

BRIEF REPORT

RESISTOME AND COMPARATIVE GENOMICS OF CLINICAL ISOLATES OF DIARRHEAGENIC *Escherichia coli* FROM LIMA, PERUWilli Quino^{1,a,e}, Orson Mestanza^{1,b,d}, Junior Caro-Castro^{1,b,c}, Carmen Verónica Hurtado^{1,b,c}, Ronnie G. Gavilán^{1,2,b,f}¹ Instituto Nacional de Salud, Lima, Perú.² Professional School of Human Medicine, Universidad Privada San Juan Bautista, Lima, Perú.^a Medical Technologist; ^b Biologist; ^c Bachelor in Microbiology and Parasitology; ^d Master in Bioinformatics;^e Master in Microbiology; ^f Doctor of Biochemistry and Molecular Biology

ABSTRACT

In order to describe the genomic characteristics related to antimicrobial resistance and comparative genomics of diarrheagenic *Escherichia coli* (DEC), 14 DEC isolates from the strain collection of the Instituto Nacional de Salud (INS) were subjected to genome sequencing. We used bioinformatic procedures to analyze the obtained sequences in order to look for microbial resistance genes and genetic regions related to pathotypes and phylogroups. Several antimicrobial resistance determinants were detected, but the production of beta-lactamases and mutations associated to quinolone resistance were the most relevant. Additionally, we observed isolates of the same pathotype grouped in different phylogroups. The comparative genomics analysis showed a greater number of orthologous genes in isolates from the same pathotype and phylogroup. In conclusion, DEC isolates from Lima, Peru, showed resistance to multiple drugs; likewise, molecular and phylogenetic diversity was observed in several pathotypes and phylogroups.

Keywords: *Escherichia coli*; Drug resistance; Phylogeny; Genomics (Source: MeSH NLM).

INTRODUCTION

Bacterial resistance is a major global health problem associated with increased loss of economic productivity and human mortality. Approximately 23,000 people die annually in the United States due to infections caused by bacteria resistant to antibiotics⁽¹⁾, e.g., *Escherichia coli*, a bacterium intrinsically susceptible to almost all relevant antibiotics, with a high capacity to acquire resistance genes by horizontal gene transfer⁽²⁾.

The group of *E. coli* that cause intestinal and extra-intestinal infections are called diarrheagenic *Escherichia coli* (DEC). According to their pathogenesis, this group is classified into seven pathotypes: Enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC), diffusely adherent *E. coli* (DAEC) and adherent-invasive *E. coli* (AIEC), all of which are responsible for 40% of the cases of acute diarrhea in developing countries⁽³⁾.

En el Perú, la vigilancia de la resistencia bacteriana se realiza principalmente mediante métodos fenotípicos como el de Kirby-Bauer o la concentración mínima inhibitoria (MIC) recomendadas por el Instituto de Estándares Clínicos y de Laboratorio (CLSI). La mayoría de los estudios se avocan a la detección fenotípica de betalactamasas de espectro extendido (BLEE) y resistencia a quinolonas, aunque en años recientes se viene implementando el estudio de genes de resistencia por métodos moleculares, como el PCR y la secuenciación Sanger⁽⁴⁾.

In recent years, the use of whole genome sequencing has been employed to characterize virulence genes of different DECs and to identify phylogenetic relationships between pathotypes⁽⁵⁾. Since this aspect has been little explored in Peru, the aim of this research is to des-

Cite as: Quino W, Mestanza O, Caro-Castro J, Hurtado CV, Gavilán RG. Resistome and comparative genomics of clinical isolates of diarrheagenic *Escherichia coli* from Lima, Peru. Rev Peru Med Exp Salud Publica. 2020;37(4):705-10. doi: <https://doi.org/10.17843/rpmesp.2020.374.5240>.

Correspondence: Willi Quino Sifuentes; Cápac Yupanqui 1400, Jesús María, Lima, Perú; williqs@yahoo.com

Received: 11/02/2020

Approved: 19/08/2020

Online: 18/11/2020

cribe the characteristics of the resistome and comparative genomics of different DEC from the city of Lima.

THE STUDY

We worked on 14 *E. coli* isolates that caused acute diarrhea in patients from Lima between 2017 and 2018. The isolates were recovered by the Instituto Nacional de Salud (INS). The selection criteria were based on the antimicrobial sensitivity profile by diffusion disc previously evaluated by the laboratory (Appendix 1).

The isolates were reactivated in tryptic soybean broth (Merck, Germany) at 37 °C for 68 hours. Subsequently, they were seeded in MacConkey (Merck, Germany) agar plates and incubated at 37 °C for 18 to 24 hours. Confirmation of *E. coli* was made with conventional biochemical tests. The pathotypes were characterized based on primers according to Appendix 2.

DNeasy Blood & Tissue Kit (Qiagen, Germany) was used to extract DNA from the isolates. Spectrophotometry (Denovix, USA) and Qubit 3.0 fluorometer (Invitrogen, Malaysia) were used to evaluate DNA concentration and quality. Sequencing libraries were created with Nextera XT kit (Illumina, USA), and genomic sequencing with MiSeq equipment (Illumina, USA).

We evaluated the quality of the sequences using FastQC v0.11.5. Adapters and low-quality nitrogenous bases were removed with Trimmomatic v0.38⁽⁶⁾. The sequences were assembled *de novo* using A5-miseq pipeline⁽⁷⁾. We used Kraken for gender identification and removal of contaminated contigs⁽⁸⁾. Allelic profiles and sequence types (ST) of the obtained genomes were assigned according to the Multilocus Sequence Typing (MLST) database for *E. coli* according to Achtman (<http://pubmlst.org/>) and using the MLST v2.10 program. The assignment of clone complexes and the creation of a minimum spanning tree (MST) was carried out using BioNumerics v7.5 (Applied Maths).

The BLAST tool was used to search for the genes associated with the virulence of *E. coli* pathotypes, according to Appendix 2. Prediction of coding sequences was made using the program Prodigal v2.6.3. The homologous genes of the sequences were identified from a gene library built using all the complete genomes available in GenBank, and the BLAST algorithm to select the alignments greater than 90% for identity and greater than 60% for coverage than the reference. The code used for the annotation is available at <http://github.com/OrsonMM/Blast-score-for-genomics>.

For the assignment of the phylogroups, we used Clermont typing scheme⁽⁹⁾, based on an *in silico* PCR using

KEY MESSAGES

Motivation for the study: There has been an increase in gastrointestinal infections by antibiotic-resistant diarrheagenic *Escherichia coli* (DEC) in recent years, whose genomic characterization has been scarcely researched in Peru.

Main findings: The pathotypes of DEC from Lima, Peru, with determinants of resistance to beta-lactams and quinolones are diverse, both molecularly and phylogenetically.

Implications: Genome sequencing allows a deep understanding of the continuous changes and adaptations that pathogens like DEC undergo, so its implementation would be important for the surveillance of antimicrobial resistance in Peru.

primers designed for each phylogroup, with the software available at <https://github.com/A-BN/Clermon-Typing>. For the detection of antibiotic resistance genes, we used online database CARD⁽¹⁰⁾, which includes chromosomal and plasmid genes. Finally, all the sequences obtained during the study were stored in GenBank database (Bioproject: PRJ-NA650130).

For comparative genomics, genomes assembled with Prokka v1.12⁽¹¹⁾ were scored with an e value of 10⁻⁹. The annotation results were uploaded to the BPGA v1.3⁽¹²⁾ program to calculate the number of genes in the central, accessory, and single genes with an identity value of >0.95. As a final step, the resulting matrix of pathotype annotation was used for intergroup comparative analysis using VennPainter⁽¹³⁾.

FINDINGS

All 14 isolates were reconfirmed as *E. coli* by conventional microbiological procedures. PCR confirmation of pathotypes resulted in two EHEC, two EIEC, three EPEC and seven ETEC (Table 1).

The sequencing error rate of 0.35% was calculated using a PhiX internal control. The Q30 used to select good quality readings was >80% per 150 bp. The fourteen genomes sequenced had an average of 184 high quality contigs. Each genome was identified as *E. coli* with more than 85% identity, obtaining a GC-content of 50.4%. Ten different STs were detected by MLST (Figure 2), the main findings were that ST11 was associated to EHEC, ST311 to EIEC and ST69 to EPEC.

The *in silico* pathotype analysis found a direct relationship with the results obtained by PCR (Figure 1). The EHEC

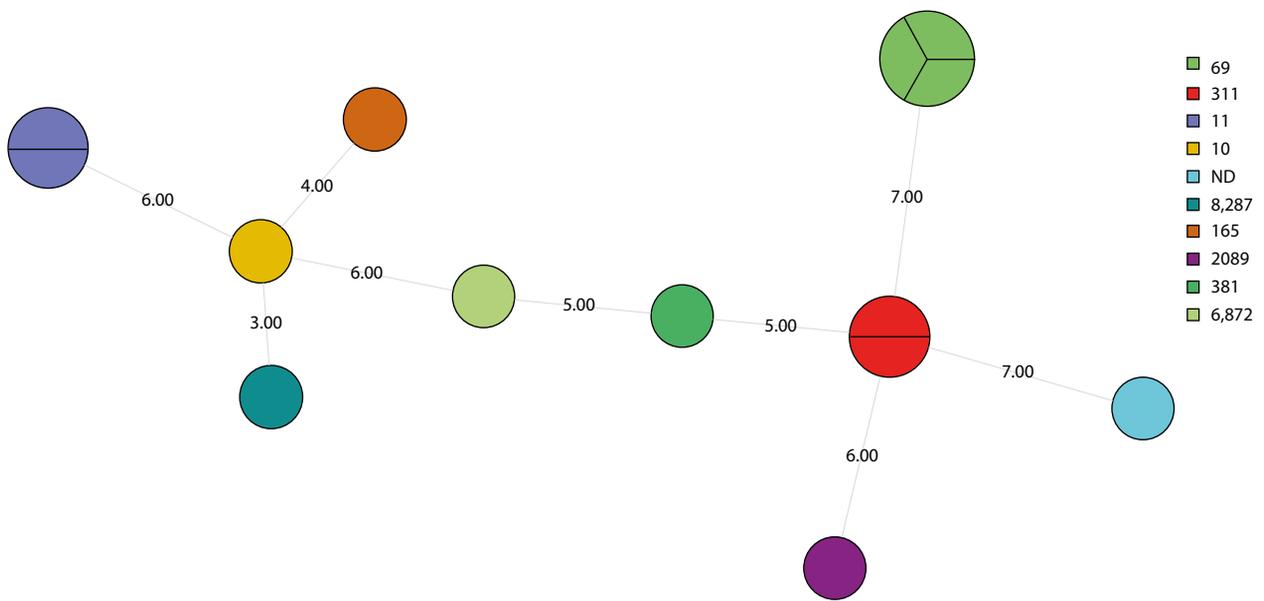


Figure 2. Minimum spanning tree of 14 allelic profiles of the multilocus sequence typing (MLST) of the diarrheagenic *Escherichia coli* included in this study, according to Achtman and designed with the BioNumerics v7.5 program. The caption indicates each sequence differentiated by color. Each circle represents a genotype of the MLST, and the size is proportional to the number of strains included in each one.

Finally, we detected a large number of ortho-genes between EHEC and EIEC isolates, as well as a small number of genes shared within EPEC and ETEC, for which there is no additional information on their name or function, being noted as hypothetical proteins.

DISCUSSION

Diarrheagenic *E. coli* is the leading cause of healthcare-related and community-acquired infections ⁽¹⁴⁾ and has a high capacity to develop antimicrobial resistance. It is important

Table 1. Genomic data of diarrheagenic *Escherichia coli* isolates in Lima, Peru.

Pathotype	Isolates	ST	CDS number	Number of genes in the central genome	Number of accessory genes	Number of single genes
EHEC	O157H7-1	11	5,270	2,950	2,133	16
	O157H7-2	11	5,215	2,950	2,104	9
EIEC	1-756-17	311	5,330	2,950	1,856	172
	2-683-18	311	5,112	2,950	1,804	15
EPEC	2-492-17	381	5,248	2,950	1,588	471
	2-537-17	ND	4,721	2,950	938	581
	2-734-18	2,089	4,922	2,950	1,503	336
	2-675-17	10	4,661	2,950	1,369	273
	2-858-17	165	5,003	2,950	1,552	386
ETEC	2-442-18	6,872	4,991	2,950	1,473	449
	2-462-18	69	5,053	2,950	2,015	0
	2-526-18	69	5,056	2,950	2,016	6
	2-539-18	69	5,055	2,950	2,011	7
	2-727-18	8,287	4,593	2,950	1,199	294

EHEC: Enterohemorrhagic *E. coli*, EIEC: Enteroinvasive *E. coli*, EPEC: Enteropathogenic *E. coli*, ETEC: Enterotoxigenic *E. coli*, ST: Sequence type, ND: Not defined, not found in the Multilocus Sequence Typing (MLST) database.

to report the genomic characteristics of pathogenic isolates to understand bacterial resistance and pathogenesis features that will contribute to monitoring and implementation of new policies related to local public health.

In this study, nine *E. coli* isolates with antibiotic resistance genes were detected by genomic sequencing. The most important ones were CTX-M-15, TEM1 and *ampC*, because they are related to beta-lactam resistance. Studies indicate the presence of *bla* genes in Peruvian ESBL positive *E. coli* isolates, with high frequency rates of CTX-M (54.7%) and TEM (13.2%)⁽¹⁵⁾. On the contrary, the detection of *ampC*-type beta-lactamase is little explored in our region, and usually only a low-frequency is phenotypically detected⁽¹⁶⁾.

Resistance to quinolones by the S83L mutation in *gyrA* was observed in two ETCEs, while the presence of *qnr* in two EIECs, one EPEC and one ETEC. Quinolone resistance has already been reported in Peru, including commensal strains due to the strong influence of persistent exposure to these antimicrobials⁽¹⁷⁾; however, few studies focused on the use of molecular markers to determine quinolone resistance. Furthermore, genomic sequencing has led to the discovery of additional genes related to resistance to other antibiotics. For example, two EIECs and two EPECs with STX resistance and nine isolates with tetracycline resistance have been reported.

The analysis of the relationship between pathotypes (non-taxonomic) and phylogroups (taxonomic) of *E. coli* showed that the mostly reported EPECs were from phylogroups B1 and B2⁽¹⁸⁾, as discussed here. The EIEC are grouped in three phylogroups: A, E and B1⁽¹⁴⁾, the isolates we used were from the phylogroup B1. On the other hand, the ETCEs

are mostly from the A and D phylogroups⁽¹⁹⁾, the seven isolates evaluated were also from this phylogroups. In contrast, the EHECs were the most consistent group and were only associated with phylogroup E, and there is an evolutive explanation: the predominant serotype in this group, O157: H7, derives from the gene acquisition of a non-pathogenic isolate O55: H7, called preEHEC⁽²⁰⁾.

Pan-genomic analysis of diarrheagenic *E. coli* using KEGG categories allowed us to detect 2.17% of the genes (from the central genome) that are involved in antimicrobial resistance. Although the previous analysis only showed information regarding 14 genes associated with resistance, we were able to detect additional resistance-associated genes (among mutations and other effectors); however, due to the little information available regarding these genes in *E. coli* isolates, we did not study those molecules in depth.

The comparative analysis between pathotypes showed that EHEC and EIEC share a large number of genes ($n = 246$), while EPEC and ETEC had the lowest number of shared genes ($n = 2$). The latter result is interesting, since it contrasts with the work done by Hazen *et al.*⁽¹⁸⁾, who found hybrid isolations of these pathotypes. On the contrary, most comparative genome studies are only based on the analysis of differences and similarities within the same pathotypes, so our research offers a first look at comparative genomes between different pathotypes and is one of the first studies on this subject in Peru.

In conclusion, the isolates of diarrheagenic *E. coli* from Lima, Peru, have resistance genes for several drugs. We detected pathotypes, ST and phylogroups with molecular and phylogenetic diversity. We suggest to continue monitoring the antimicrobial resistance of these bacteria with modern sequencing techniques, since this information is useful for local public health.

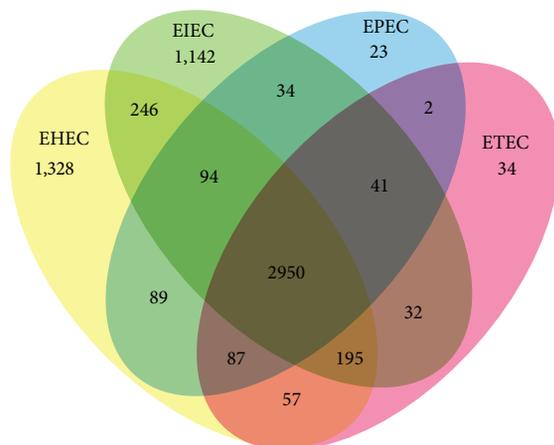
Acknowledgements: To the members of Laboratorio Nacional de Referencia de Enteropatógenos/INS, especially to María Luz Zamudio, Gustavo Bellido and Ana María Meza for their technical assistance.

Authors' contribution: WQ and RG participated in the conception, hypothesis formulation and study design. OM, JC, and CH participated in the analysis, data interpretation and writing of the article. WQ and RG participated in the critical review of the article.

Funding: The research was funded by the Instituto Nacional de Salud in the context of the project "Desarrollo de plataformas para el diagnóstico y vigilancia molecular de infecciones relacionadas con síndrome febril y síndrome diarreico agudo" (Development of platforms for the diagnosis and molecular surveillance of infections related to febrile syndrome and acute diarrheal syndrome), approved with RD 2012016-OGITT-OPE/INS.

Conflicts of interest: All authors declare that they have no conflict of interest regarding this publication.

Supplementary material: Available in the electronic version of the RPMESP.



EHEC: Enterohemorrhagic *E. coli*, EIEC: Enteroinvasive *E. coli*, EPEC: Enteropathogenic *E. coli*, ETEC: Enterotoxigenic *E. coli*

Figure 3. Comparative genomics of *Escherichia coli* pathotypes. The central intersection indicates the genes shared by the four pathotypes (central genome).

REFERENCES

- Marston HD, Dixon DM, Knisely JM, Palmore TN, Fauci AS. Antimicrobial Resistance. *JAMA*. 2016 20;316(11):1193-1204. doi: 10.1001/jama.2016.11764.
- Poirel L, Madec JY, Lupo A, Schink AK, Kieffer N, Nordmann P, *et al*. Antimicrobial Resistance in *Escherichia coli*. *Microbiology Spectrum*. 2018;6(4). doi: 10.1128/microbiolspec.ARBA-0026-2017.
- O’Ryan M, Prado V, Pickering LK. A millennium update on pediatric diarrheal illness in the developing world. *Semin Pediatr Infect Dis*. 2005;16(2):125-36. doi: 10.1053/j.spid.2005.12.008.
- Ochoa TJ, Ruiz J, Molina M, Valle LJ, Vargas M, Gil AI, *et al*. High Frequency of Antimicrobial Drug Resistance of Diarrheagenic *Escherichiacoli* in Infants in Peru. *AmJTropMedHyg*.2009;81(2):296-301. doi: 10.4269/ajtmh.2009.81.296.
- Noll LW, Worley JN, Yang X, Shridhar PB, Ludwig JB, Shi X, *et al*. Comparative genomics reveals differences in mobile virulence genes of *Escherichia coli* O103 pathotypes of bovine fecal origin. *PLoS One*. 2018;13(2):e0191362. doi: 10.1371/journal.pone.0191362.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20. doi: 10.1093/bioinformatics/btu170.
- Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*. 2015;31(4):587-9. doi: 10.1093/bioinformatics/btu661.
- Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*. 2014;15(3):R46. doi: 10.1186/gb-2014-15-3-r46.
- Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O. ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb Genom*. 2018;4(7):e000192. doi: 10.1099/mgen.0.000192.
- Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, *et al*. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2017;45(D1):D566-73. doi: 10.1093/nar/gkw1004.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-9. doi: 10.1093/bioinformatics/btu153.
- Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. *Scientific Reports*. 2016;6:24373. doi: 10.1038/srep24373.
- Lin G, Chai J, Yuan S, Mai C, Cai L, Murphy RW, *et al*. VennPainter: A Tool for the Comparison and Identification of Candidate Genes Based on Venn Diagrams. *PLoS ONE*. 2016;11(4):e0154315. doi: 10.1371/journal.pone.0154315.
- Oteo J, Lázaro E, de Abajo FJ, Baquero F, Campos J. Antimicrobial-resistant Invasive *Escherichia coli*, Spain. *Emerg Infect Dis*. 2005;11(4):546-53. doi: 10.3201/eid1104.040699.
- García C, Astocondor L, Banda C. Enterobacterias productoras de beta-lactamasas de espectro extendido: Situación en América Latina y en el Perú. *Acta Med Peru*. 2012;29(3):163-9.
- Galván F, Agapito J, Bravo N, Lagos J, Tamariz J. Caracterización fenotípica y molecular de *Escherichia coli* productoras de β -Lactamasas de espectro extendido en pacientes ambulatorios de Lima, Perú. *Rev Med Hered*. 2016;27(1):22-9. doi: 10.20453/rmh.v27i1.2780.
- Pons MJ, Mosquito S, Ochoa TJ, Vargas M, Molina M, Lluque A, *et al*. Niveles de resistencia a quinolonas y otros antimicrobianos en cepas de *Escherichia coli* comensales en niños de la zona periurbana de Lima, Perú. *Rev Peru Med Exp Salud Publica*. 2012;29(1):82-6.
- Hazen TH, Daugherty SC, Shetty AC, Nataro JP, Rasko DA. Transcriptional Variation of Diverse Enteropathogenic *Escherichia coli* Isolates under Virulence-Inducing Conditions. *mSystems*. 2017;2(4):e00024-17. doi: 10.1128/mSystems.00024-17.
- Lu S, Jin D, Wu S, Yang J, Lan R, Bai X, *et al*. Insights into the evolution of pathogenicity of *Escherichia coli* from genomic analysis of intestinal *E. coli* of Marmota himalayana in Qinghai–Tibet plateau of China. *Emerg Microbes Infect*. 2016;5(12):e122. doi: 10.1038/emi.2016.122.
- Zhou Z, Li X, Liu B, Beutin L, Xu J, Ren Y, *et al*. Derivation of *Escherichia coli* O157:H7 from Its O55:H7 Precursor. *PLoS One*. 2010;5(1):e8700. doi: 10.1371/journal.pone.0008700.