Study of The Diversity of 16S–23S rDNA Internal Transcribed Spacer (ITS) Typing of Escherichia Coli Strains Isolated From Various Biotopes in Tunisia

Souhir Badi  
Centre de Recherches et des Technologies des Eaux

Rim Ammeri (✉ rim.werheni@gmail.com)  
Centre de Recherches et des Technologies des Eaux

Mohamed Abbassi  
Institute of Veterinary Research of Tunisia, Tunis, Tunisia.

Mejdi Snoussi  
Centre de Recherches et des Technologies des Eaux

Paola Cremosini  
Istituto di Biologia e biotecnologia Agraria, CNR, Lodi, Italy.

Mario Luini  
IZSLER, Lodi, Italy

Bianca Castiglioni  
Istituto Zooprofilattico Sperimentale delle Venezie

Abdennaceur Hassen  
Centre de Recherches et des Technologies des Eaux

Research Article

Keywords: Escherichia coli, Environmental biotopes, Phylogroups, Genotyping, RS-PCR, ITS region

Posted Date: July 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-702436/v1

License: ☇ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

Version of Record: A version of this preprint was published at Archives of Microbiology on December 19th, 2021. See the published version at https://doi.org/10.1007/s00203-021-02684-x.
Abstract

We investigated the 16 S-23S rRNA intergenic spacer region (ISR)-PCR and the phylogenetic PCR analyzes of 150 *Escherichia coli* isolates as tools to explore their diversity, according to their sampling origins, and their relative dominance in these sampling sources. So, these genetic markers are used to explore phylogenetic and genetic relationships of these 150 *E. coli* isolates recovered from different environmental sources (water, food, animal, human and vegetables). These isolates are tested for their biochemical pattern and later genotyped through the 16S–23S rRNA intergenic spacer PCR amplification and their polymorphism investigation of PCR-amplified 16S-23S rDNA ITS. The main results of the pattern band profile revealed one to 4 DNA fragments. Distributing 150 *E. coli* isolates according to their ITS and by using RS-PCR, revealed 4 genotypes and 4 subtypes. The DNA fragment size ranged from 450 to 550 bp. DNA band patterns analysis revealed considerable genetic diversity in interspecies. Thus, the 450 and 550 bp size of the common bands in all *E. coli* isolates are highly diversified. Genotype I appeared as the most frequent with 77.3% (116 isolates), genotype II with 12% (18 isolates); genotype III with 9.7% (14 isolates), and the IV rarely occurred with 4% (2 isolates). Distributing the *E. coli* phylogroups showed 84 isolates (56%) of group A, 35 isolates (23.3%) of group B1, 28 isolates (18.7%) of group B2 and only 3 isolates (2%) of group D.

Introduction

*Escherichia coli* is associated with a variety of intestinal diseases in humans and animals (Fröhlicher et al., 2008; Martinez-Medina and Garcia-Gil, 2014). Some pathogenic Escherichia coli can produce adherent and destructive lesions, which are characterized by the bacteria tightly adhering to intestinal epithelial cells and destroying the underlying cytoskeleton (Jang et al. 2017). Similarly, *E. coli* is considered a normal resident of the intestines of humans and most animals. Some *E. coli* strains can cause a variety of intestinal and extra-intestinal diseases, such as diarrhea, urinary tract infections, sepsis, and neonatal meningitis (Gomes et al. 2016). The typing method to distinguish different bacterial isolates of the same species is an indispensable epidemiological tool in infection prevention and control. Traditional typing systems based on phenotypes (such as serotypes, biotypes, phage types or antibacterial profiles) have been used for many years (Fratamico et al. 2016). However, more advanced molecular methods have recently been developed to examine the affiliation of microbial isolates, and these methods have changed our ability to accurately distinguish bacterial types and subtypes (Sabat et al., 2013). Since the last decade, several molecular techniques for microbial characterization have been developed. The study of non-coding RNA is important for finding functions or roles in cells (Harris et al. 2018). To understand its function, we may find a derivative structure. The tRNA family is a form of RNA molecule with a special function that can convert amino acids into protein-building machinery. 16S, 23S, ITS, gyrase, RNA polymerase and DNA ligase are highly conserved genes in bacteria and can be used for molecular identification. The internal transcription spacer (ITS) is also called ISR and exists between the 16S and 23S rDNA regions of ribosomal genes. The arrangement of complete ribosomal gene units (such as 16 S-ITS-23S-ITS-5S) is scattered in the genome of bacteria, and its copy number is between 1 and 15 (Tacao
et al. 2005). The recently developed DNA fingerprinting method is based on the repeated intergenic consensus sequence (ERIC) of intestinal bacteria, and has described repeated foreign palindrome (REP) and BOX elements for distinguishing bacterial strains (Xin Wang et al. 2015; Tacao et al. 2005). According to reports, the length polymorphism in the internal transcription spacer (ITS) of 16S-23S ribosomal DNA is a stable genetic marker for studying bacterial phylogeny. Although rRNA (and rRNA genes) are highly conserved, nucleotide variation between rDNA sequences is usually large enough to be used to estimate the relationship between bacterial phylogenetic profiles (Gutellet et al. 1994). The usefulness of rDNA sequence as a classification tool has been shown in bacteria, among which bacterial 16S rRNA sequence analysis has redefined the phylogenetic relationship, and it depended too on cell metabolism before (Fox et al. 1980). The size and number of DNA fragments generated by PCR amplification can be achieved to quickly identify a wide range of bacteria. In this work, we used 16S-23S ITS genetic markers to study the phylogenetic relationships of some E. coli strains isolated from various environmental biomes (water, animals, humans, and vegetables). Several biochemical markers have been studied in advance to identify these isolates. The relationship between these E. coli isolates and their system group members and their origins will be studied.

Materials And Methods

Collection, Bacterial Isolates Media and Chemicals

All samples collected from various biotopes (animal organs and meats, soil, water, feces of varied animals, food, humans, nosocomial and abattoir environment) were cultured onto 3 specific media, either eosin-methylene blue (EMB) agar, Chromagar™ and MacConkey agar for 24h at 37°C for the detection of \textit{E. coli}. A single colony from positive samples was sub-cultured on nutrient agar for 18–24 h at 37°C. The cultures were kept in the freezer (+4°C) before the different tests and sub-cultured on a new nutrient agar when considered needed. The specific colony of presuming \textit{E. coli} was preliminarily identified by the characteristic green metallic sheen on the EMB or blue color on the Chromagar™ or brick red on the MacConkey. Colonies with typical \textit{E. coli} morphology were selected and identified by some standard specific morphological and biochemical tests such as Gram stain, catalase, oxidase, indole, methyl-red-Voges-Proskauer, citrate and urease, and confirmed by the Api 20E system (BioMérieux, La Balme Les Grottes, France). The final identification of all the isolates was made by the polymerase chain reaction (PCR) with specific genes of \textit{E. coli} and by using the 2 primers as shown: (Altschul et al. 1997). Ert2. F: 5′-ACT GGA ATA CTT CGG ATT CAG ATA CGT-3’ and Ert2. R: 5′–ATC ACA GAT TCA TTC CAC GAA a-3’. All \textit{E. coli} isolates were stored at -80°C in brain-heart infusion broth containing 20% of glycerol.

Extraction of Genomic DNA

The method adopted is based on the ability of silica resin to bind DNA in the presence of a high concentration of guanidine thiocyanate chemotropic agent that guaranteed an excellent disruption of bacterial cells, collected from the MacConkey plates. Purified DNA was recovered from cell lysates using 2 sequential chloroform phenol extraction and ethanol precipitation steps (Jenson et al. 1993). DNA is
typically determined by spectrophotometer at 260 nm, and one absorbance unit \((A_{260})\) corresponded to 50 mg DNA/ml. The purity may also be estimated by spectrophotometer from the relative absorbance determination at 260 and 280 nm, respectively \((A_{260}/A_{280})\). Due to the variation between individual starting DNA materials, the expected range of \(A_{260}/A_{280}\) ratios will be around 1.6–1.8.

**Phylogenetic Grouping**

PCR was performed with a Perkin-Elmer Gene Amp 9600 thermocycler under these conditions: denaturation for 5 min at 94°C; 30 cycles of the 30s at 94°C, 30s at 55°C and 30s at 72°C, and a final extension step of 7 min at 72°C (Clermont et al. 2000). Phylogroups and subtypes were identified according to Clermont et al. (2000) and Escobar et al. (2004).

**PCR Amplification of 16 S-23S rDNA ITS and Reaction Conditions**

The method of Jensen et al. (1993) was used in operating RS-PCR genotyping that is based on the amplification of the 16 S-23S rRNA ISR. G1 and L1 primers defined by Jensen et al. (1993) were used in operating genotyping. The first primer G1 was selected from a highly conserved region immediately adjacent to the 16S-23S spacer. This oligonucleotide contains the sequence 16F: GAAGTCGTAACAAGG and it is about 30 to 40 nucleotides upstream from the spacer boundary (Fig.1). The second primer L1 was chosen from the 5 bacterial and 4 plant chloroplasts 23S sequences compiled by Getell et al. (1988). This sequence 23R: CAAGGCATCCACCGT, is the most conserved 23S sequence immediately following the spacer, and it is situated approximately 20 feet downstream from the spacer boundary (Jensen et al. 1993). Primers for both the 16S and 23S regions were restricted to a length of 15 bases because of variations in the sequences beyond these highly conserved regions. The protocol of DNA amplification adopted was as recommended by Fournier (2008). Each reaction of a total volume of 25 µl contained 1 Hot Sar Taq Master Mix (Qiagen), 800 nM of each primer (G1 and L1 primer) and 30 ng NA. The PCR profile was: 95°C for 15 min, followed by 27 cycles at 94°C for 1 min, followed by 2 min ramp and annealing at 55°C for 7 minutes. Then a further 2 min ramp, the extension was done at 72°C for 2 min, PCR was terminated as described.

**Construction and Analysis of Dendrogram**

Fingerprints, discontinuous noise and the overall density of fingerprints, ITS-PCR patterns, a band-matching algorithm (match-matching tolerance of 1.0%) were used to calculate the pairwise similarity matrix with similarity coefficients. Cluster analysis of similarity matrices was performed by UPGMA (Unweighted Pair Group Method with Arithmetic Mean). Major DNA bands were needed for constructing the phylogenetic tree with TFPGA (Tools for Population Genetic Analyses). Each isolate of *E. coli* was one population. Hence there were 4 total populations and 2 loci were considered for constructing the dendrogram.

**Results**
Morphological and Biochemical Characteristics of All E. coli isolates

This study was conducted between January and July 2012, and 150 E. coli isolates are collected from various biotopes, distributed: 22 from animal organs and meats, 21 from varied kinds of soil, 45 from different types of water, 16 from feces of varied animals, 27 from different foods, 5 from humans, and 14 from different nosocomial and abattoir environment.

According to the results of morphological and biochemical tests, all isolates were as Gram, urease, catalase and oxidase negative, and 97% were indole positive, 75% were mobile, 92% were citrate of Simmons negative, and 99% were gas glucose positive (Table 1).

Phylogenetic Grouping

Distributing the phylogroups of 150 E. coli isolates showed that 84 strains belonged to group A (56%), including subgroups A1 (37, 24.7%) and A0 (47, 31.3%), and 35 strains (23.3%) belonged to the phylogroup B1 (Figure 1, Table 2). Therefore, phylogroups A and B1 accounted for around 79.3% of the total isolates. The number of 28 isolates (18.7%) belonged to the phylogenetic group B2, and 19 isolates (12.7%) were detected and subtyped as B2, and 9 isolates as B23 (6%). Phylogroup D was infrequent, with only 3 isolates (2%).

Distribution of Genes Encoding RS-PCR Groups According to the Origins of E. coli isolates.

The result of Table 3 showed the dominance of water as the most important source of E. coli isolates with around 37.3% (56/150), the food animals, vegetables, feces, and soil arrive after with a respective frequency of 0.18% (27/150), 0.17% (26/150), 0.15% (23/150) and 0.11% (17/150). The lowest one is observed from a human source with a frequency of 0.01% (2/150). Distributing phylogroups according to the origins of E. coli isolates showed that water and food animals appeared as the primary sources of the different phylogroups registered in this study since all known E. coli phylogroups were found with water and food animals. The soil and vegetables, sampling origins came in the second range with a smaller number of phylogroup types. While E. coli isolates from feces appeared as presenting the lowest number of known E. coli phylogroups.

Cloning and Sequencing of 16 S-23S RDNA ITS

In the present study, 16S–23S rRNA ISRs of 150 E. coli strains were successfully amplified by PCR. Besides the required presence of the nuc gene, all the 150 E. coli isolates presented the Ert2 genes, confirming the strain identity as E. coli. Analysis of the 16 S-23S-region revealed various sized amplicons of 100, 150, 290, 320, 450, 550 bp, respectively; the most frequent ones ranged between 450 and 550 pb. The configuration ITS patterns reflected a developed level phylogenetic grouping:

Amplicons of nearly 750 pb were calculated to comprise approximately 550 pb of the 3’portion of the 16S rRNA genes and 450 bp of the 5’ portion of the 23S rRNA gene; a selection of 12 strains is shown in
Figure 2. Patterns generated by RS-PCR and the miniaturized electrophoresis were well reproducible to show that 4 distinct groups of *E. coli* isolates were found (Figure 3).

The dendrogram showing the clustering of the amplification patterns of *E. coli* with RS-PCR is generated by using the squared Euclidean distance measure and the average linkage clustering method with the program SPSS 22 for Windows (Figure 4).

Therefore, genotype I appeared as the most frequent with 77.3% (116 isolates), genotype II with 12% (18 isolates); genotype III with 9.7% (14 isolates), and the IV rarely occurred in 4% (2 isolates).

Analysis of the 16S–23S rRNA intergenic spacer region by RS-PCR revealed 4 genotypes and 4 subtypes. On one side, the genotype I appeared as the most frequent in 116 isolates among 150 ones and represented 77.3% of all the isolates. By against, 18 isolates representing 12% of all isolates belonged to genotype B1; therefore, phylogroups A and B1 accounted for the whole 89.3% (134 isolates) of the isolates. The number of 14 isolates (9.7%) belonged to genotype B2, and the genotype D rarely occurred with 4% (2 isolates). For further analysis, the rare genotypes were grouped and named other genotypes (OG). The dendrogram confirmed the dissimilarities between the different genotypes, in particular for strains of genotype B2 and D (Figure 4, 6).

Comparisons of aligned 16S–23S rDNA space region sequences revealed that rRNA processing motifs are highly conserved within the 16S–23S rDNA space regions of all the isolates (Figure 3). In the sequence region between the tRNA genes, the number of nucleotide positions varied from 7 (in genomovars 1 and 5) to 31.

*Taq*I restriction profiles of the 16S–23S rDNA space region amplicons, using the PCR primers 16F and 23R, included 4 to 7 bands of sizes ranging from 450 to 550 bp (Figure 3). Identical TaqI digestion always profiles were found for strains belonging to a genomovar (as defined by genomic DNA similarities). All *E. coli* genomovars presented 2 characteristic bands of 450 and 550 bp related to their *Taq*I restriction patterns (Figures 2 and 3). The 16S–23S rDNA space region restriction patterns of strains of *E. coli*, generated by *Taq*I digestion, were clustered by UPGMA (Figure 5). Branching dichotomies, because of 16S–23S rDNA space region polymorphisms, resulted in clusters of strains at the species level.

**Discussion**

The phylogenetic analysis of 150 isolates showed that in our study, most of the isolates belonged to phylogenetic groups A and B1. Isolates belonging to these 2 system groups are considered symbionts of animal or human origin, because their genes encoding virulence factors are few and unrelated, and they are found in human or animal naturally infectious *E. coli* isolates. The frequency of phylogroups is lower (Carlos et al. 2010; Jakobsen et al. 2010). However, based on several studies, based on the presence of several virulence genes, B2 and D phylogroups strains appear to be more toxic (Kilani et al. 2017). Interestingly, no significant relationship was found between antibiotic resistance and the members of the phylogenetic group of the isolates. Fecal contamination is mainly caused by the dominance of
Escherichia coli, which constitutes a serious environmental problem and may affect many coastal and inland waters around the world (Anderson et al. 1997). Point source discharges such as raw sewage, stormwater, and combined sewer overflow, effluents from wastewater treatment plants and agro-alimentary industry sources, are the major contributors to fecal pollution and contamination of natural environmental systems (Griffin et al. 2001).

Thus, 16S–23S rDNA space regions may be good targets of genomovar- and species-specific probes for environmental monitoring. The fewer conserved 16S–23S rDNA space regions can be applied as a high-resolution indicator of the evolutionary divergence of *E. coli* strains.

Despite the observation that every species of *E. coli* analyzed presented a unique restriction pattern, more strains of each species will need to be analyzed before arriving at general conclusions about the utility of 16S–23S rDNA space region restrictions for the identification of strains at the species level. Whereas comparisons of 16S rRNA gene sequences are restricted in their power to resolve closely related species of a genus (Fox et al. 1992; Martinez-Murcia et al., 1992; Hauben et al., 1997, 1998), spacer regions within the 16S and 23S genes in prokaryotic rRNA genetic loci exhibit significant length and sequence polymorphisms in different species and are flanked by highly conserved sequences (Jenson et al. 1983).

Our results confirm that RS-PCR can be used as a rapid test for molecular typing of *E. coli* strains isolated from various biological communities, and can identify genetic subtypes with specific virulence. In the study of Fournier et al. (2008), the authors concluded that the bovine Staphylococcus aureus isolates are genetically heterogeneous using the 16S-23S rDNA spatial region. Maeda et al. (2000) showed that the 16S-23S rRNA intergenic regions contained different tRNA compositions, and the similarities in the nucleotide sequence of the non-coding regions flanking the tRNA gene have been noted. The phylogenetic information variable site is only located in the non-coding region. The sequence analysis results of the 16S-23S rDNA spatial region maintain and correlate with the clear phylogenetic relationship in the phylogenetic group, providing an alternative tool for genotype and *E. coli* species differentiation (Hin-Choung et al., 2001). Therefore, amplification using primers considered based on these flanking sequences will produce polymorphic fingerprints that could distinguish bacterial strains at the species and subspecies levels (Bidet et al. 2000). However, since the RS-PCR patterns are more simply visible visually than the REP-PCR or ERIC-PCR patterns, they may be a practical technique for routine usage (Hin-Choung et al. 2001). Also, the sensitivity and the specificity of the RSS-PCR method were 100 and 96%, respectively (Kimura et al 2000). Although variances are detected in the number and size of the PCR-6S–23S rDNA space region products attained from different strains, these characteristics alone could not be used for an overall difference of all genomovars or *E. coli* species.

**Conclusion**

In this study we evaluated the diversity of 150 *E. coli* strains isolated from different sites. Isolated and selected strains were submitted to identification, virulence gene and antibiotic resistance analyses. The results obtained can be used in epidemiological, diagnostic, virulence and molecular taxonomy studies.
The study of the diversity of E.coli strains in this study was performed by the RS-PCR technique. RS-PCR is a suitable rapid typing method for E. coli isolates. This method was implemented in its high discriminative power. But RS-PCR may be a more practical method because fewer amplification bands and patterns are generated, simplifying reviews and interpretation of data. It is a rapid, easily workable and achievable and reproducible method appropriate for genotyping of A. hydrophila, for example at the strain level. 16 S-23S rRNA and phylogroups analysis of E. coli isolates revealed the potential for identifying sources of E. coli environmental contamination. A fairly small number of isolates are necessary to find the candidate source-specific E. coli that is stable and unchanging under the simulated environmental conditions. The results achieved by the RS-PCR technique will be invaluable for developing extra typing strategies and the optimization of traditional typing methods, such as the triplex PCR of phylogeny group approaches. The ribosomal spacer PCR (RS-PCR) appeared as a highly resolving and robust genotyping method for E. coli of moderate costs and suitable to be used for routine. Water and food animals appeared as the most sampling sources of E. coli isolation, showing high diversity. Soil, vegetables and feces came in the second rank, while the human origin showed the least E. coli diversity.

**Declarations**

**Ethics approval and consent to participate:** Not applicable

**Consent for publication:** Not applicable

**Author's Contributions:** "SB analyzed and interpreted the patient data. RWA major contributor in writing the manuscript, and was a corresponding author, MSA major contributor in E. coli strain selection, MS analyzed the strain characterization, PC, ML and BC a major contributor in molecular analysis, AH a major contributor in writing the manuscript. All authors read and approved the final manuscript."

**Availability of data and materials:** Not applicable

**Funding:** Not applicable

**Competing interests:** Not applicable

**References**

1. Altschul, S. F. *et al.* DJ(1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

2. Anderson, S. A., Turner, S. J. & Lewis, G. D. Enterococci in the New Zealand environment: implications for water quality monitoring. *Water Sci. Technol*, **35**, 325–331 (1997).

3. Bidet, P. *et al.* Comparison of PCR ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing Clostridium difficile. *J Clin Microbiol*, **38**, 2484–2487 (2000).

4. Cedergren, R. J., Abel, G. Y. & Sankoff, D. 1(988) Evolutionary relationships among known life forms. *J. Mol. Evol*. 28:98–112. 6.
5. Dams, E., Hendricks, L., ClermontO, Bonacorsi, S. & Bingen, E. Rapid and simple determination of the Escherichia coli phylogenetic groups. *Environ. Microbiol*, 66, 4555–4558 (2000).

6. Fournier, C. et al. Bovine Staphylococcus aureus: Association of virulence genes, genotypes and clinical outcomes. *Research in veterinary Science*, 85, 439–448 (2008).

7. Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P. How close is close: 16S rRNA sequence identity may not guarantee species identity? *Int. J. Syst. Bacteriol*, 42, 166–170 (1992).

8. Fratamico, P. M. et al. (2016) Advances in Molecular Serotyping and Subtyping of Escherichia coli. Front in Microbiol 7

9. Frohlicher, E., Krause, G., Zweifel, C., Beutin, L. & Stephan, R. Characterization of attaching and effacing Escherichia coli (AEEC) isolated from pigs and sheep. *BMC Microbiology*, 8, 144 (2008).

10. Gomes, T. A. T. et al. (2016) Diarrheagenic Escherichia coli. Braz. J of Microbiol 47, 1, 3–30.

11. Griffin, D. W., Lipp, E. K., McLaughlin, M. R. & Rose, J. B. (2001) *Marine recreation and public health microbiology: the quest for the ideal indicator*. BioScience 51:817–825.

12. Gutell, R. R. & Fox, G. E. (1988) A compilation of large subunit RNA sequences in a structural format. Nucleic Acids Res. 16 (Suppl.), r175-r201.

13. Gutell, R. R., Larsen, N. & Woese, C. R. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev*, 58, 10–26 (1994).

14. Harris, K. A. & Breaker, R. R. (2018) Large noncoding RNAs in bacteria. Microbiol Spectr. 6(4)

15. Hauben, L. et al. Phylogenetic position of phytopathogens within the Enterobacteriaceae. *Syst. Appl. Microbiol*, 21, 384–397 (1998).

16. Hauben, L., Vauterin, L., Swings, J. & Moore, E. R. B. Comparison of 16S ribosomal DNA sequences of all Xanthomonas species. *Int. J. Syst. Bacteriol*, 47, 328–335 (1997).

17. Hing, C. & Chih, C. (2001) *Evaluation of Typing of Vibrio parahaemolyticus by Three PCR Methods Using Specific Primers*. J. Clin. Microbiol. 39(12):4233–4240.

18. Jang, J. et al. Environmental Escherichia coli: Ecology and public health implications—A review. *J. Appl. Microbiol*, 123, 570–581 https://doi.org/10.1111/jam.13468 (2017).

19. Jensen, M. A., Webster, J. A. & Straus, N. Rapid identification of bacteria based on the polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Appl. Environ. Microbiol*, 59, 945–952 (1993).

20. Kimura, R. et al. Restriction-Site-Specific PCR as a Rapid Test To Detect Enterohemorrhagic Escherichia coli O157:H7 Strains in Environmental Samples. *Appl Environ Microbiol*, 66 (6), 2513–2519 https://doi.org/10.1128/aem.66.6.2513-2519 (2000).

21. Maeda, T., Takada, N., Furushita, M. & Shiba, T. Structural variation in the 16 S-23S rRNA intergenic spacers of Vibrio parahaemolyticus. *FEMS Microbiol. Lett*, 192, 73–77 (2000).

22. Martinez-Medina, M. & Garcia-Gil, L. J. Escherichia coli in chronic inflammatory bowel diseases: An update on adherent invasive Escherichia coli pathogenicity. *World J Gast Pathol*, 15 (3), 213–227 (2014). 10.4291/wjgp. v5.i3.213
23. Martinez-Murcia, A. J., Benlloch, S. & Collins, M. D. Phylogenetic interrelationships of members of the neural Aeromonas and Plesiomonas as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridization. *Int. J. Syst. Bacteriol*, 42, 412–421 (1992).

24. Orskv, F. & Orskv, I. Escherichia coli serotyping and disease in man and animals. *Can. J. Microbiol*, 38, 699–704 (1992).

25. Sabat, A. J. *et al.* Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro surveillance*, 18, 4 (2013).

26. Tacao, M., Alves, A., Saavedra, M. J. & Correia, A. BOX–PCR Is an Adequate Tool for Typing Aeromonas spp. *Antonie van Leeuwenhoek*, 88, 2173–2179 (2005).

27. Wang, X., Jordan, I. K. & Mayer, L. W. (2015) *A Phylogenetic Perspective on Molecular Epidemiology.* *Mol. Med. Microbiol* 517–536. Doi:10.1016/b978-0-12-397169-2$400029-9

**Tables**

**Table 1:** Principal biochemical characters of the different *E. coli* isolated from the various biotopes examined.
| Biochemical character | Isolates number | Results |
|-----------------------|----------------|---------|
| **Oxidase**           | 0              | -       |
| **Test ONPG**         | 259            | +       |
| **Lactose**           | 217            | + or (-)|
| **H₂S**               | -              | -       |
| **LDC**               | -              | D       |
| **ODC**               | -              | D       |
| **ADH**               | -              | D       |
| **Indole**            | 259            | +       |
| **Urease**            | 2              | -       |
| **TDA, PDA**          | 13             | -       |
| **Growth on Simmons Citrate** | 11         | -    |
| **Malonate**          | 23             | -       |
| **VP**                | 19             | -       |
| **Gaz-glucose**       | 229            | +       |
| **Gelatinase**        | 21             | -       |
| **Saccharose**        |                | D       |
| **Inositol**          | 9              | -       |
| **Adinitol**          | 4              | -       |
| **Galacturonate**     | 259            | +       |

+: Positive; -: Negative d: Not detected.

**Table 2:** Distribution of different *E. coli* isolates into phylogenetic groups according to their origins
| Phylogroups | Isolates \(n=150\) | Soil \((n=20)\) \(\%\) | Wastewater \((n=75)\) \(\%\) | Animals/PA \((n=105)\) \(\%\) | Vegetables \((n=44)\) \(\%\) | Hunan \((n=15)\) \(\%\) | \(P\) |
|-------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|---|
| \(A_0\)    | 97 (37.4)        | 7               | 27              | 39              | 17              | 7               | 0.24 |
| \(A_1\)    | 51 (19.7)        | 5               | 9               | 25              | 10              | 2               | 0.22 |
| \(B_1\)    | 59 (22.7)        | 5               | 18              | 22              | 10              | 4               | 0.22 |
| \(B_{22}\) | 25 (9.6)         | 2               | 5               | 15              | 2               | 1               | 0.24 |
| \(B_{23}\) | 9 (3.4)          | 0               | 3               | 1               | 5               | 1               | 0.24 |
| \(D_1\)    | 14 (5.4)         | 1               | 9               | 1               | 2               | 0               | 0.24 |
| \(D_2\)    | 4 (1.5)          | 0               | 4               | 0               | 0               | 0               | 0.29 |

**Table 3:** Distribution of genes encoding RS-PCR Groups, the phylogenetic groups among their origins of 150 *E. coli* strains.

| RS-PCR groups \((n)\) | Origin of strains \((n)\) | Phylogroups |
|------------------------|---------------------------|-------------|
| I (116)                | Water (41), soil (13), VG (17), f (22), fa (21), hum (2) | \(A_0\) (40), \(A_1\) (29), \(B_1\) (25), \(B_{22}\) (13), \(B_{23}\) (7), \(D\) (2) |
| II (14)                | Water (5), soil (1), VG (4), f (1), fa (3) | \(A_0\) (1)*, \(A_1\) (5), \(B_1\) (3), \(B_{22}\) (3), \(B_{23}\) (2) |
| III (18)               | Water (8), soil (3), VG (5), fa (2) | \(A_0\) (5), \(A_1\) (5), \(B_1\) (2), \(B_{22}\) (5), \(D\) (1) |
| IV (2)                 | Water (1), fa (1) | \(A_1\) (1), \(B_1\) (1) |

Legends: Vg: Vegetables, hum: Human, fa: Food animals, f: Feces, *: \(A_0\) (1) = Phylogroup A0+ (1) Number of isolates;

**Figures**
Figure 1

Representative gel electrophoresis of triplex genotyping PCR assay conducted on E. coli using ATCC strain 25,922 as a positive control. Expected product sizes are 279 pb for the chutA gene, 211 pb for the yjaA gene, and 152 pb for the TSPE4 fragment.

Figure 2

Schematic representation of ITS region position between the 16S and 23S rDNA gene region obtained for some selected E. coli isolates; Lane M: Gene ruler expresses DNA ladder 100 bp (Fermentas), Line 1: Isolate 132, Line 2: Isolate 133, Line 3: Isolate 134, Line 4: Isolate 135, Line 5: Isolates 136, Line 6:
Isolates 137, Line 7: Isolate 138, Lane 8: Isolates 139, Line 3: Isolate 140, Line 3: Isolate 141, Line 3: Isolate 142 and Line 3: Isolate 143.

Figure 3

Examples of genotype electrophoresis of various PCR products of the 16S – 23S rRNA ITS for some selected E. coli isolates
Figure 4

Dendrogram, derived from calculated evolutionary distances, depicting the estimated evolutionary relationships between the 16S–23S rDNA space region sequences of genomovars of E. coli. Sequences for tRNA genes were removed from the calculations. The branch scale bar shows a means of 10 substitutions at any nucleotide position per 100 positions.
Figure 5

Dendrogram constructed based on the DNA bands in ITS region present in different isolates of E. coli using the UPGMA method. The 150 E. coli strains were isolated between January to July 2013 from different environmental biotopes. The letters of the external nodes specify the genotype. These numbers indicate the different clusters of each genotype.
Figure 6

Heat map of E. coli strains isolated from different origins

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- floatimage1.png