 2. Frequência dos filo-tipos nos hospedeiros analisados.

Escobár-Páramo *et al.* (2006) observaram a prevalência dos grupos D/B1 em aves, A/B1 em mamíferos não-humanos e A/B2 em humanos. Nesse trabalho, os autores concluíram que a principal força que modela a estrutura genética das populações de *E. coli* nos diferentes hospedeiros é a domesticação. Gordon & Cowling (2003) encontraram uma prevalência do filo-grupo B2 nos mamíferos herbívoros e onívoros e uma prevalência do grupo B1 nos mamíferos carnívoros e nas aves. Unno *et al.* (2009) analisaram 138 linhagens de *E. coli* isoladas de

humanos e 376 de animais domésticos de uma província da Coréia do Sul e não encontraram o grupo B2 em humanos e animais domesticados. Assim sendo, a distribuição observada neste trabalho dos grupos filogenéticos entre os hospedeiros analisados difere das distribuições já descritas na literatura, o que suporta a hipótese que a distribuição dos grupos filogenéticos de *E. coli* entre os animais seja dependente de efeitos geográficos. Gordon & Cowling (2003) ainda sugerem que os grupos A e B1 são mais “generalistas” e que os grupos B2 e D são mais “especialistas”, o que está de acordo com os dados observados neste trabalho, já que os grupos A e B1 foram encontrados em todos os hospedeiros analisados e o grupo B2 foi mais predominante em humanos. Portanto, pode-se sugerir a utilização do filo-tipo B2₃ como indicador de contaminação fecal de origem humana no Estado de São Paulo.

Tabela 3. Distribuição dos filo-tipos de *E. coli* entre os hospedeiros analisados.

Filo-tipo	Humanos	Bois	Galinhas	Porcos	Ovelhas	Cabras
A ₀	12	12	10	4	4	1
A ₁	65	2	3	17	0	2
B1	11	29	2	9	20	13
B2 ₂	6	0	1	2	0	0
B2 ₃	22	0	0	0	0	0
D ₁	41	4	0	5	3	0
D ₂	17	3	0	2	2	0
Total	174	50	16	39	29	16

As amostras de humanos e suínos apresentaram índices de diversidade de Shannon e de Simpson semelhantes e apresentaram um índice de similaridade de Pianka de 87,8% (Tabela 4 e 5). A amostra de caprinos foi a que apresentou os menores índices de diversidades e os mamíferos herbívoros (bois, cabras e ovelhas) apresentaram uma similaridade média, quanto à distribuição de filo-tipos, de 96%. Bois, cabras e ovelhas são mamíferos ruminantes que diferem

em muitas características intestinais dos outros animais. Humanos e porcos apresentam características intestinais muito semelhantes, pois são ovívoros e monogástricos (Apajalahti, 2005). Os resultados parecem indicar que essas características do sistema digestório e de hábitos alimentares são fatores importantes na modelagem da estrutura populacional de *E. coli* nos hospedeiros analisados.

Tabela 4. Índices de diversidade de Shannon e de Simpson de cada hospedeiro analisado.

Índice de diversidade	Humanos	Bois	Galinhas	Porcos	Ovelhas	Cabras
Shannon	0,726	0,482	0,452	0,652	0,412	0,261
Simpson	0,769	0,579	0,555	0,724	0,489	0,320

Tabela 5. Índice de similaridade de Pianka entre os hospedeiros analisados.

	Aviária	Suína	Ovina	Caprina	Humana
Bovina	0,533	0,541	0,981	0,932	0,279
Aviária	-	0,508	0,362	0,298	0,384
Suína	-	-	0,507	0,574	0,878
Ovina	-	-	-	0,966	0,246
Caprina	-	-	-	-	0,258

Os dados obtidos neste trabalho enfatizam a importância de se analisar linhagens de *E. coli* isoladas no Estado de São Paulo para estabelecer parâmetros para um monitoramento mais efetivo dos recursos hídricos no Estado. Os resultados obtidos sugerem que as linhagens de *E. coli* isoladas de fezes de humanos apresentam maior diversidade, o que pode ser devido aos hábitos alimentares e estilo de vida dos humanos. Já os animais de criação estão submetidos a condições mais controladas de alimentação e estilo de vida, o que pode levar a uma menor variabilidade de linhagens de *E. coli*.

A Tabela 6 e a figura 3 mostra a distribuição dos filo-tipos entre os pontos de amostragem de linhagens ambientais. Duas linhagens do ponto Ani-1 não puderam ser classificadas em um dos grupos filogenéticos principais, pois apresentaram a seguinte combinação de marcadores: *chuA*-, *yjaA*+, Tspe4.C2+. Apenas três das 110 linhagens de origem ambiental pertencem ao grupo B1, sendo que uma foi isolada do rio Tietê, uma da ETE Jesus Netto e uma da ETE Barueri. Walk *et al.* (2007) demonstraram que linhagens de *E. coli* isoladas do ambiente formam um grupo distinto e bem estruturado. Segundo esses autores, a maior parte das linhagens ambientais pertence ao grupo B1. Neste trabalho poucas linhagens ambientais foram classificadas como pertencentes ao grupo B1 (Tabela 6), o que pode indicar que no Brasil o grupo A pode ter ocupado esse nicho.

Tabela 6. Distribuição dos filo-tipos de *E. coli* nos esgotos, rios e reservatórios analisados.

Filo-tipo	Ani-1	Ani-2	Ani-3	Hum-1	Hum-2	Hum-3	Hum-4	Esg-1	Esg-2
A ₀	1	4	8	2	1	0	0	3	6
A ₁	2	2	0	4	0	1	6	9	5
B1	1	0	0	0	0	0	0	1	1
B2 ₂	0	2	0	0	1	0	0	0	0
B2 ₃	2	1	0	2	9	8	2	2	3
D ₁	0	3	1	0	0	0	0	5	1
D ₂	0	0	1	0	0	0	0	4	4
Total	6	12	10	8	11	9	8	24	20

A maioria das linhagens ambientais pertence ao grupo A (54) e, ao contrário do que foi observado nas linhagens de animais, cerca de 26% das linhagens de origem ambiental pertencem ao grupo B2₃, sendo que a maioria delas foram isoladas das represas Billings e Guarapiranga. Orsi *et al.* (2007) verificaram a distribuição dos grupos filogenéticos em 131 linhagens isoladas

dessas represas e encontraram uma predominância de apenas 6% do grupo B2. Esses autores sugerem que a distribuição desses grupos pode ser alterada na estação chuvosa ou seca. As linhagens das represas Billings e Guarapiranga utilizadas neste trabalho foram isoladas no mês de julho de 2009, estação seca, o que pode indicar uma menor influência de fontes de contaminação difusa, predominando assim a fonte de contaminação pontual principal, provavelmente humana.

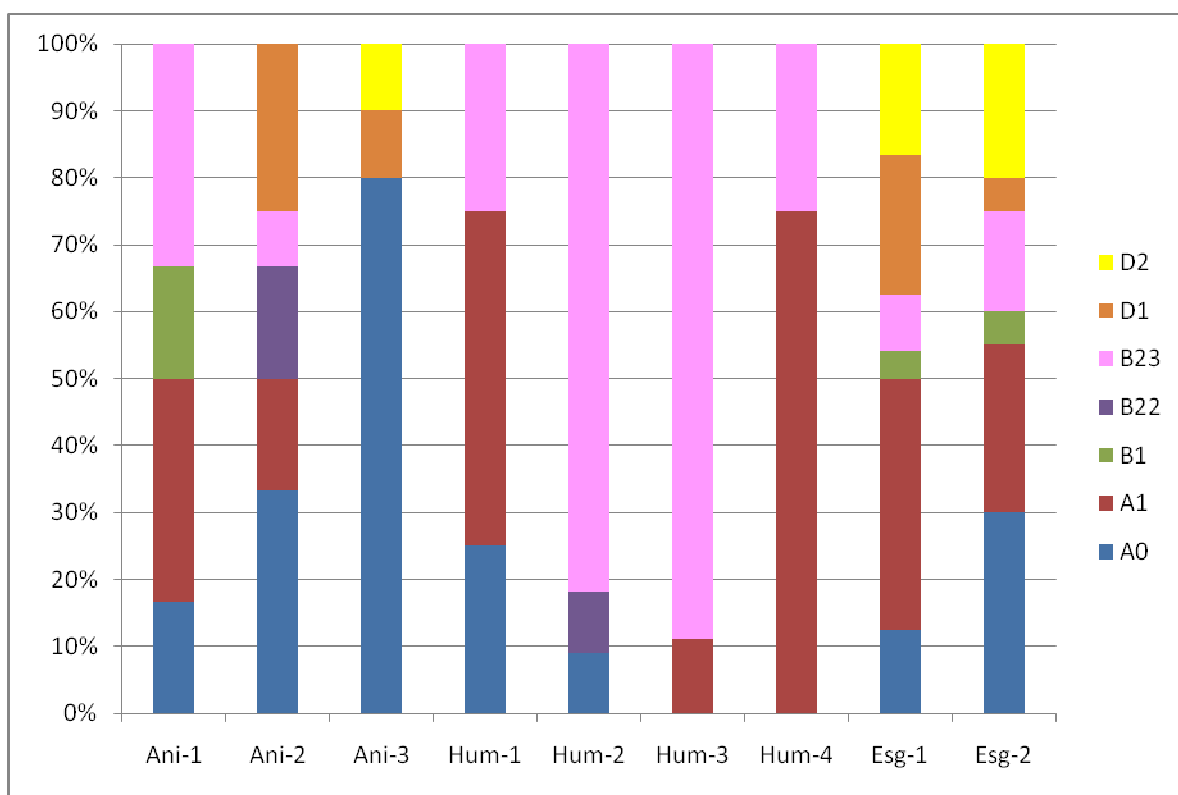


Figura 3. Frequência dos filo-tipos nos pontos de amostragem.

Para avaliar se a distribuição dos grupos filogenéticos poderia prever a origem animal de um grupo de linhagens foi construído um modelo de análise de correspondência com os dados da Tabela 3. A análise de correspondência (CA) é uma técnica descritiva, baseada em valores de

Qui-quadrado, que permite a exploração da estrutura dos dados. A Figura 4 mostra a representação bidimensional desta análise. Neste modelo, o eixo horizontal (*Dimension 1*) é responsável por 73,49% da variância total dos dados e o eixo vertical (*Dimension 2*) é responsável por 20,35% da variância total dos dados. Animais que apresentam distribuição de filo-tipos semelhante aparecem mais próximos entre si. Na Figura 4, pode-se observar que os mamíferos herbívoros apresentam uma distribuição dos filo-tipos mais semelhante entre si, assim como os mamíferos onívoros são mais semelhantes entre si.

Os dados da Tabela 6 foram inseridos no modelo da Figura 4 como pontos suplementares. Estes pontos não participam da construção do modelo. O posicionamento espacial destes pontos na representação bi-dimensional permite a comparação da distribuição dos filo-tipos com os dados utilizados na construção do modelo.

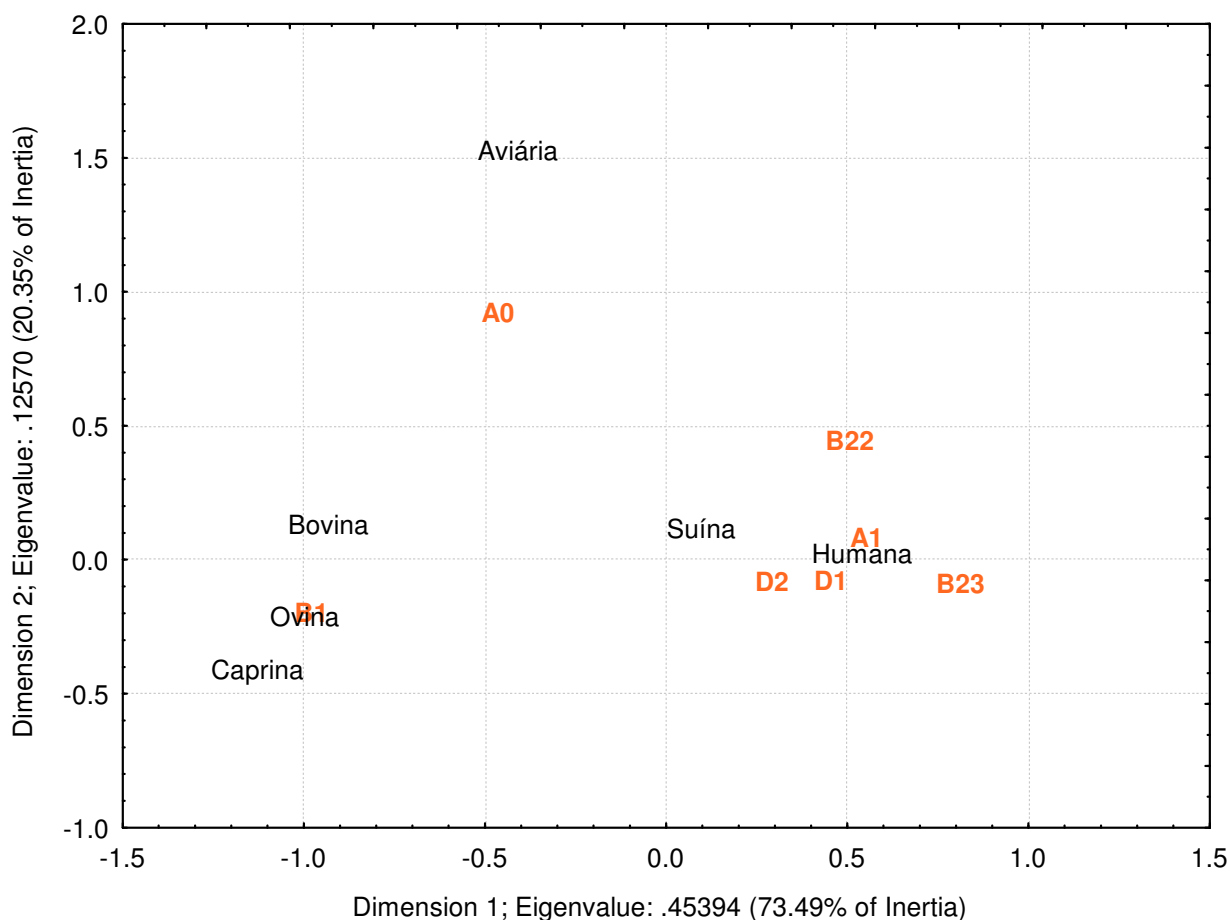


Figura 4. Análise de correspondência: representação bidimensional da frequência de filo-tipos nos hospedeiros. Os filo-tipos estão em laranja.

A Figura 5 mostra a comparação da distribuição de filo-tipos das amostras ambientais com as amostras de origem animal conhecida. As amostras Hum-1, Hum-2, Hum-3 e Hum-4 ficaram mais próximas da amostra de humanos do que de qualquer outro animal, assim como a amostra Esg-1. Esg-2, Ani-1 e Ani-2 ficaram entre a amostra de origem humana e a de origem suína e Ani-3 ficou próxima da amostra de origem aviária. A contaminação predita pelo modelo CA (Figura 5) foi a mesma que a esperada na maioria dos pontos (Hum-1, Hum-2, Hum-3, Hum-4, Esg-1 e Ani-3).

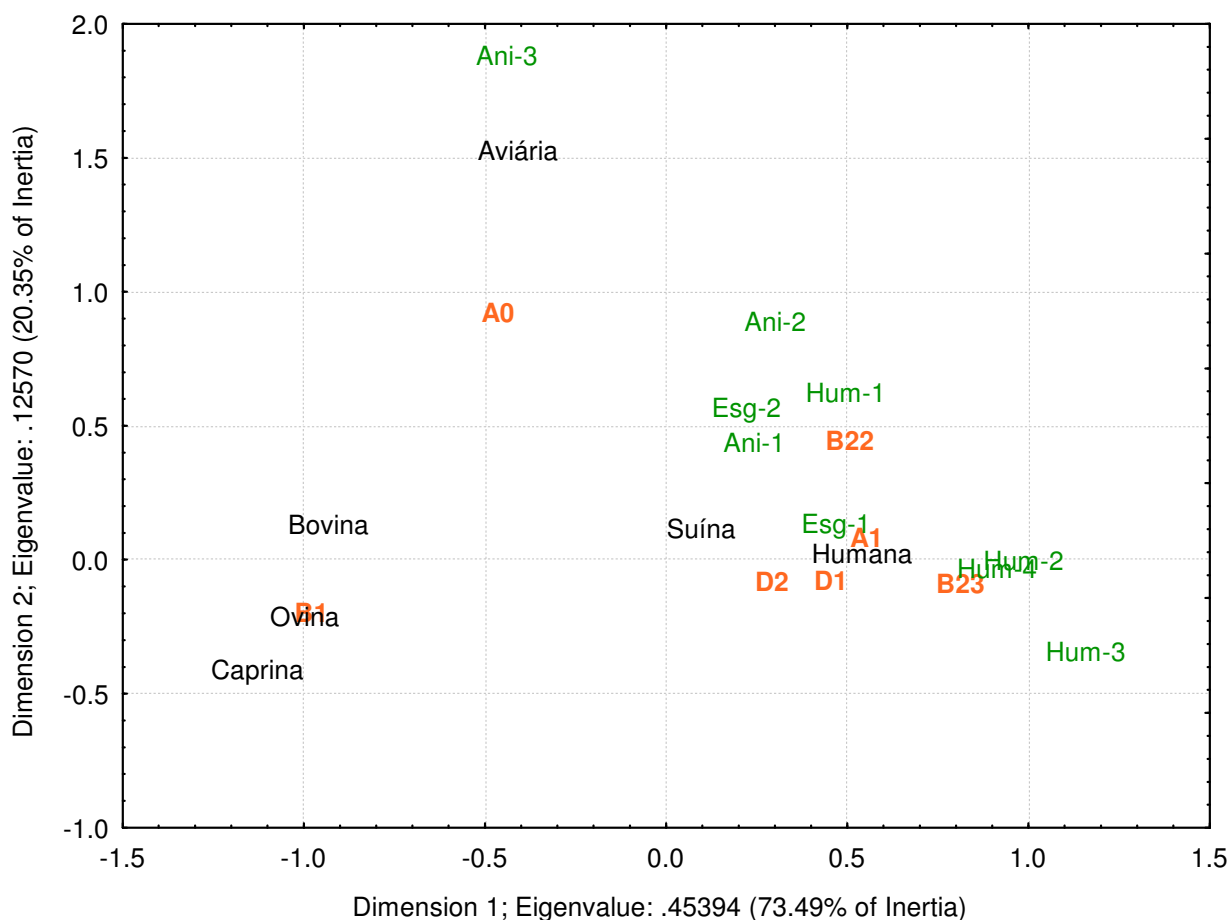


Figura 5. Análise de correspondência: representação bidimensional da frequência de filo-tipos nos hospedeiros. Os filo-tipos estão em laranja e os pontos suplementares estão em verde.

Assim sendo, a análise da distribuição de filo-tipos de *E. coli* parece ser uma boa ferramenta para auxiliar na identificação de fontes de contaminação fecal. Entretanto, não existem métodos para avaliar a significância estatística da análise de correspondência (Statsoft, 2007).

É muito difícil identificar marcadores hospedeiro-específicos em *E. coli*, o que dificulta os estudos de rastreamento de fonte microbiana (Gordon, 2001; Clermont *et al.*, 2008). Neste trabalho é proposta uma nova abordagem para a identificação da origem animal da fonte de

contaminação fecal, a qual se baseia na especificidade da estrutura populacional de *E. coli* ao invés da especificidade de linhagens. Essa abordagem é simples, barata e, apesar de não apresentar um alto poder de resolução, pode ajudar no direcionamento para utilização de técnicas mais laboriosas e com maior nível de discriminação, como por exemplo, o rep-PCR e a espectroscopia no infravermelho com transformada de Fourier.

REFERÊNCIAS

- Apajalahti J (2005) Comparative gut microflora, metabolic challenges, and potential opportunities. *J Appl Poult Res* 14:444-453.
- Ayres M, Ayres JRM, Ayres DL, Santos AS (2005) *BioEstat 4.0: Aplicações estatísticas nas áreas das ciências biológicas e médicas*. Belém: Sociedade Civil Mamirauá, CNPq.
- Baldy-Chudzik K, Mackiewicz P, Stosik M (2008) Phylogenetic background, virulence gene profiles, and genomic diversity in commensal *Escherichia coli* isolated from ten mammal species living in one zoo. *Vet Microbiol* 131:173-184.
- Bingen E, Picard B, Brahimi N, Mathy S, Desjardins P, Elion J, Denamur E (1998) Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strain. *J Infect Dis* 177:642-650.

Bush AO, Lafferty KD, Lotz JM, Shostak AW (1997) Parasitology meets ecology on its own terms: Margolis *et al.* revised. J Parasitol 83:575-583.

Clermont O, Bonacorsi S, Bingen E (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol 66:4555-4558.

Clermont O, Lescat M, O'Brien CL, Gordon DM, Tenaillon O, Denamur E (2008) Evidence for a human-specific *Escherichia coli* clone. Environ Microbiol 10:1000-1006.

Escobar-Páramo P, Le Menac'h A, Le Gall T, Amorin C, Gouriou S, Picard B, Skurnik D, Denamur E (2006) Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. Environ Microbiol 8:1975-1984.

Gomes TAT, Ramos SRTS, Rodrigues D, Rassi V, Toledo MRF, Vieira MAM, Gatti SV, Holck P, Griffin PM (1994) Etiology of acute diarrhea in children 1 to 5 years old in São Paulo, Brazil, abstr. C-299, p. 543. In Abstracts of the 94th General Meeting of the American Society for Microbiology, 1994. American Society for Microbiology, Washington, D.C.

Gordon DM (2001) Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiology 147:1079-1085.

Gordon DM, Cowling A (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. Microbiology 149:3575-3586.

- Herzer PJ, Inouye S, Inouye M, Whittan TS (1990) Phylogenetic distribution of branched RNS-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J Bacteriol 172:6175-6181.
- 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.
- Johnson JR, Delavari P, Kuskowski M, Stell AL (2001) Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. J Infect Dis 183:78–88.
- Ochman H, Selander RK (1984) Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol 157:690–693.
- Orsi RH, Stoppe NC, Sato MI, Gomes TA, Prado PI, Manfio GP, Ottoboni LMM (2007) Genetic variability and pathogenicity potential of *Escherichia coli* isolated from recreational water reservoirs. Res Microbiol 158:420-7.
- Orsi RH, Stoppe NC, Sato MIZ, Prado PI, Ottoboni LMM (2008) Phylogenetic group distribution among *Escherichia coli* isolated from rivers in São Paulo State, Brazil. World J Microbiol Biotechnol 24:1573-1577.
- Pianka ER (1973) The structure of lizard communities. Annu Rev Ecol Syst 4:53-74.

- 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* 65:2685-2692.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a laboratory manual*. 2nd ed. N.Y., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press.
- Selander RK, Caugant DA, Ochman H, Musser M, Gilmour MN, Whittam TS (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* 51:873-884.
- Silveira WD, Ferreira A, Brocchi M, Hollanda LM, Castro AFP, Yamada AT, Lancelloti M (2002) Biological characteristics and pathogenicity of avian *Escherichia coli* strains. *Vet Microbiol* 85:47-83.
- StatSoft Inc (2007) *Electronic Statistics Textbook*. Tulsa, OK: StatSoft. WEB: <http://www.statsoft.com/textbook/stathome.html>.

United States Environmental Protection Agency (2002) Method 1603 *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* Agar (Modified m-TEC), EPA 821-R-02-023, U.S Environmental Protection Agency, Office of Research and Development: Cincinnati, OH.

Unno T, Han D, Jang J, Lee SN, Ko G, Choi HY, Kim JH, Sadowsky MJ, Hur HG (2009) Absence of *Escherichia coli* phylogenetic group B2 strains in humans and domesticated animals from Jeonnam Province, Republic of Korea. *Appl Environ Microbiol* 75:5659-66.

Vicente HIG, Amaral LA, Cerqueira AMF (2005) Shigatoxigenic *Escherichia coli* serogroup O157, O111 and O113 in feces, water and milk samples from dairy farms. *Braz J Microbiol* 36:217-222.

Walk ST, Alm EW, Calhoun LM, Mladonicky JM, Whittam TS (2007) Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ Microbiol* 9:2274–2288.

CAPÍTULO 2

Utilização de BOX e (GTG)₅-PCR na discriminação de *Escherichia coli* de diferentes animais

INTRODUÇÃO

Entre os métodos de rastreamento de fonte microbiana, o rep-PCR é considerado relativamente rápido e pouco dispendioso (Meays *et al.*, 2004). Este método, que engloba os métodos de REP-, ERIC- e BOX-PCR, baseia-se na amplificação de sequências repetitivas do genoma. Os elementos repetitivos REP (*Repetitive Extragenic Palindromes*) foram caracterizados inicialmente em *E. coli* e *Salmonella typhimurium* (Stern *et al.*, 1984). O número de cópias desses elementos no cromossomo de *E. coli* e *S. typhimurium* é de aproximadamente 500-1000 cópias, equivalente a aproximadamente 1% do genoma da bactéria (Stern *et al.*, 1984; Martin *et al.*, 1992). Os elementos ERIC (*Enterobacterial Repetitive Intergenic Consensus*) foram descritos em várias espécies de bactérias (Hulton *et al.*, 1991). O número de cópias desse elemento em *E. coli* é de aproximadamente 30-50 e, em *S. typhimurium*, de 150 (Sharples & Lloyd, 1990; Hulton *et al.*, 1991). Os elementos BOX foram descritos em *Streptococcus pneumoniae* (Martin *et al.*, 1992). Eles são compostos de 3 subunidades (BOX A, BOX B e BOX C), com número variável de repetições (Martin *et al.*, 1992). O papel destes elementos repetitivos no genoma das bactérias ainda não está claro. Foi sugerido que eles teriam importância na determinação da estrutura e topologia do DNA, além de poderem estar envolvidos em regulação (van Belkum *et al.*, 1998).

Os elementos repetitivos têm sido utilizados em análises de variabilidade de diversas espécies de bactérias. Com a utilização de rep-PCR, Sadowsky *et al.* (1996) obtiveram padrões distintos de amplificação de DNA para a maioria das linhagens de *Streptomyces* analisadas. Dombek *et al.* (2000) analisaram 125 linhagens de *E. coli* isoladas de fezes humanas e de diversos animais através de BOX- e REP-PCR observando padrões complexos e distintos de bandas para as linhagens analisadas. Orsi *et al.* (2007) identificaram as fontes de contaminação

fecal de duas represas do Estado de São Paulo por rep-PCR. Mohapatra *et al.* (2007a) diferenciaram, por rep-PCR, linhagens de *E. coli* provenientes de 11 hospedeiros diferentes, incluindo humanos, animais selvagens e domésticos. O oligonucleotídeo (GTG)₅ também vem sendo utilizado com sucesso no estudo genotípico de linhagens bacterianas (Svec *et al.*, 2005; Mohapatra *et al.*, 2007b; Mohapatra *et al.*, 2008).

MATERIAIS E MÉTODOS

Linhagens utilizadas

Neste trabalho foram utilizadas linhagens do banco descrito no capítulo 1. Foram analisadas 100 linhagens de origem humana (crianças), 50 de origem bovina, 39 de origem suína, 16 de origem aviária, 29 de origem ovina e 16 de origem caprina. Além disso, foi realizado o BOX-PCR de 32 linhagens de humanos adultos e 27 das estações ETE Jesus Netto e Barueri (Tabela 1).

Tabela 1. Origem e número de linhagens de *E. coli* utilizadas neste trabalho.

Origem	Número de linhagens
Humana (construção do banco)	100
Humana (validação)	32
Bovina	50
Aviária	16
Suína	39
Caprina	16
Ovina	29
Esgoto	27
Total	309

Análise genotípica utilizando BOX- e (GTG)₅-PCR

Aproximadamente 10 µL da cultura permanente das linhagens do banco foram inoculados em 3 mL de meio LB líquido (Sambrook *et al.*, 1989) e mantidas em um agitador orbital a 250 rpm, 37°C por 18-20 horas. O DNA foi isolado com o Wizard Genomic DNA Purification Kit (Promega), segundo especificações do fabricante. O DNA foi utilizado nas análises de BOX- e (GTG)₅-PCR.

BOX-PCR: A análise do genoma das bactérias através de BOX-PCR foi feita segundo modificações na metodologia descrita por Sadowsky *et al.* (1996). Para isto, foram realizadas reações de amplificação em um volume final de 25 µL contendo: aproximadamente 10 ng de DNA, 50 pmoles do *primer* BOXA1R (5'-CTACGGCAAGGCGACGCTGAGG-3'), tampão da *Taq* polimerase 1X concentrado, 0,1 mM de cada dNTP, 2,0 mM de MgCl₂ e 0,5 unidades de *Taq* polimerase (Invitrogen). Todas as reações foram feitas em duplicata. As condições de amplificação foram: desnaturação inicial a 95°C por 7 min, seguida por 30 ciclos de amplificação (94°C – 1 min, 53°C – 1 min, 56°C – 4 min) e extensão final a 65°C por 16 min.

(GTG)₅-PCR: A análise através de (GTG)₅-PCR foi feita segundo modificações na metodologia descrita por Mohapatra *et al.* (2007b). O *primer* utilizado foi: (GTG)₅ (5'-GTGGTGGTGGTGGTG-3'). As reações de amplificação foram realizadas em um volume final de 20 µL contendo: aproximadamente 40 ng de DNA, 50 pmoles de *primer*, tampão da *Taq* polimerase 1X concentrado, 0,2 mM de cada dNTP, 2,5% de glicerol, 2,0 mM de MgCl₂ e 0,5 unidades de *Taq* polimerase (Invitrogen). Todas as reações foram feitas em duplicata. As condições de amplificação foram: desnaturação inicial a 95°C por 2 min, seguida de 30 ciclos de

amplificação (94°C – 3 s e 92°C – 30 s, 40°C – 1 min, 65°C – 8 min) e extensão final a 65°C por 8 min.

Os produtos de amplificação (BOX- e GTG₅-PCR) foram visualizados por eletroforese em gel de 2,0% de agarose em tampão TAE 1X (Sambrook *et al.*, 1989). Como marcador de peso molecular foi utilizado no gel o *Ladder* 1 Kb e 100 pb (Invitrogen). Após a eletroforese, o gel foi corado por 20 min com brometo de etídio (0,5 µg/mL), descorado por 1 h em água destilada e fotografado sob luz UV.

Análise de padrões de BOX e (GTG)₅-PCR

Os padrões de bandas das linhagens obtidos por BOX- e (GTG)₅-PCR foram analisados com o programa GelCompar II v. 5.1 (*Applied Maths*, Kortrijk, Bélgica). O programa identifica os padrões de bandas das diferentes linhagens e constrói dendrogramas baseados nested padrões.

Nas análises dos perfis de BOX-PCR foram utilizados o coeficiente de Jaccard, com tolerância de posição de 0,48% e otimização de 0,73%, o método de agrupamento UPGMA (Sokal & Michener, 1958). Para verificar a estabilidade interna dos grupos foi realizada a análise discriminante *Jackknife*, os valores são apresentados na forma de taxa de classificação correta (TCC) e taxa global de classificação correta (TGCC), que é a porcentagem total de linhagens corretamente classificadas.

Nas análises dos perfis de (GTG)₅-PCR foram utilizados o coeficiente de Jaccard, com tolerância de posição de 1,28% e otimização de 0%, o método de agrupamento UPGMA (Sokal & Michener, 1958). Para verificar a estabilidade interna dos grupos foi realizada a análise discriminante *Jackknife*, os valores são apresentados na forma de TCC e TGCC.

Os valores de tolerância de posição e otimização utilizados foram os que apresentaram a melhor separação das linhagens de acordo com a origem animal.

RESULTADOS E DISCUSSÃO

BOX-PCR

As linhagens de origem aviária, bovina, caprina, humana, ovina e suína foram submetidas à análise de perfil genotípico utilizando o *primer* BOXA1R (BOX-PCR). A Figura 1 exemplifica os perfis encontrados com esse *primer*.

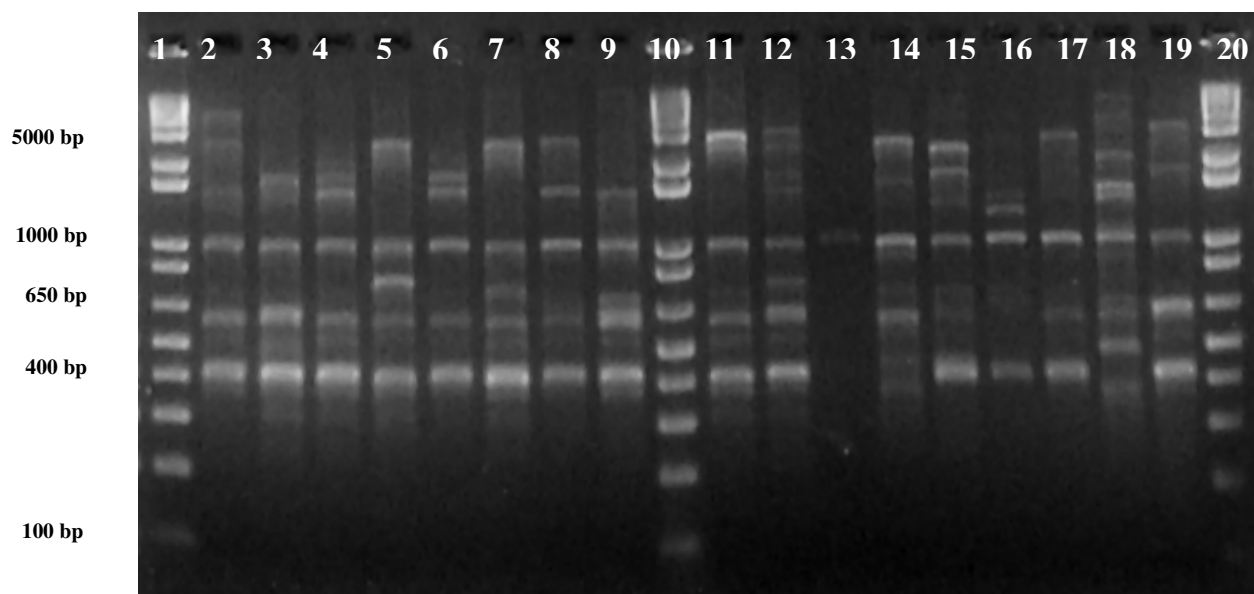


Figura 1. Padrão de bandas (perfil genotípico) obtido a partir de BOX-PCR. As amostras 1, 10 e 20 correspondem ao *Ladder* 1 Kb Plus (Invitrogen). As amostras de 2 a 9 correspondem às linhagens Bs18 a Bs25 e as amostras de 11 a 19 correspondem às linhagens Bs26 a Bs34.

A fim de se mensurar a reprodutibilidade do BOX-PCR, foram feitas triplicatas (três culturas independentes) de uma linhagem de humano, uma de boi e uma de galinha. Foi encontrada uma similaridade maior que 95% entre as triplicatas (coeficiente de Jaccard e tolerância de posição 0,48%), indicando que o método é reprodutível.

Os perfis obtidos por BOX-PCR foram submetidos a uma comparação utilizando o coeficiente de similaridade de Jaccard, que leva em consideração a presença/ausência e o método de agrupamento UPGMA. Foram excluídas das análises estatísticas as linhagens que não produziram bandas e as linhagens que apresentaram uma similaridade maior do que 95% com outra linhagem de mesma origem e mesmo filo-tipo, para evitar a super-estimação da taxa de classificação correta (Hassan *et al.*, 2005). Assim sendo, um total de 223 linhagens, sendo 11 de origem aviária, 43 de origem bovina, 15 de origem caprina, 94 de origem humana, 29 de origem ovina e 31 de origem suína foram utilizadas nas análises estatísticas.

Na análise discriminante *Jackknife* as linhagens foram separadas de quatro formas. A separação A foi feita entre linhagens de origem aviária, bovina, caprina, humana, ovina e suína. A separação B foi feita entre os grupos filogenéticos A, B1, B2 e D. A separação C foi feita entre linhagens de origem herbívora e onívora. E a separação D foi feita entre linhagens de origem humana e origem não humana. A Tabela 3 mostra os valores da taxa global de classificação correta para cada um dos casos. A separação D foi a que obteve os melhores valores de taxa classificação correta (TCC).

Segundo Hassan *et al.* (2005) quando a similaridade máxima é utilizada considera-se uma abordagem epidemiológica e quando a similaridade média é utilizada considera-se uma abordagem de genética de populações. Neste trabalho, como os perfis obtidos são muito

divergentes, pode-se concluir que o uso da similaridade máxima pode ser mais vantajoso na identificação da fonte animal.

Tabela 3. Taxa global de classificação correta pela análise discriminante *Jackknife* usando as similaridades média e máxima dos perfis obtidos por BOX-PCR.

Separação	Taxa global de classificação correta usando similaridade média	Taxa global de classificação correta usando similaridade máxima
A	57,11%	63,70%
B	39,62%	38,16%
C	68,77%	82,97%
D	75,86%	84,13%

A Tabela 4 mostra a TCC das linhagens de cada hospedeiro utilizando a similaridade média. Com esta abordagem, apenas 26,26% das linhagens de origem humana foram corretamente classificadas como humanas. As linhagens de origem aviária foram as que apresentaram a melhor TCC, 84,62%. Aparentemente, linhagens isoladas de um mesmo hospedeiro não formam um grupo muito coeso, pelo menos em relação ao perfil genotípico gerado por BOX-PCR.

Tabela 4. Taxa de classificação correta (TCC), em porcentagem, obtida a partir da análise discriminante *Jackknife* usando a similaridade média dos perfis obtidos por BOX-PCR.

Classificação	Humana	Bovina	Suína	Ovina	Caprina	Aviária
Humana	26,26	0	3,13	0	0	0
Bovina	24,24	65,96	6,25	13,79	13,33	7,69
Suína	14,14	8,51	71,88	24,14	13,33	7,69
Ovina	7,07	4,26	3,13	44,83	13,33	0
Caprina	4,04	2,13	6,25	6,9	53,33	0
Aviária	24,24	19,15	9,38	10,34	6,67	84,62

A Tabela 5 mostra a taxa de classificação de cada origem animal utilizando a similaridade máxima. Neste caso, 75,76% das linhagens de humanos foram corretamente classificadas como de origem humana e a TCC das linhagens de origem aviária caiu para 53,85%.

Tabela 5. Taxa de classificação correta (TCC), em porcentagem, obtida a partir da análise discriminante *Jackknife* usando a similaridade máxima dos perfis obtidos por BOX-PCR.

Classificação	Humana	Bovina	Suína	Ovina	Caprina	Aviária
Humana	75,76	6,38	9,38	3,45	6,67	15,38
Bovina	11,11	76,6	9,38	20,69	0	23,08
Suína	7,07	8,51	50	6,9	20	7,69
Ovina	1,01	2,13	21,88	58,62	13,33	0
Caprina	0	0	3,13	6,9	60	0
Aviária	5,05	6,38	6,25	3,45	0	53,85

Apesar de relativamente baixas, as TCCs obtidas neste trabalho com o BOX-PCR são similares a de outros trabalhos (Somarelli *et al.*, 2007; Hassan *et al.*, 2007). Até o momento, não existe um método padrão bem estabelecido para estudos de rastreamento de fonte microbiana, e segundo Harwood *et al.* (2000), qualquer método que apresentar uma TCC entre 60 e 70% já pode ser considerado útil para o desenvolvimento e implantação de planos de manejo para prever futuras contaminações fecais.

(GTG)₅-PCR

As linhagens de origem aviária, bovina, caprina, humana, ovina e suína foram submetidas à análise de perfil genotípico utilizando o *primer* (GTG)₅. A Figura 2 exemplifica os perfis encontrados com esse *primer*. Em geral, os perfis obtidos com o *primer* (GTG)₅ apresentam mais bandas do que os perfis gerados com o *primer* BOXA1R.

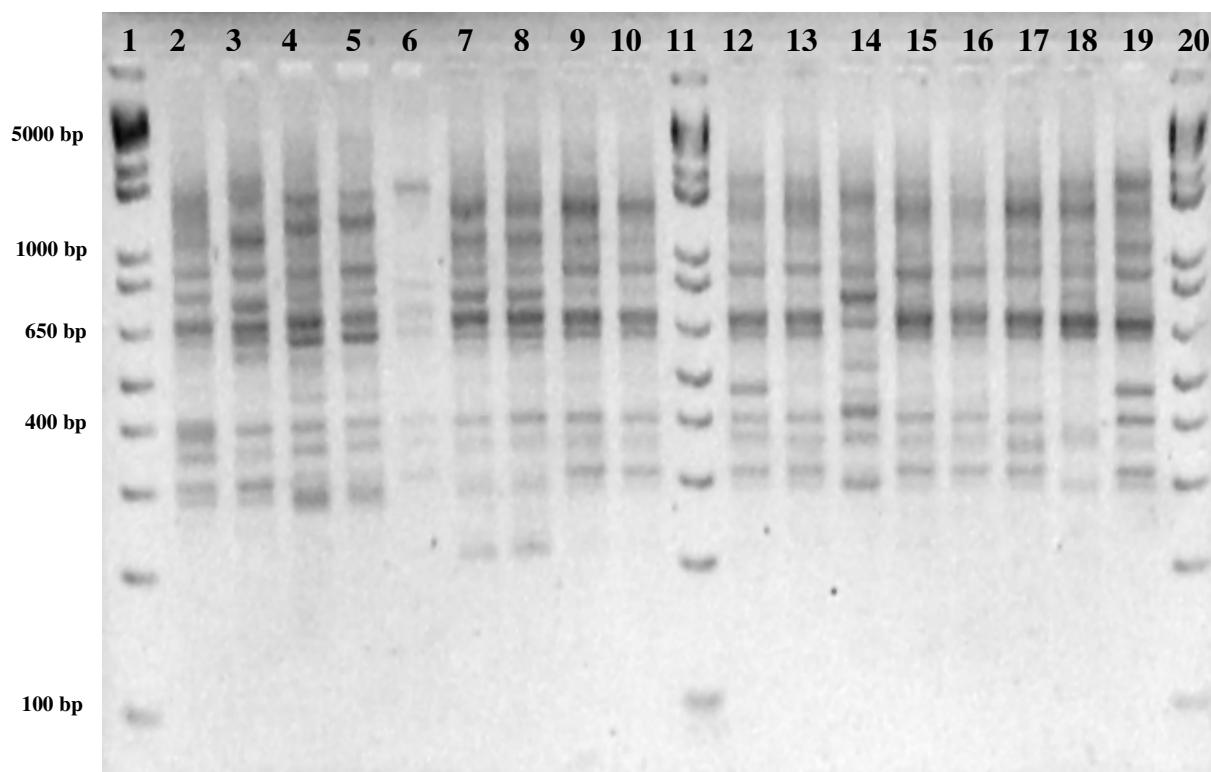


Figura 2. Padrão de bandas (perfil genotípico) obtido a partir de (GTG)₅-PCR. As amostras 1, 11 e 20 correspondem ao *Ladder* 1 Kb Plus (Invitrogen). As amostras de 2 a 10 correspondem às linhagens Hc81 a Hc89 e as amostras de 12 a 19 correspondem às linhagens Hc90 a Hc97.

A fim de se mensurar a reprodutibilidade do (GTG)₅-PCR, foram feitas triplicatas (três culturas independentes) de uma linhagem de humano, uma de boi e uma de galinha. Foi encontrada uma similaridade maior que 91% entre as triplicatas (coeficiente de Jaccard e tolerância de posição 1,28%) indicando que o método é reprodutível.

Os perfis obtidos por (GTG)₅-PCR foram submetidos a uma comparação utilizando o coeficiente de similaridade de Jaccard, que leva em consideração a presença/ausência e o método de agrupamento UPGMA (Sokal & Michener, 1958).

Assim como na análise de BOX-PCR, foram excluídas das análises as linhagens que não produziram bandas, e as linhagens que apresentaram uma similaridade maior do que 91% com outra linhagem da mesma origem e o mesmo filo-tipo, para evitar a super-estimação da taxa de classificação correta. Assim sendo, um total de 231 linhagens, sendo 16 de origem aviária, 43 de origem bovina, 15 de origem caprina, 95 de origem humana, 29 de origem ovina e 34 de origem suína foram utilizadas nas análises estatísticas.

Pode-se observar na Tabela 6 que os perfis obtidos por (GTG)₅-PCR foram menos eficientes em discriminar as linhagens de acordo com a origem animal (separação A) do que o BOX-PCR. Entretanto, as taxas globais de classificação correta em grupos filogenéticos (separação B) e em origem humana e não humana foram maiores do que as obtidas por BOX-PCR, tanto usando a similaridade média quanto usando a similaridade máxima para a classificação. Esses resultados parecem indicar que o (GTG)₅-PCR pode ser mais informativo sobre relações filogenéticas do que o BOX-PCR.

Tabela 6. Taxa global de classificação correta, em porcentagem, obtida pela análise discriminante *Jackknife* usando as similaridades média e máxima dos perfis obtidos por (GTG)₅-PCR.

Separação	Taxa global de classificação correta usando similaridade média	Taxa global de classificação correta usando similaridade máxima
A	45,56%	49,10%
B	51,60%	47,01%
C	72,83%	79,33%
D	84,87%	85,96%

A Tabela 7 mostra a taxa de classificação de cada origem animal utilizando a similaridade média, 68,42% das linhagens de humanos foram corretamente classificadas. Aparentemente, as linhagens de origem humana são mais semelhantes entre si em relação ao perfil de (GTG)₅-PCR. Nenhuma linhagem de origem bovina foi corretamente classificada, 39,5% delas foram classificadas como ovinas e 27,91% delas como caprinas. As linhagens de origem ovina foram as que apresentaram a maior TCC.

Tabela 7. Taxa de classificação correta (TCC), em porcentagem, obtida a partir da análise discriminante *Jackknife* usando a similaridade média dos perfis obtidos por (GTG)₅-PCR.

Classificação	Humana	Bovina	Suína	Ovina	Aviária	Caprina
Humana	68,42	4,65	0	0	0	0
Bovina	0	0	2,86	0	0	0
Suína	1,05	18,6	57,14	7,41	18,75	6,67
Ovina	25,26	39,53	14,29	77,78	25	60
Aviária	2,11	9,3	17,14	0	50	13,33
Caprina	3,16	27,91	8,57	14,81	6,25	20

A Tabela 8 mostra a taxa de classificação de cada origem animal utilizando a similaridade máxima. Neste caso, 46,51% das linhagens de origem bovina foram corretamente classificadas e apenas 26,67% das linhagens de cabras foram corretamente classificadas.

Tabela 8. Taxa de classificação correta (TCC), em porcentagem, obtida a partir da análise discriminante *Jackknife* usando a similaridade máxima dos perfis obtidos por (GTG)₅-PCR.

Classificação	Humana	Bovina	Suína	Ovina	Aviária	Caprina
Humana	80	16,28	0	11,11	6,25	0
Bovina	5,26	46,51	2,86	29,63	18,75	20
Suína	2,11	4,65	74,29	11,11	25	13,33
Ovina	10,53	18,6	11,43	29,63	12,5	40
Aviária	1,05	4,65	8,57	3,7	37,5	0
Caprina	1,05	9,3	2,86	14,81	0	26,67

Mohapatra *et al.* (2008) utilizaram (GTG)₅-PCR para diferenciar *E. coli* de oito fontes distintas e obtiveram uma taxa global de classificação correta de 83,67%. Mohapatra *et al.* (2007b) compararam os cinco métodos de PCR de regiões repetitivas, REP-, ERIC1-, ERIC2-, BOX- e (GTG)₅-PCR, e observaram uma maior eficiência dos dois últimos métodos. Apesar de ter sido utilizado neste trabalho um número de linhagens por hospedeiro muito semelhante ao do trabalho citado, a eficiência não foi semelhante. Isso pode ser devido à grande variabilidade das linhagens de *E. coli* de região para região.

As mesmas análises também foram feitas utilizando-se os perfis gerados pelo BOX- e (GTG)₅-PCR concomitantemente. Na Tabela 9 são apresentados os resultados principais. Utilizando-se os dois perfis, 92,54% das linhagens foram corretamente classificadas em origem humana e não humana. Entretanto, a discriminação entre os animais de origem foi menor do que

a obtida utilizando-se apenas o perfil do BOX-PCR (Tabela 3) e maior do que a obtida com o (GTG)₅-PCR (Tabela 9).

Tabela 9. Taxa global de classificação correta pela análise discriminante *Jackknife* utilizando similaridades média e máxima dos perfis obtidos por BOX- e (GTG)₅-PCR.

Separação	Taxa global de classificação correta usando similaridade média	Taxa global de classificação correta usando similaridade máxima
A	51,64%	57,61%
B	52,07%	47,15%
C	72,10%	81,29%
D	92,54%	89,48%

Sendo assim, o uso desses dois *primers* pode ser promissor para a discriminação de linhagens de *E. coli* de origem humana e não humana. Entretanto, a grande variabilidade de perfis de BOX- e (GTG)₅-PCR encontrada em todos os animais analisados pode dificultar a utilização destes métodos em estudos de rastreamento de fonte de contaminação fecal no Estado de São Paulo.

Validação externa do BOX-PCR

Como o BOX-PCR apresentou a melhor taxa de classificação correta para separar as linhagens em suas origens animais e apresentou maior reprodutibilidade do que o (GTG)₅-PCR, foi feita a validação externa deste método com linhagens não utilizadas na construção do banco. Foram utilizadas 32 linhagens de origem humana, denominadas HUM, e 27 linhagens isoladas de esgoto, denominadas AMB. A validação foi feita pela comparação dos perfis de BOX-PCR por

quatro ferramentas estatísticas: similaridade média, similaridade máxima, *K-nearest neighbor* e redes neurais (Tabela 10). O *score* máximo da ferramenta *K-nearest neighbor* é 10, enquanto que o *score* máximo das demais ferramentas é 100.

Das 32 linhagens de humanos usadas para validação, 13 não foram classificadas como de origem Humana por nenhuma ferramenta, dez foram classificadas como de origem humana por uma ferramenta, quatro linhagens foram classificadas como de origem humana por duas ferramentas, quatro foram classificadas como de origem humana por três ferramentas e uma linhagem foi classificada como de origem humana pelas quatro ferramentas (Tabela 10). A baixa taxa de classificação correta obtida por todas as ferramentas é mais um indício da grande variabilidade de perfil de BOX-PCR das linhagens de *E. coli* de humanos. As classificações pela similaridade máxima e por *K-nearest neighbor* foram as que apresentaram as maiores taxas de classificação correta, 37,5% e 40,7%, respectivamente.

As linhagens de esgoto foram submetidas à identificação através da comparação dos perfis de BOX-PCR dessas linhagens com os perfis das linhagens do banco, mais as 32 linhagens HUM, perfazendo um total de 126 linhagens de origem humana. Os resultados são apresentados na Tabela 11. Das 27 linhagens de esgoto analisadas, quatro não foram classificadas como de origem humana por nenhuma ferramenta, oito foram classificadas como de origem humana por uma das ferramentas, 11 foram classificadas como de origem humana por duas ferramentas e quatro foram classificadas como humana por três ferramentas.

Tabela 10. Classificação das linhagens de *E. coli* de humanos utilizando quatro ferramentas estatísticas.

	<i>K-Nearest Neighbor</i>		Rede neural		Similaridade média		Similaridade máxima	
Linhagem	Classificação	Score	Classificação	Score	Classificação	Score	Classificação	Score
HUM077	Humana	6	Humana	88,3	Bovina	20,2	Humana	44,4
HUM082	Suína	6	Aviária	35,8	Suína	19,7	Suína	41,7
HUM083	Humana	4	Aviária	55,7	Suína	19,8	Humana	44,4
HUM084	Bovina	6	Aviária	35,8	Caprina	19,4	Humana	40
HUM085	Bovina	4	Suína	27,1	Suína	26,9	Suína	66,7
HUM086	Suína	4	Aviária	57,6	Aviária	27,7	Humana	60
HUM087	Bovina	5	Caprina	54,3	Aviária	24,3	Suína	45,5
HUM088	Humana	5	Aviária	57,4	Bovina	31,9	Bovina	75
HUM092	Bovina	4	Humana	98,7	Bovina	30,3	Bovina	60
HUM099	Humana	7	Humana	79,8	Humana	17	Humana	50
HUM101	Bovina	5	Aviária	35,8	Aviária	23,6	Suína	45,5
HUM102	Humana	4	Aviária	32,9	Bovina	19,5	Bovina	44,4
HUM103	Humana	7	Humana	28,6	Bovina	31,1	Humana	75
HUM104	Bovina	5	Caprina	54,8	Aviária	31,8	Suína	75
HUM109	Humana	4	Aviária	57,7	Bovina	30	Humana	60
HUM111	Suína	4	Humana	82,2	Bovina	27	Humana	60
HUM112	Aviária	4	Aviária	57,7	Aviária	23,2	Bovina	40
HUM113	Humana	5	Humana	92,8	Suína	13,5	Humana	40
HUM115	Ovina	4	Humana	71,7	Ovina	20,5	Ovina	66,7
HUM116	Humana	9	Bovina	61,7	Humana	9,8	Humana	50
HUM117	Humana	3	Ovina	46,8	Aviária	22,4	Aviária	44,4
HUM118	Ovina	5	Ovina	52,4	Caprina	14,1	Ovina	27,3
HUM119	Bovina	4	Caprina	79,7	Ovina	18,4	Humana	66,7
HUM121	Aviária	3	Aviária	35,8	Aviária	22,9	Bovina	62,5
HUM122	Bovina	5	Caprina	66,5	Aviária	36	Bovina	100
HUM124	Suína	4	Aviária	33,7	Aviária	26,5	Bovina	60
HUM125	Suína	4	Ovina	46,9	Suína	13,9	Humana	42,9
HUM126	Humana	5	Aviária	35,8	Suína	15	Aviária	33,3
HUM127	Bovina	6	Caprina	71,4	Bovina	26,2	Bovina	60
HUM128	Suína	5	Ovina	46,8	Ovina	29,6	Suína	55,6
HUM129	Bovina	6	Aviária	35,5	Bovina	35,8	Aviária	66,7
HUM132	Humana	4	Suína	28,5	Caprina	16,3	Humana	66,7

A classificação das linhagens de esgoto pela similaridade média foi a que apresentou menor eficiência. Já as classificações pela similaridade máxima e por *K-nearest neighbor* apresentaram uma maior taxa de acerto em relação à fonte esperada. Apesar disto, os *scores* de classificação foram relativamente baixos, em média 5,5 por *K-nearest neighbor* e 61,5 por similaridade máxima.

Quando um agrupamento é realizado apenas com as 126 linhagens isoladas de humanos e com as 27 linhagens de esgoto, a similaridade mínima encontrada entre os perfis é de 3,22%, o que indica uma grande variabilidade. A análise *Jackknife* utilizando a similaridade máxima classificou apenas 15 linhagens de esgoto como de origem humana (55%), o que pode sugerir a existência de populações de *E. coli* que utilizam o esgoto como habitat primário.

Muitos estudos de RFM são feitos utilizando-se linhagens de esgoto na construção do banco de linhagens, pressupondo-se que sejam de origem humana (Haznedaroglu *et al.*, 2007; Duran *et al.*, 2009). Entretanto, nossos resultados demonstram que as linhagens de esgoto podem apresentar perfis de BOX-PCR muito distintos das linhagens isoladas diretamente de fezes humanas, podendo existir populações capazes de utilizar este ambiente como habitat primário (Walk *et al.*, 2007). Portanto, as linhagens isoladas de esgoto podem não ser bons indicadores de contaminação por fezes humanas.

Tabela 11. Classificação das linhagens de esgoto utilizando quatro ferramentas estatísticas e adicionando as linhagens HUM ao banco.

	<i>K-Nearest Neighbor</i>		Rede neural		Similaridade média		Similaridade máxima	
Linhagem	Classificação	Score	Classificação	Score	Classificação	Score	Classificação	Score
AMB001	Bovina	6	Humana	38,4	Bovina	34,2	Humana	66,7
AMB003	Humana	5	Ovina	32,8	Ovina	22,5	Humana	44,4
AMB004	Humana	7	Ovina	32,8	Suína	20,8	Humana	54,5
AMB005	Bovina	5	Bovina	36,1	Bovina	25,2	Suína	55,6
AMB006	Humana	5	Humana	38,4	Bovina	26,9	Humana	80
AMB007	Suína	4	Bovina	36,1	Suína	26,7	Suína	50
AMB009	Bovina	5	Humana	38,4	Bovina	26,1	Suína	50
AMB010	Humana	4	Bovina	36,1	Bovina	27,1	Ovina	50
AMB011	Bovina	5	Humana	38,4	Bovina	30,8	Suína	71,4
AMB012	Humana	6	Suína	52,3	Aviária	25,1	Humana	61,5
AMB013	Humana	5	Humana	84,2	Bovina	29,7	Humana	60
AMB014	Bovina	6	Ovina	30,6	Bovina	24,8	Suína	50
AMB015	Humana	9	Humana	38,4	Bovina	32,7	Humana	60
AMB016	Bovina	4	Humana	98	Bovina	23,7	Bovina	57,1
AMB017	Humana	5	Humana	80,4	Bovina	19,2	Bovina	36,4
AMB019	Suína	4	Bovina	36,1	Suína	8,3	Humana	33,3
AMB020	Bovina	4	Bovina	36,1	Suína	25,3	Humana	42,9
AMB022	Humana	4	Bovina	36,1	Suína	22,7	Humana	44,4
AMB025	Bovina	5	Bovina	42,1	Aviária	20,9	Aviária	45,5
AMB028	Humana	8	Bovina	36,1	Suína	23,7	Humana	80
AMB032	Humana	6	Ovina	32,4	Suína	25,5	Humana	58,3
AMB059	Humana	4	Humana	98,9	Suína	24,4	Bovina	60
AMB060	Ovina	3	Humana	80,4	Ovina	16,5	Ovina	50
AMB062	Humana	5	Bovina	36,1	Bovina	36,3	Humana	100
AMB063	Humana	6	Humana	78,5	Bovina	28	Humana	75
AMB065	Humana	7	Bovina	35,9	Bovina	24,7	Ovina	66,7
AMB066	Humana	2	Humana	38,4	Aviária	33,8	Ovina	57,1

REFERÊNCIAS

- 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.
- Duran M, Yurtsever D, Dunaev T (2009) Choice of indicator organism and library size considerations for phenotypic microbial source tracking by FAME profiling. *Water Sci Technol* 60:2659-68.
- Harwood VJ, Wiggins B, Hagedorn C, Ellender RD, Gooch J, Kern J, Samadpour M, Chapman ACH, Robinson BJ (2003) Phenotypic library-based microbial source tracking methods: Efficacy in the California collaborative study. *J Water Health* 1:153-166.
- Hassan WM, Ellender RD, Wang SY (2007) Fidelity of bacterial source tracking: *Escherichia coli* vs *Enterococcus* spp and minimizing assignment of isolates from nonlibrary sources. *J Appl Microbiol* 102:591-8.
- Haznedaroglu BZ, Yurtsever D, Lefkowitz JR, Duran M (2007) Phenotypic characterization of *Escherichia coli* through whole-cell fatty acid profiling to investigate host specificity. *Water Res* 41:803-809.

- Hulton CS, Higgins, CF, Sharp PM (1991) ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. Mol Microbiol 5:825-834.
- Martin B, Humbert O, Camara M, Guenzi E, Walker J, Mitchell T, Andrew P, Prudhomme M, Alloing G, Hakenbeck R, Morrison DA, Boulnois GJ, Claverys JP (1992) A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res 20:3479-3483.
- Meays CL, Broersma K, Nordin R, Mazumder A (2004) Source tracking fecal bacteria in water: a critical review of current methods. J Environ Manage 73:71-9.
- Mohapatra B, Broersma K, Nordin R, Mazumder A (2007a) Evaluation of repetitive extragenic palindromic-PCR for discrimination of fecal *Escherichia coli* from humans, and different domestic- and wild-animals. Microbiol Immunol 51:733-740.
- Mohapatra BR, Broersma K, Mazumder A (2007b) Comparison of five rep-PCR genomic fingerprinting methods for differentiation of fecal *Escherichia coli* from humans, poultry and wild birds. FEMS Microbiol Letters 277:98-106.
- Mohapatra BR, Broersma K, Mazumder A (2008) Differentiation of fecal *Escherichia coli* from poultry and free-living birds by (GTG)₅-PCR genomic fingerprint. Int J Med Microbiol 298:245-252.

- Sadowsky MJ, Kinkel LL, Bowers JH, Schottel JL (1996) Use of repetitive intergenic DNA sequences to classify pathogenic and disease-suppressive *Streptomyces* strains. *Appl Environ Microbiol* 62:3489-3493.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor.
- Sharples GJ, Lloyd, RG (1990) A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucleic Acids Res* 18:6503-6508.
- Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. *Kans Univ Sci Bull* 38:1409-1438.
- Somarelli JA, Makarewicz JC, Sia R, Simon R (2007) Wildlife identified as major source of *Escherichia coli* in agriculturally dominated watersheds by BOX A1R-derived genetic fingerprints. *J Environ Manage* 82:60-65.
- Stern MJ, Ames, GF, Smith NH, Robinson EC, Higgins CF (1984) Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* 37:1015-1026.
- Svec P, Vancanneyt M, Seman M, Snauwaert C, Lefebvre K, Sedlacek I, Swings J (2005) Evaluation of (GTG)₅-PCR for identification of *Enterococcus* spp. *FEMS Microbiol Letters* 247:59-63.

van Belkum A, Sherer S, Van Alphen L, Verbrugh H (1998) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* 62:275-293.

Walk ST, Alm EW, Calhoun LM, Mladonicky JM, Whittam TS (2007) Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ Microbiol* 9:2274–2288.

CAPÍTULO 3

**Utilização de espectroscopia no infravermelho com transformada de
Fourier (FT-IR) na discriminação de *Escherichia coli* de diferentes
animais**

INTRODUÇÃO

A espectroscopia no infravermelho por transformada de Fourier (FT-IR) tem sido utilizada com sucesso na identificação e classificação de microrganismos nos níveis de espécie e subespécies (Naumann *et al.*, 1991). Esta técnica mede as propriedades vibracionais das ligações químicas quando excitadas pela absorção da radiação infravermelha. Quando aplicada a células intactas, o espectro infravermelho reflete a composição bioquímica qualitativa das células (Mouwen *et al.*, 2005).

Mouwen *et al.* (2005) conseguiram bons resultados com FT-IR na discriminação de *Campylobacter coli* e *Campylobacter jejuni* quando comparados à diferenciação por ERIC-PCR. Gilbert *et al.* (2009) demonstraram que é possível discriminar linhagens de *E. coli*, apesar da grande semelhança bioquímica. A maioria dos estudos que aplicam a espectroscopia FT-IR em microbiologia utilizam a região média do infravermelho ($4000 - 600 \text{ cm}^{-1}$) (Naumann *et al.*, 1995; Yu & Irudayaraj, 2005; Mouwen *et al.*, 2005; Gilbert *et al.*, 2009). Entretanto, alguns estudos concluíram que a espectroscopia na região próxima do infravermelho ($10000-4000 \text{ cm}^{-1}$), pode ser uma ferramenta rápida, barata e não destrutiva na identificação de microrganismos (Rodriguez-Saona *et al.*, 2001; Rodriguez-Saona *et al.*, 2004; Alexandrakis *et al.*, 2008).

A espectroscopia FT-IR nunca foi aplicada na identificação da fonte animal de um microrganismo. Sendo assim, a objetivo deste trabalho foi discriminar linhagens de *E.coli* de origem aviária, bovina e humana por espectroscopia FT-IR.

MATERIAIS E MÉTODOS

Linhagens utilizadas

Foram utilizadas 20 linhagens de *E. coli* isoladas de humanos, 15 isoladas de bois e 16 isoladas de galinhas. Essas linhagens pertencem ao banco descrito no capítulo 1.

Espectroscopia FT-IR

Cinco μL da cultura permanente das linhagens foram inoculados em 2 mL de meio LB líquido (Sambrook *et al.*, 1989) e incubados por 18-20 h a 37°C, 250 rpm. A seguir, 5 μL da cultura diluída em água 1:1 foi plaqueada em meio LB sólido e as placas foram mantidas por 16 h a 37°C. De cinco a sete colônias foram transferidas para lâminas de vidro. Após secar a temperatura ambiente, as amostras foram analisadas em um espectrômetro SPOTLIGHT 400N (Perkin Elmer). Os parâmetros utilizados foram: modo imagem, transmitância, 64 scans por pixel, resolução 4 cm^{-1} , faixa espectral de 6000 a 2000 cm^{-1} , área de 100 x 100 micrometros. Os experimentos foram feitos em triplicada e um total de 192 espectros foi obtido de cada linhagem.

Tratamento dos dados, modelo e validação

Os espectros foram analisados com PLS toolbox 3.5 para MATLAB 7.0 (Eigenvector Research Inc.). Foi utilizada nas análises a média dos espectros das triplicadas. Para obtenção do

melhor modelo de PLS-DA vários pré-processamentos foram testados. O melhor modelo foi aquele no qual os espectros foram cortados entre 2600 e 3300 cm^{-1} . Para ajustar a linha de base foi utilizada a segunda derivada dos dados e para eliminar informações desnecessárias ao modelo foi utilizado o algoritmo OSC (*orthogonal signal correction*). Neste procedimento, a matriz **X** é corrigida pela subtração da variação que é ortogonal a um vetor **y**, o qual possui as classes correspondentes a cada amostra. Após a aplicação do OSC os espectros foram novamente cortados, agora entre 2816 e 3026 cm^{-1} .

A validação do método foi feita separando algumas amostras dos conjuntos originais (5 das amostras de bactérias provenientes de fezes de bovinos, 5 das provenientes das fezes de galinhas e 7 das provenientes das fezes de humanos) e construindo um novo modelo PLS-DA com as amostras restantes, utilizando os mesmos pré-processamentos do melhor modelo anterior. As amostras separadas foram colocadas então no modelo e através da separação na primeira variável latente foi possível ver que a previsão de amostras desconhecidas foi bastante satisfatória. As amostras de 1 a 34 são as que foram usadas na construção do modelo (amostras de calibração) e as amostras de 35 a 51 foram as amostras de validação, sendo, as amostras de 35 a 39 referentes a bactérias de fezes de bovinos, de 40 a 44 referentes a bactérias de fezes de galinhas e de 45 a 51 referentes a bactérias de fezes humanas.

Comparação com BOX-PCR

Um dendrograma foi construído, utilizando-se apenas as linhagens usadas na análise FT-IR com os perfis obtidos por BOX-PCR descritos no Capítulo 2, para comparação entre as duas técnicas.

RESULTADOS E DISCUSSÃO

Espectroscopia FT-IR

Na análise das 51 linhagens de *E. coli* por espectroscopia FT-IR, os espectros foram obtidos na região entre 2000 a 6000 cm^{-1} (Figura 1a). Análises preliminares indicaram que a região entre 2600 e 3300 cm^{-1} seria mais informativa para separar as linhagens de acordo com a origem animal (Figura 1b).

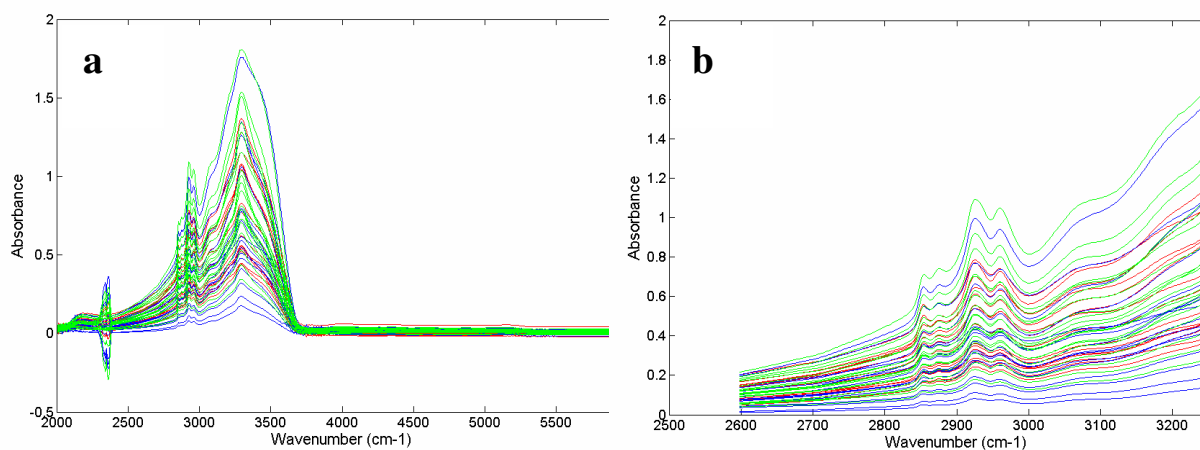


Figura 1. a) Espectros médios das bactérias entre 2000 e 6000 cm^{-1} e **b)** Espectros cortados entre 2600 e 3300 cm^{-1} . As linhas verdes representam os espectros das linhagens de *E. coli* de origem humana, as linhas vermelhas representam os espectros das linhagens de *E. coli* de origem aviária e, as linhas azuis representam os espectros das linhagens de *E. coli* de origem bovina.

O algoritmo OSC foi utilizado para eliminar informações desnecessárias e os espectros foram novamente cortados na região entre 2816 e 3026 cm^{-1} . Após o novo corte foi aplicada a segunda derivada nos espectros (Figura 2). As principais diferenças podem ser observadas entre os espectros das linhagens de humanos e de bois. Os espectros das linhagens de galinhas foram, aparentemente, mais variáveis.

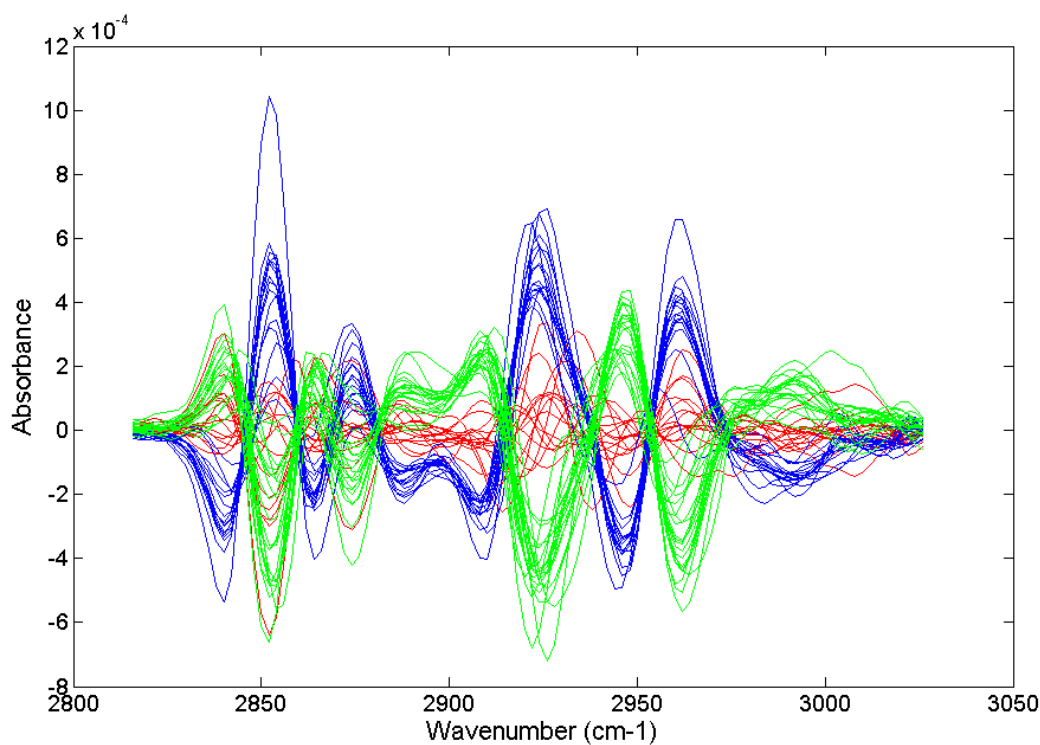


Figura 2. *Orthogonal signal correction* (OSC) da segunda derivada dos espectros entre 2816 e 3026 cm^{-1} . As linhas verdes representam os espectros das linhagens de *E. coli* de origem humana, as linhas vermelhas representam os espectros das linhagens de *E. coli* de origem aviária e, as linhas azuis representam os espectros das linhagens de *E. coli* de origem bovina.

Na Figura 3 estão duas variáveis latentes (VL1 e VL2) que juntas explicam 83,45% da variância total dos dados. A primeira variável (VL1) é responsável por separar as linhagens de acordo com a origem animal e a segunda variável (VL2) é responsável pela variância dentro de cada grupo.

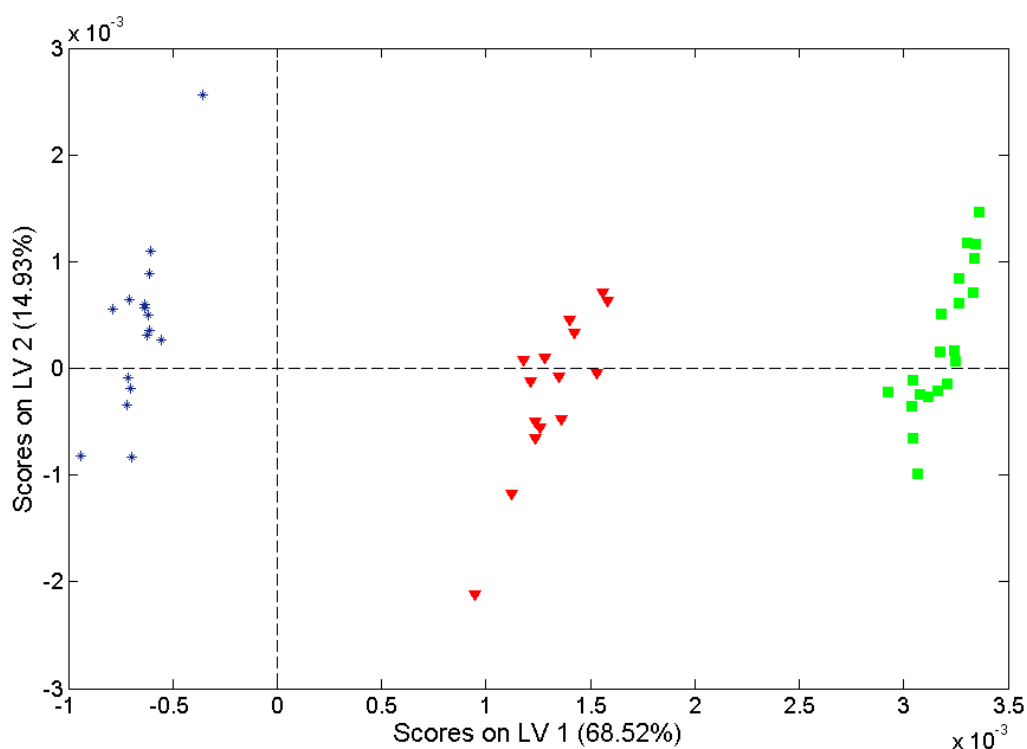


Figura 3. Representação bi-dimensional mostrando a separação obtida pelo modelo PLS-DA utilizando a segunda derivada dos espectros na região entre 2816 e 3026 cm^{-1} . Os quadrados verdes representam as linhagens de *E. coli* de origem humana, os triângulos vermelhos representam as linhagens de *E. coli* de origem aviária e, os asteriscos azuis representam os espectros as *E. coli* de origem bovina. A primeira variável latente (LV1) é responsável por

68,52% da variância total dos dados e a segunda variável latente (LV2) é responsável por 14,93% da variância total dos dados.

Uma aprendizagem supervisionada foi realizada para a validação do modelo PLS-DA. Pode-se observar na Figura 4 que todas as amostras de validação tiveram os mesmos escores na primeira variável latente que seus respectivos conjuntos de validação, onde as amostras referentes a bactérias de fezes de bovinos ficaram com os escores entre $0,5$ e $-0,5 \times 10^{-3}$, as amostras referentes a bactérias de fezes de galinhas ficaram com os escores entre $-1,5$ e $-2,5 \times 10^{-3}$ e as amostras referentes a bactérias de fezes de humanos ficaram com os escores entre $1,25$ e 2×10^{-3} .

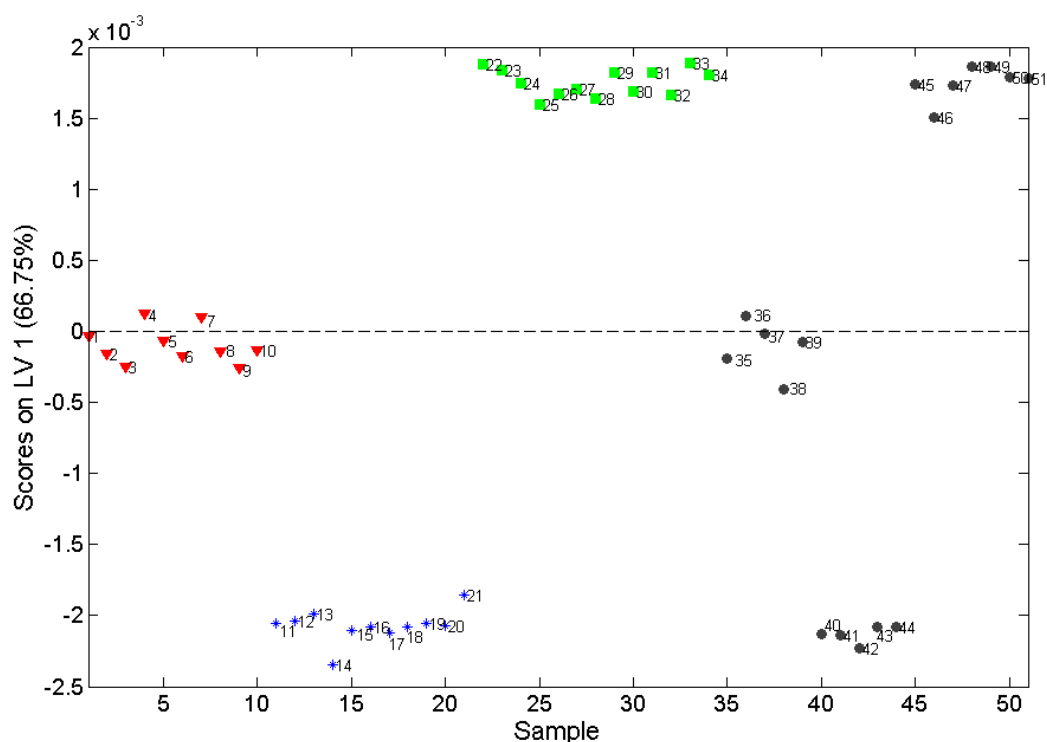


Figura 4. Representação bi-dimensional do modelo de PLS-DA e validação (LV1 versus número das amostras). Os quadrados verdes representam as linhagens de *E. coli* de origem humana, os triângulos vermelhos representam as linhagens de *E. coli* de origem aviária e, os asteriscos azuis representam os espectros de *E. coli* de origem bovina. Os círculos pretos são as amostras de validação. As amostras de 1 a 34 são as que foram usadas na construção do modelo (amostras de calibração) e as amostras de 35 a 51 foram as amostras de validação, sendo, as amostras de 35 a 39 referentes a bactérias de fezes de galinhas, de 40 a 44 referentes a bactérias de fezes de bovinos e de 45 a 51 referentes a bactérias de fezes humanas. A primeira variável latente (LV1) é responsável por 66,75% da variância total dos dados.

Os perfis bioquímicos obtidos pela espectroscopia FT-IR foram submetidos a análise multivariada, PLS-DA. O melhor resultado foi obtido construindo o modelo PLS-DA com a segunda derivada do espectro na região entre 2816 a 3026 cm^{-1} , o qual permitiu a separação completa das linhagens de *E. coli* de origem aviária, bovina e humana. As variáveis mais importantes na VL1 correspondem às bandas de FT-IR 2852, 2924, 2946 e 2962 cm^{-1} (Tabela 2). A região do espectro que permitiu a discriminação das linhagens de *E. coli* de acordo com sua origem animal é descrita como a região de vibrações de estiramento dos grupos funcionais C-H₃ e C-H₂ com características espectrais de cadeias de ácidos graxos e outros anfifílicos de membrana (Naumann *et al.*, 1991). Haznedaroglu *et al.* (2007), utilizando perfil de ácidos graxos (FAME), encontraram alguns marcadores específicos para diferentes origens de *E. coli* (bovina, esgoto, suína, aviária, córregos e animais selvagens). Esses ácidos graxos podem desempenhar um papel importante nas relações microrganismo-hospedeiro. Sendo assim, as bandas nessa região do espectro são boas candidatas a marcadores fenotípicos de *E. coli* de diferentes hospedeiros.

Tabela 2. Correspondência, segundo Chalmers e Griffiths (2002), das bandas mais importantes na construção do modelo PLS-DA.

Número de onda (cm^{-1})	Correspondência
2852	C-H estiramento simétrico de CH ₂ em ácidos graxos
2924, 2946	C-H estiramento assimétrico de CH ₂
2960	C-H estiramento assimétrico de CH ₃ em ácidos graxos

BOX-PCR

Foram analisadas 51 linhagens de *E. coli* por BOX-PCR. Três linhagens de galinha não produziram um perfil de bandas e foram excluídas das análises de BOX-PCR. O perfil de bandas de 20 linhagens de humanos, 15 de bois e 13 de galinhas foram comparados utilizando o coeficiente de similaridade de Jaccard e o método de agrupamento UPGMA no software GelCompar II. A Figura 5 mostra o dendrograma obtido. Os coeficientes de similaridade variaram de 14,47% a 100%. A estrutura do dendrograma foi complexa, sendo possível observar diversos grupos.

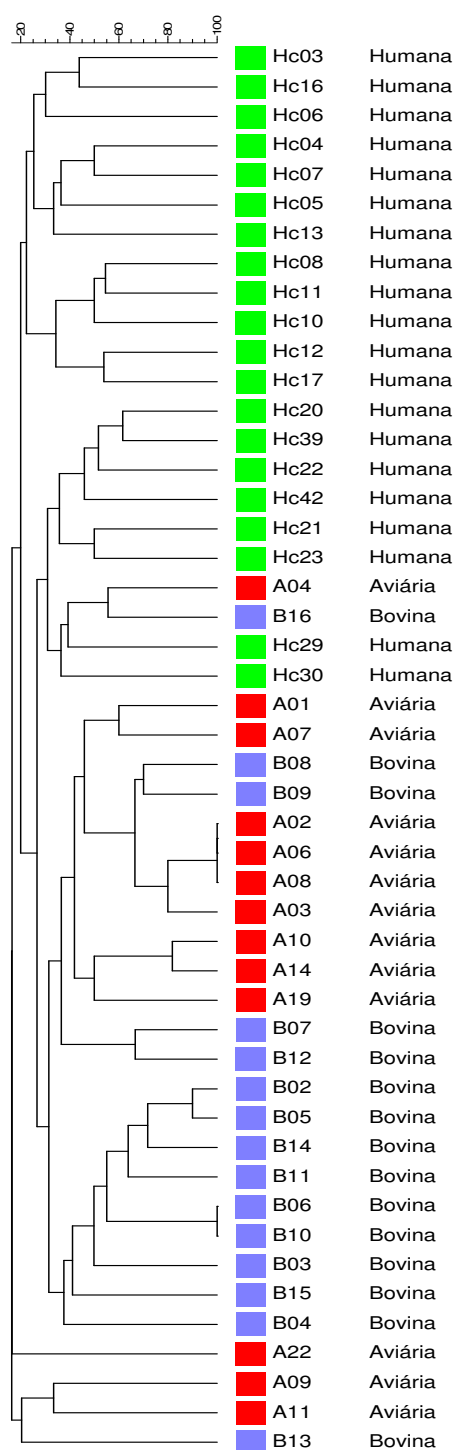


Figura 5. Dendrograma obtido com o coeficiente de similaridade de Jaccard e método de agrupamento UPGMA dos perfis de banda de BOX-PCR. Em verde estão as linhagens de *E. coli*

de origem humana, em vermelho estão as linhagens de *E. coli* de origem aviária e em azul estão as linhagens de *E. coli* de origem bovina.

A análise discriminante *Jackknife* foi realizada a fim de se avaliar a taxa de classificação correta obtida por BOX-PCR na predição da origem animal dessas linhagens de *E. coli*. A Tabela 1 mostra a porcentagem de linhagens de *E. coli* classificadas em cada origem. A taxa global de classificação correta foi de 77,92%, ou seja, 77,92% das linhagens de *E. coli* analisadas foram corretamente classificadas de acordo com sua origem animal. As linhagens de origem aviária apresentaram uma TCC de 93,75%, as linhagens de origem boviana apresentaram um TCC de 80% e as linhagens de origem humana apresentaram uma TCC de 60%. Aproximadamente 35% das linhagens de origem humana foram classificadas como de origem aviária e 5% como de origem bovina. A TCC obtida neste trabalho com o BOX-PCR é similar a algumas encontradas na literatura (Somarelli *et al.*, 2007; Hassan *et al.*, 2007).

Tabela 1. Classificação das linhagens de *E. coli* estimada pela análise *Jackknife* usando as similaridades médias^a.

Origem	Humana	Aviária	Bovina
Humana	60	0	6,67
Aviária	35	93,75	13,33
Bovina	5	6,25	80

^a Os valores são apresentados em porcentagem. Valores em negrito representam a taxa de classificação correta para cada origem. A taxa global de classificação correta foi 77,92%.

Comparação entre BOX-PCR e FT-IR

Existem diversos fatores do hospedeiro que podem selecionar a microbiota intestinal, como por exemplo, o sistema imune, a temperatura corpórea, a dieta e outras características bioquímicas, fisiológicas e comportamentais (Apajalahti, 2005). O uso de algumas características fenotípicas pode ser mais vantajoso para diferenciar *E. coli* de diferentes hospedeiros, uma vez que foi observado que alguns perfis genotípicos são mais variáveis do que alguns perfis fenotípicos (Houser *et al.*, 2008). Recentemente, Grasselli *et al.* (2008) mostraram evidências de transferência horizontal de genes entre *E. coli* comensal de origem humana e não humana, o que também pode dificultar o rastreamento de contaminação fecal por métodos genotípicos.

Os métodos genotípicos, como o BOX-PCR, dependem menos das condições de cultivo do que a técnica de FT-IR. Entretanto, a reprodutibilidade do BOX-PCR é influenciada por um grande número de passos de protocolo e as intensidades das bandas podem variar muito, o que dificulta a interpretação e análise dos perfis. Já a técnica de FT-IR é simples, apesar da grande necessidade de padronizações e dos complexos processamentos de dados (Mouwen *et al.*, 2005). A principal vantagem da espectroscopia FT-IR é a grande amplitude do nível de discriminação da técnica, que pode ser utilizada para discriminar bactérias de diferentes gêneros (Alexandrakis *et al.*, 2008), espécies (Lamprell *et al.*, 2006) e linhagens de uma mesma espécie (Gilbert *et al.*, 2009). Sendo, portanto possível ajustar o modelo ao nível de discriminação necessário, apenas mudando a região do espectro analisada e as análises multivariadas aplicadas. Além disso, após a construção do banco de linhagens, a identificação bacteriana por FT-IR é possível mesmo em culturas mistas e em amostras líquidas, e o cultivo não é necessário (Rodriguez-Saona *et al.*, 2001, Rodriguez-Saona *et al.*, 2004). Outra vantagem é que aparentemente, os perfis de BOX-

PCR são mais variáveis do que os perfis bioquímicos de FT-IR e, essa maior diversidade reflete na necessidade de um banco de linhagens muito maior para BOX-PCR (USEPA, 2005).

Este é o primeiro trabalho que usa espectroscopia FT-IR para discriminar *E. coli* de diferentes hospedeiros. Os resultados obtidos são animadores e indicam que esta poderá ser uma boa técnica para estudos de rastreamento de fonte de contaminação fecal.

REFERÊNCIAS

Alexandrakis D, Downey G, Scannell AG (2008) Detection and identification of bacteria in an isolated system with near-infrared spectroscopy and multivariate analysis. *J Agric Food Chem* 56:3431-3437.

Apajalahti J (2005) Comparative gut microflora, metabolic challenges, and potential opportunities. *J Appl Poult Res* 14:444-453.

Chalmers MJ, Griffiths PR (2002) *Handbook of Vibrational Spectroscopy*. Vol. 5; John Wiley & Sons, Ltd; New York.

Gilbert MK, Frick C, Wodowski A, Vogt F (2009) Spectroscopy imaging for detection and discrimination of different *E. coli* strains. *Applied Spectroscopy* 63:6-13.

Grasselli E, François P, Gutacker M, Gettler B, Benagli C, Convert M, Boerlin P, Schrenzel J, Piffaretti JC (2008) Evidence of horizontal gene transfer between human and animal

commensal *Escherichia coli* strains identified by microarray. FEMS Immunol Med Microbiol 53:351-8.

Hassan WM, Ellender RD, Wang SY (2007) Fidelity of bacterial source tracking: *Escherichia coli* vs *Enterococcus* spp and minimizing assignment of isolates from nonlibrary sources. J Appl Microbiol 102:591-8.

Harwood VJ, Wiggins B, Hagedorn C, Ellender RD, Gooch J, Kern J, Samadpour M, Chapman ACH, Robinson BJ (2003) Phenotypic library-based microbial source tracking methods: Efficacy in the California collaborative study. J Water Health 1:153-166.

Haznedaroglu BZ, Yurtsever D, Lefkowitz JR, Duran M (2007) Phenotypic characterization of *Escherichia coli* through whole-cell fatty acid profiling to investigate host specificity. Wat Res 41:803-809.

Houser BA, Donaldson SC, Padte R, Sawant AA, DebRoy C, Jayarao BM (2008) Assessment of phenotypic and genotypic diversity of *Escherichia coli* shed by healthy lactating dairy cattle. Foodborne Pathog Dis 5:41-51.

Lamprell H, Mazerolles G, Kodjo A, Chamba JF, Noël Y, Beuvier E (2006) Discrimination of *Staphylococcus aureus* strains from different species of *Staphylococcus* using Fourier transform infrared (FTIR) spectroscopy. Int J Food Microbiol 108:125-129.

- Moussa SH, Massengale RD (2008) Identification of the sources of *Escherichia coli* in a watershed using carbon-utilization patterns and composite data sets. J Water Health 6:197-207.
- Mouwen DJM, Weijtens MJB, Capita R, Alonso-Calleja C, Prieto M (2005) Discrimination of Enterobacterial Repetitive Intergenic Consensus PCR Types of *Campylobacter coli* and *Campylobacter jejuni* by Fourier Transform Infrared Spectroscopy. Appl Environ Microbiol 71:4318-4324.
- Naumann D, Helm D, Labischinski H (1991) Microbiological characterizations by FT-IR spectroscopy. Nature 351:81-82.
- Naumann D, Keller S, Helm D, Schultz C, Schrader B (1995) FT-IR spectroscopy and FT-Raman spectroscopy are powerful analytical tools for the non-invasive characterization of intact microbial cells. Journal of Molecular Structure 347:399-406.
- Rodriguez-Saona LE, Khambaty FM, Fry FS, Calvey EM (2001) Rapid detection and identification of bacterial strains by Fourier transform near-infrared spectroscopy. J Agric Food Chem 49:574-579.
- Rodriguez-Saona LE, Khambaty FM, Fry FS, Dubois J, Calvey EM (2004) Detection and identification of bacteria in a juice matrix with Fourier transform–near infrared spectroscopy and multivariate analysis. J Food Protect 67:2555-2559.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: a laboratory manual. 2nd ed. N.Y., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press.

Somarelli JA, Makarewicz JC, Sia R, Simon R (2007) Wildlife identified as major source of *Escherichia coli* in agriculturally dominated watersheds by BOX A1R-derived genetic fingerprints. J Environ Manage 82:60-5.

United States Environmental Protection Agency (2005) Microbial source tracking guide document. EPA/600/R-05/064; U.S. Environmental Protection Agency, Office of Research and Development: Cincinnati, OH.

Yu C, Irudayaraj J (2005) Spectroscopy characterization of microorganisms by Fourier transform infrared microspectroscopy. Biopolymers 77:368-377.

CONCLUSÕES

- A distribuição dos grupos filogenéticos de *E. coli* diferiu entre os hospedeiros analisados.
- Essa diferença pode ser utilizada na predição da fonte principal de contaminação de um corpo hídrico. Essa é uma abordagem rápida, barata e que pode direcionar estudos com técnicas com maior resolução, mais caras e laboriosas.
- Os métodos de BOX- e (GTG)₅-PCR não apresentaram uma boa eficiência na identificação da fonte de contaminação fecal, pois os perfis obtidos apresentaram grande variabilidade.
- Os métodos de perfis genotípicos analisados são laboriosos, requerem muitas padronizações e apresentam baixa taxa de classificação correta, o que pode limitar seu uso em estudos de rastreamento de fonte microbiana. Entretanto, são bons marcadores de variabilidade genética.
- Aparentemente as linhagens de esgoto formaram um grupo distinto, sugerindo que este ambiente pode ser o habitat primário de muitas delas.
- As diferenças entre os espectros obtidos por FT-IR puderam ser observadas com o uso de análises estatísticas multivariadas.
- Os perfis de FT-IR apresentaram menor variabilidade que os perfis genotípicos.

- A espectroscopia FT-IR é uma técnica rápida, robusta, barata após a aquisição do equipamento e independente de cultura após a criação do banco de linhagens.
- Apesar dos resultados apresentados serem preliminares, a espectroscopia FT-IR é uma técnica promissora para futuros estudos de rastreamento de fonte microbiana.

REFERÊNCIAS

BRASIL. Ministério da Saúde. Agência Nacional de Vigilância Sanitária. Portaria 518 de 25 de março de 2004. Estabelece os procedimentos e responsabilidades relativos ao controle e vigilância da qualidade da água para consumo humano e seu padrão de potabilidade e dá outras providências. Diário Oficial da União - República Federativa do Brasil, Poder Executivo, Brasília, DF, 26 mar. 2004. Seção 1. Disponível em: <<http://www.anvisa.gov.br/e-legis>>. Acesso em: mar. 2007.

BRASIL. Ministério da Saúde. Agência Nacional de Vigilância Sanitária. Resolução RDC nº 275 de 22 de setembro de 2005. Aprova o regulamento técnico de características microbiológicas para água mineral natural e água natural. Diário Oficial da União - República Federativa do Brasil, Poder Executivo, Brasília, DF, 23 set. 2005. Seção 1. Disponível em: <<http://www.anvisa.gov.br/e-legis>>. Acesso em: mar. 2007.

Brenner DJ (1983) Impact of modern taxonomy on clinical microbiology. ASM News 49:58-63

Brenner DJ, Falkow S (1971) Molecular relationships among members of the *Enterobacteriaceae*. Adv Genet 16:81-118.

Casarez EA, Pillai SD, Mott JB, Vargas M, Dean KE, Di Giovanni GD (2007) Direct comparison of four bacterial source tracking methods and use of composite data sets. J Appl Microbiol 103:350-364.

- 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.
- Edberg SC, Rice EW, Karlin RJ, Allen MJ (2000) *Escherichia coli*: the best biological drinking water indicator for public health protection. *J Appl Microbiol* 88:106-116.
- Field KG, Samadpour M (2007) Fecal source tracking, the indicator paradigm, and managing water quality. *Wat Res* 41:3517-3538.
- Harwood VJ, Wiggins B, Hagedorn C, Ellender RD, Gooch J, Kern J, Samadpour M, Chapman ACH, Robinson BJ (2003) Phenotypic library-based microbial source tracking methods: Efficacy in the California collaborative study. *J Water Health* 1:153-166.
- Hartl DL, Dykhuizen DE (1984) The population genetics of *Escherichia coli*. *Ann Rev Genet* 18:31-68.
- Health and Welfare Canada (1992) Guidelines for Recreational Water Quality. National Health and Welfare. Ottawa, Canada.
- Jiang SC, Chu W, Olson BH, He JW, Choi S, Zhang J, Le JY, Gedalanga PB (2007) Microbial source tracking in a small southern California urban watershed indicates wild animals and growth as the source of fecal bacteria. *Appl Microbiol Biotechnol* 76:927-934.

- Kuhnert P, Boerlin P, Frey J (2000) Target genes for virulence assessment of *Escherichia coli* isolates from water, food and the environment. FEMS Microbiol Rev 24:107-117.
- Lessa P (1990) Multidimensional scaling of geographic genetic structure. Syst Zool 39:242-252.
- Meays CL, Broersma K, Nordin R, Mazumder A (2004) Source tracking fecal bacteria in water: a critical review of current methods. J Environ Manage 73:71-79.
- Mohapatra BR, Broersma K, Nordin R, Mazumder A, (2007) Evaluation of repetitive extragenic palindromic-PCR for discrimination of fecal *Escherichia coli* from humans, and different domestic- and wild-animals. Microbiol Immunol 51:733-740.
- Moussa SH, Massengale RD (2008) Identification of the sources of *Escherichia coli* in a watershed using carbon-utilization patterns and composite data sets. J Water Health 6:197-207.
- Nelson M, Jones SH, Edwards C, Ellis JC (2008) Characterization of *Escherichia coli* populations from gulls, landfill trash, and wastewater using ribotyping. Dis Aquat Organ 81:53-63.
- Parveen S, Portier KM, Robinson K, Edmiston L, Tamplin ML (1999) Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Appl Environ Microbiol 65:3142–3147.

- Price B, Venso E, Frana M, Greenberg J, Ware A (2007) A comparison of ARA and DNA data for microbial source tracking based on source-classification models using classification trees. *Wat Res* 41:3515-3584.
- Rabinovici SJM, Berhknopf RL, Wein AM, Coursey DL, Whitman RL (2004) Economic and health risks trade-offs of swim closures at a Lake Michigan Beach. *Environm Sci Technol* 38:2737-2745.
- Santo Domingo JW, Bambic DG, Edge TA, Wuertz S (2007) Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. *Wat Res* 41:3539-3552.
- Scott TM, Rose JB, Jenkins TM, Farrah SR, Luhasik J (2002) Microbial source tracking: current methodology and future directions. *Appl Environ Microbiol* 68:5796-803.
- Siegrist TJ, Anderson PD, Huen WH, Kleinheinz GT, McDermott CM, Sandrin TR (2007) Discrimination and characterization of environmental strains of *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). *J Microbiological Methods* 68:554-562.
- Simpson JM, Santo Domingo JW, Reasoner DJ (2002) Microbial source tracking: state of the science. *Environ Sci Technol* 36:5279-88.

United States Environmental Protection Agency (2005) Microbial source tracking guide document. EPA/600/R-05/064; U.S. Environmental Protection Agency, Office of Research and Development: Cincinnati, OH.

United States Environmental Protection Agency (2002) Method 1603 *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* Agar (Modified m-TEC), EPA 821-R-02-023, U.S. Environmental Protection Agency, Office of Research and Development: Cincinnati, OH.

World Health Organization (2007) Water supply, sanitation and hygiene development. Disponível em: http://www.who.int/water_sanitation_health/hygiene/en/index.html. Acesso em 06.07.2007.

World Health Organization (2004) Guidelines for drinking water quality. Volume 1. Recommendations. 3rd ed. World Health Organization. Geneva, Disponível em http://www.who.int/water_sanitation_health/dwq/em/index.html.

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação/tese de Mestrado intitulada “Identificação de marcadores moleculares hospedeiro-específicos de *Escherichia coli* de águas superficiais do Estado de São Paulo”:

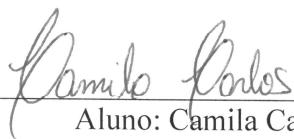
() não se enquadra no § 3º do Artigo 1º da Informação CCPG 01/08, referente a bioética e biossegurança.

() está inserido no **Projeto CIBio/IB/UNICAMP** (Protocolo nº _____), intitulado _____;

() tem autorização da **Comissão de Ética em Experimentação Animal/IB/UNICAMP** (Protocolo nº _____);

(X) tem autorização do **Comitê de Ética para Pesquisa com Seres Humanos/FCM/UNICAMP** (Protocolo nº 049/2009);

() tem autorização de comissão de bioética ou biossegurança externa à UNICAMP.
Especificar: _____



Aluno: Camila Carlos



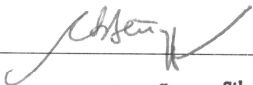
Orientador: Laura M. M. ottoboni

Para uso da Comissão ou Comitê pertinente:

(X) Deferido () Indeferido

Nome: _____

Função: _____


Prof. Dra. Carmen Silvia Bertazzo
Vice-Presidente do
Comitê de Ética em Pesquisa
FCM/UNICAMP
Matrícula: 28856-6