

## ORIGINAL BREVE

RESISTOMA Y GENÓMICA COMPARATIVA DE AISLADOS CLÍNICOS DE *Escherichia coli* DIARREOGÉNICA EN LIMA, PERÚWilli Quino<sup>1,a,e</sup>, Orson Mestanza<sup>1,b,d</sup>, Junior Caro-Castro<sup>1,b,c</sup>, Carmen Verónica Hurtado<sup>1,b,c</sup>, Ronnie G. Gavilán<sup>1,2,b,f</sup><sup>1</sup> Instituto Nacional de Salud, Lima, Perú.<sup>2</sup> Escuela Profesional de Medicina Humana, Universidad Privada San Juan Bautista, Lima, Perú<sup>a</sup> Tecnólogo médico; <sup>b</sup> biólogo/a; <sup>c</sup> bachiller en Microbiología y Parasitología; <sup>d</sup> magíster en Bioinformática; <sup>e</sup> magíster en Microbiología; <sup>f</sup> doctor en Bioquímica y Biología Molecular

## RESUMEN

Con el objetivo de describir las características genómicas relacionadas con la resistencia antimicrobiana y genómica comparativa de *Escherichia coli* diarreogénica (DEC), se sometieron a secuenciación genómica catorce aislamientos de DEC del banco de cepas del Instituto Nacional de Salud (INS). Las secuencias obtenidas se analizaron mediante procedimientos bioinformáticos a fin de buscar genes de resistencia microbiana y regiones genéticas relacionadas con patotipos y filogrupos. Se detectaron diversos determinantes de resistencia antimicrobiana, se destaca la producción de betalactamasas y mutaciones asociadas a la resistencia a quinolonas. Además, se observaron aislamientos de un mismo patotipo agrupados en distintos filogrupos. El análisis de genómica comparativa mostró un mayor número de genes ortólogos en aislamientos que pertenecían al mismo patotipo y filogrupo. Sobre la base de lo estudiado, los aislamientos de DEC en Lima, Perú, presentan resistencia a múltiples fármacos, y se detectaron varios patotipos y filogrupos con diversidad molecular y filogenética.

**Palabras clave:** *Escherichia coli*; Resistencia a Medicamentos; Filogenia; Genómica (Fuente: DeCS BIREME).

RESISTOME AND COMPARATIVE GENOMICS OF CLINICAL ISOLATES OF DIARRHEAGENIC *Escherichia coli* FROM LIMA, PERU

## 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).

## INTRODUCCIÓN

La resistencia bacteriana se constituye como un importante problema de salud mundial, asociada a un incremento en la pérdida de productividad económica y mortalidad humana. Se estima el deceso anual de aproximadamente 23 000 personas en los Estados Unidos como resultado de infecciones por bacterias resistentes a antibióticos<sup>(1)</sup>, destacando entre ellas *Escherichia coli*, una bacteria intrínsecamente susceptible a casi todos los antibióticos relevantes, pero con alta capacidad de adquirir genes de resistencia por transferencia genética horizontal<sup>(2)</sup>.

*E. coli* puede causar infecciones intestinales y extraintestinales, limitando su carácter patógeno a un grupo denominado *Escherichia coli* diarreogénica (DEC), que se clasifica de-

**Citar como:** Quino W, Mestanza O, Caro-Castro J, Hurtado CV, Gavilán RG. Resistoma y genómica comparativa de aislados clínicos de *Escherichia coli* diarreogénica en Lima, Perú. Rev Peru Med Exp Salud Publica. 2020;37(4):705-710. doi: <https://doi.org/10.17843/rpmesp.2020.374.5240>.

**Correspondencia:** Willi Quino Sifuentes; Cápac Yupanqui 1400, Jesús María, Lima, Perú; [williqs@yahoo.com](mailto:williqs@yahoo.com)

**Recibido:** 11/02/2020  
**Aprobado:** 19/08/2020  
**En línea:** 18/11/2020

pendiendo de su patogénesis en siete patotipos: *E. coli* enteropatógena (EPEC), enterohemorrágica (EHEC), enterotoxigénica (ETEC), enteroinvasiva (EIEC), enteroagregativa (EAEC), adherente difusa (DAEC) y adherente invasiva (AIEC), responsables del 40% de los casos de diarrea aguda en países en vías de desarrollo <sup>(3)</sup>.

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 <sup>(4)</sup>.

En años recientes, se viene apostando por el uso del secuenciamiento del genoma completo para caracterizar genes de virulencia de diversas DEC e identificar relaciones filogenéticas entre patotipos <sup>(5)</sup>. Esta área ha sido poco explorada en el Perú, por lo que el objetivo del presente trabajo es describir las características del resistoma y genómica comparativa de diferentes DEC de la ciudad de Lima.

## EL ESTUDIO

Se trabajó con 14 aislamientos de *E. coli* recuperados por el Instituto Nacional de Salud (INS), con antecedente de haber provocado enfermedad diarreica aguda en pacientes de la ciudad de Lima entre 2017 y 2018, tomando como criterio de selección el perfil de sensibilidad antimicrobiana por disco difusión previamente evaluada por el laboratorio (Anexo 1).

Los aislamientos se reactivaron en caldo tripticosa soya (Merck, Alemania) a 37 °C durante 68 horas. Posteriormente, se sembraron en placas de agar MacConkey (Merck, Alemania) y se incubaron a 37 °C durante 18 a 24 horas. Se confirmó *E. coli* mediante pruebas bioquímicas convencionales. Los patotipos fueron caracterizados en base a cebadores según el Anexo 2.

La extracción del ADN de los aislamientos se realizó con el kit DNeasy Blood & Tissue (Qiagen, Alemania). Se evaluaron la concentración y la calidad del ADN mediante espectrofotometría (Denovix, EE. UU.) y con un fluorómetro Qubit 3.0 (Invitrogen, Malasia). La elaboración de bibliotecas de secuenciamiento se realizó con el kit Nextera XT (Illumina, EE. UU.), y el secuenciamiento genómico con el equipo MiSeq (Illumina, EE. UU.).

Se evaluó la calidad de las secuencias obtenidas con FastQC v0.11.5. Se removieron los adaptadores y las bases nitrogenadas de baja calidad con el programa Trimmomatic v0.38 <sup>(6)</sup>. Las secuencias se ensamblaron *de novo* empleando el pipeline A5-miseq <sup>(7)</sup>. La identificación del género y remoción de *contigs* contaminados se realizó con el programa Kraken <sup>(8)</sup>. Los perfiles alélicos y los secuenciotipos (ST, por sus siglas en inglés, *sequence type*) de los genomas obtenidos se asignaron de acuerdo con la base de datos del esquema de tipificación multilocus (MLST,

## MENSAJES CLAVE

**Motivación para realizar el estudio:** Se ha observado un incremento de infecciones gastrointestinales por *Escherichia coli* diarreogénicas (DEC) resistentes a antibióticos en los últimos años, cuya caracterización genómica viene siendo poco explorada en el Perú.

**Principales hallazgos:** Los patotipos de DEC de Lima, Perú, poseedores de determinantes de resistencia a betalactámicos y quinolonas son diversos, tanto de manera molecular como filogenética.

**Implicancias:** El secuenciamiento genómico permite una comprensión profunda de los continuos cambios y adaptaciones que sufren patógenos como DEC, por lo que su implementación sería importante para la vigilancia de la resistencia antimicrobiana en el Perú.

por sus siglas en inglés, *multilocus sequence typing*) para *E. coli* según Achtman (<http://pubmlst.org/>) utilizando el programa MLST v2.10. La asignación de complejos clonales y la creación de un árbol recubridor mínimo (MST, por sus siglas en inglés, *minimum spanning tree*) se realizó mediante el programa BioNumerics v7.5 (Applied Maths).

Con la herramienta BLAST se buscaron los genes asociados a la virulencia de los patotipos de *E. coli*, según el Anexo 2. Se realizó la predicción de secuencias codificantes usando el programa Prodigal v2.6.3. Se identificaron los genes homólogos de las secuencias a partir de una biblioteca de genes construida utilizando todos los genomas completos disponibles en el GenBank, utilizando el algoritmo BLAST seleccionando los alineamientos mayores de 90% de identidad y cobertura mayor al 60% con la referencia. El código empleado para la anotación se encuentra disponible en <http://github.com/OrsonMM/Blast-score-ratio-for-genomics>.

Para la asignación de los filogrupos, se empleó el esquema de tipificación de Clermont <sup>(9)</sup>, basado en una PCR *in silico* utilizando cebadores diseñados para cada filogrupo con el *software* disponible en <https://github.com/A-BN/Clermont-Typing>. Para la detección de genes de resistencia antibiótica, se utilizó la base de datos online CARD <sup>(10)</sup> que incluye genes cromosomales y plasmídicos. Finalmente, todas las secuencias obtenidas durante el estudio se depositaron en la base de datos del GenBank (Bioproject: PRJNA650130).

En cuanto a la genómica comparativa, se anotaron los genomas ensamblados con Prokka v1.12 <sup>(11)</sup> con un valor  $e$  de  $10^{-9}$ . Los resultados de la anotación se subieron al programa BPGA v1.3 <sup>(12)</sup> para calcular el número de genes del genoma central, accesorio y genes únicos con un valor  $>0,95$  de identidad. Como último paso, se utilizó la matriz resultante de anotación por patotipo para el análisis comparativo entre grupos utilizando VennPainter <sup>(13)</sup>.

### HALLAZGOS

Los 14 aislamientos fueron reconfirmados como *E. coli* mediante procedimientos microbiológicos convencionales. La confirmación de patotipos por PCR dio como resultado dos EHEC, dos EIEC, tres EPEC y siete ETEC (Tabla 1).

La tasa de error de secuenciamiento evaluado mediante el uso de un control interno PhiX indicó una tasa del 0,35%. El Q30 empleado para seleccionar lecturas de buena calidad fue >80% por cada 150 pb. Los catorce genomas secuenciados tienen en promedio 184 *contigs* de alta calidad. Cada genoma se identificó como *E. coli* con más del 85% de identidad, obteniendo un contenido de GC del 50,4%. Se detectaron 10 ST diferentes por MLST (Figura 2), teniendo como principales hallazgos al ST11 asociado a EHEC, ST311 a EIEC y ST69 a EPEC.

El análisis *in silico* de patotipos encontró una relación directa con los resultados obtenidos por PCR (Figura 1). Los

aislamientos EHEC presentaron los genes de la toxina Shiga (*stx1* y *stx2*). Cuatro ETEC presentaron el gen de la enterotoxina termolábil (*elt*), dos presentaron la enterotoxina termoestable (*st*) y solo uno presentó ambos genes. La anotación reveló la presencia del gen *eaeA* dentro de las EPEC, y finalmente las EIEC poseían el gen *ipaH*.

El esquema de tipificación de Clermont asignó cuatro genomas de *E. coli* al filogrupo A, cuatro a B1, uno a B2, tres a D y dos a E. Una revisión adicional reveló que los aislamientos EHEC peruanos pertenecían al filogrupo E, las EIEC a B1, las ETEC a A y D, y las EPEC a los filogrupos B1, B2 y C (Figura 1).

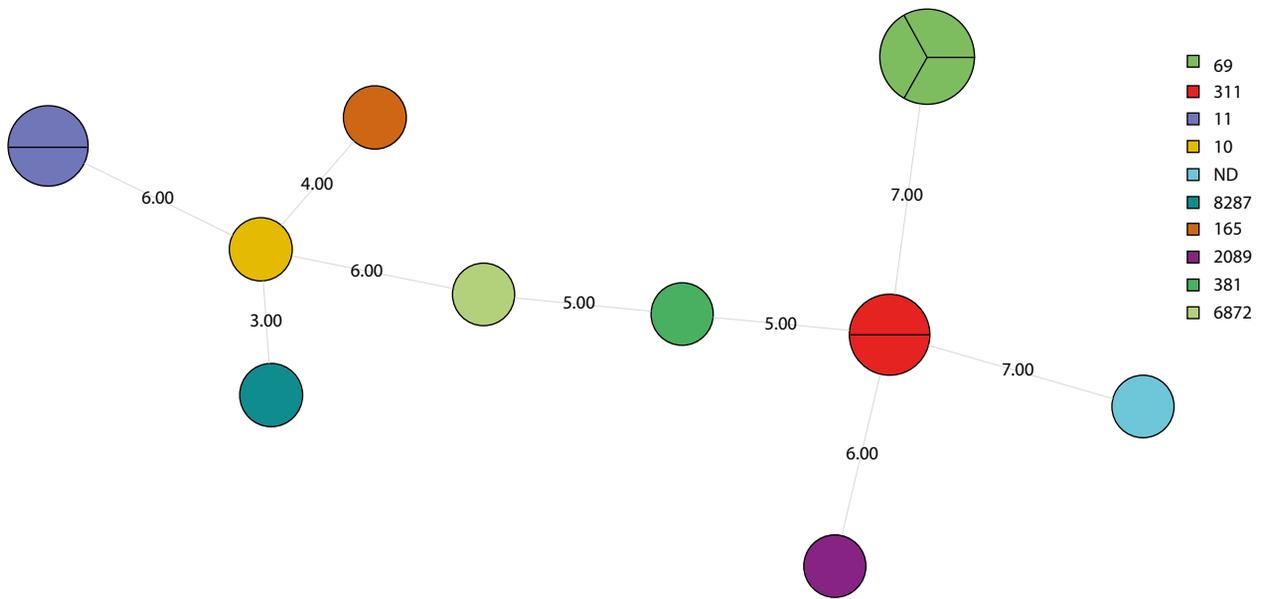
En el análisis relacionado con el resistoma, se detectaron genes que codifican betalactamasas; nueve genomas presentaron el gen *ampC*; cuatro poseían *bla* TEM1, y solo uno presentó *bla* CTX-M-15. Además, se detectaron genes de resistencia a las quinolonas. El gen *gyrA* con la mutación S83L se detectó en dos ETEC. El gen *qnrS1* se encontró en

Aislados	Patotipo	Filogrupo	ST	Identificación de determinantes de resistencia antimicrobiana												Identificación alélica de patotipo						Identificación alélica de filogrupo						
				<i>bla</i> TEM1	<i>bla</i> CTX-M-15	<i>ampC</i>	<i>qnrS1</i>	<i>qnrB10</i>	<i>gyrA</i> (S83L*)	<i>dfpA8/14</i>	<i>sul2</i>	<i>marR</i> (G103S, Y137H*)	<i>nfsA</i> (Y45C*)	PBP3 (S357N, D350N*)	<i>GlpT</i> (E48K*)	EF-Tu (R234F*)	Tet(A)	<i>eaeA</i>	<i>elt</i>	<i>stx1</i>	<i>stx2</i>	<i>ipaH</i>	<i>aggR</i>	<i>arpA</i>	<i>chuA</i>	<i>yjaA</i>	TspE4-C2	
O157H7-1	EHEC	E	11																									
O157H7-2		E	11																									
1-756-17	EIEC	B1	311																									
2-683-18		B1	311																									
2-492-17	EPEC	B1	381																									
2-537-17		B2	ND																									
2-734-18		C <sup>a</sup>	2089																									
2-675-17	ETEC	A <sup>a</sup>	10																									
2-858-17		A	165																									
2-442-18		A	6872																									
2-462-18		D	69																									
2-526-18		D	69																									
2-539-18		D	69																									
2-727-18		A	8287																									

<sup>a</sup> Esto se validó usando el gen adicional *trpAgpC*.

EHEC: *E. coli* enterohemorrágica, EIEC: *E. coli* enteroinvasiva, EPEC: *E. coli* enteropatógena, ETEC: *E. coli* enterotoxigénica, ND: No definido, no encontrado en la base de datos del esquema de tipificación *multilocus*.

**Figura 1.** Identificación molecular de genes asociados a patotipos, filogrupos y genes de resistencia antimicrobiana de *Escherichia coli* diarregénica. Las casillas coloreadas indican la detección de los marcadores mientras que las incoloras indican su ausencia.



**Figura 2.** Árbol recubridor mínimo de 14 perfiles alélicos del esquema de tipificación multilocus (MLST) de *Escherichia coli* diarreogénicas según Achtman incluidos en este estudio, diseñado con el programa BioNumerics v7.5. La leyenda indica cada secuenciotipo diferenciado por colores. Cada círculo representa un genotipo del MLST y el tamaño es proporcional al número de cepas que incluye cada uno.

todas las EIEC y en una ETEC, mientras que el gen *qnrB10* se presentó en una EPEC (Figura 1).

El promedio de secuencias codificantes para los catorce genomas analizados fue de 5016. El número de genes del genoma accesorio varió entre 938 y 2133, mientras que el número de genes únicos varió entre 0 y 581. Finalmente, el número de genes del genoma central fue 2950 (Tabla 1).

El análisis funcional pangénomico detectó 2,17% de genes centrales, 2,04% de genes accesorios y 1,63% de genes únicos relacionados con la resistencia antimicrobiana usando las categorías de la enciclopedia de genes y genomas de Kyoto (KEGG) (Anexo 3). Además, la anotación empleando las categorías de clústeres de grupos ortólogos de proteínas (COG) detectaron que el 5,32% de los genes centrales, el

**Tabla 1.** Datos genómicos de los aislamientos de *Escherichia coli* diarreogénica en Lima, Perú.

Patotipo	Aislamientos	ST	Número de CDS	Número de genes del genoma central	Número de genes accesorios	Número de genes únicos
EHEC	O157H7-1	11	5270	2950	2133	16
	O157H7-2	11	5215	2950	2104	9
EIEC	1-756-17	311	5330	2950	1856	172
	2-683-18	311	5112	2950	1804	15
EPEC	2-492-17	381	5248	2950	1588	471
	2-537-17	ND	4721	2950	938	581
	2-734-18	2089	4922	2950	1503	336
	2-675-17	10	4661	2950	1369	273
	2-858-17	165	5003	2950	1552	386
	2-442-18	6872	4991	2950	1473	449
ETEC	2-462-18	69	5053	2950	2015	0
	2-526-18	69	5056	2950	2016	6
	2-539-18	69	5055	2950	2011	7
	2-727-18	8287	4593	2950	1199	294

EHEC: *E. coli* enterohemorrágica, EIEC: *E. coli* enteroinvasiva, EPEC: *E. coli* enteropatógena, ETEC: *E. coli* enterotoxigénica, ST: Sequence type, ND: No definido, no encontrado en la base de datos del esquema de tipificación multilocus (MLST).

7,32% de los genes accesorios y el 10,22% de los genes únicos estaban relacionados con la pared celular y las proteínas de membrana (Anexo 4).

La genómica comparativa entre patotipos reveló 2950 genes del genoma central. Datos adicionales de agrupación de dos o tres patotipos se detallan en la Figura 3. Finalmente, detectamos una gran cantidad de genes ortólogos entre aislados EHEC y EIEC, así como una pequeña cantidad de genes compartidos dentro de EPEC y ETEC, para los cuales no hay información adicional sobre su nombre o función, siendo anotados como proteínas hipotéticas.

## DISCUSIÓN

La *E. coli* diarreogénica es la principal causa de infecciones nosocomiales y adquiridas en la comunidad<sup>(14)</sup>, y tiene una alta capacidad para desarrollar resistencia a los antimicrobianos. Es importante reportar las características genómicas de aislamientos patógenos para comprender la resistencia y patogénesis bacteriana que contribuirán con el monitoreo y la aplicación de nuevas acciones relacionadas con la salud pública local.

En el presente estudio, se detectaron nueve aislamientos de *E. coli* poseedores de genes de resistencia antibiótica por secuenciación genómica, siendo la más importante la presencia de CTX-M-15, TEM1 y *ampC* relacionados con la resistencia a betalactámicos. Los estudios indican la presencia de genes *bla* en aislados peruanos de *E. coli* BLEE-positivo, con altas tasas de frecuencia de CTX-M (54,7%) y TEM (13,2%)<sup>(15)</sup>. Por otro lado, la detección de la betalactamasa tipo *ampC* es poco explorada en nuestra región, siendo principalmente una detección fenotípica con bajas frecuencias<sup>(16)</sup>.

La resistencia a quinolonas por la mutación S83L en *gyrA* se presentó en dos ETEC, mientras que la presencia

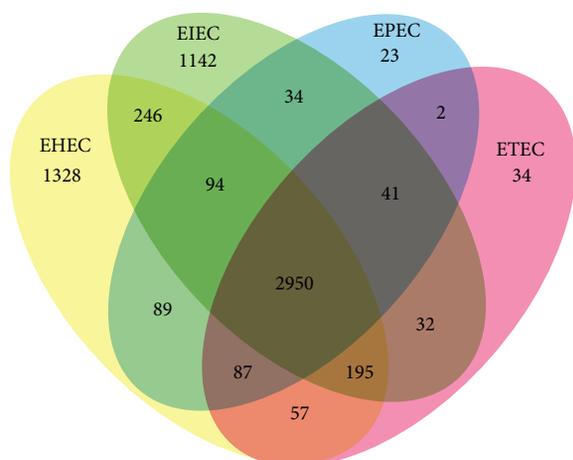
de *qnr* se detectó en dos EIEC, una EPEC y una ETEC. La resistencia a las quinolonas ya se había informado en el Perú, incluyendo cepas comensales debido a la fuerte influencia de la exposición persistente a estos antimicrobianos<sup>(17)</sup>; sin embargo, son pocos los estudios centrados en el uso de marcadores moleculares para determinar la resistencia a las quinolonas. Además, el secuenciamiento genómico permitió encontrar genes adicionales relacionados con la resistencia a otros antibióticos. Por ejemplo, se reporta dos EIEC y dos EPEC con resistencia a STX y nueve aislamientos con resistencia a tetraciclinas.

La exploración de la relación entre patotipos (no taxonómica) y filogrupos (taxonómica) de *E. coli* mostró que las EPEC han sido reportadas mayormente dentro de los filogrupos B1 y B2<sup>(18)</sup>, acorde con lo analizado aquí. Las EIEC se agrupan en tres filogrupos: A, E y B1<sup>(14)</sup>, siendo B1 donde se ubicaron los aislamientos de este estudio. Por otro lado, las ETEC se encuentra mayormente dentro de A y D<sup>(19)</sup>, donde los siete aislamientos evaluados aquí también encajaron. En contraste, las EHEC resultaron ser el grupo más uniforme, estando asociadas al filogrupo E. Esto tiene una explicación evolutiva; el serotipo predominante de este grupo, O157: H7, deriva de una adquisición de genes de un aislamiento no patógeno O55: H7, llamada preEHEC<sup>(20)</sup>.

El análisis pangenómico de *E. coli* diarreogénica usando las categorías KEGG permitió detectar un 2,17% de genes del genoma central involucrados en la resistencia antimicrobiana. Aunque en el análisis anterior solo se mostró información con respecto a 14 genes asociados con la resistencia, logramos detectar genes adicionales asociados con la resistencia, entre mutaciones y otros efectores; sin embargo, debido a la poca información disponible de estas moléculas en los aislamientos de *E. coli*, no profundizamos en el estudio de esas moléculas.

El análisis comparativo entre patotipos mostró que EHEC y EIEC comparten una gran cantidad de genes ( $n = 246$ ), mientras que EPEC y ETEC presentaron la menor cantidad de genes compartidos ( $n = 2$ ). Este último resultado es interesante, ya que contrasta con el trabajo realizado por Hazen *et al.*<sup>(18)</sup>, quienes encontraron aislamientos híbridos de estos patotipos. Por otro lado, la mayoría de los estudios de genómica comparativa solo se basan en el estudio de diferencias y similitudes dentro de los mismos patotipos, por lo que este trabajo ofrece una primera mirada a la genómica comparativa entre diferentes patotipos, siendo uno de los primeros trabajos de esta temática en el Perú.

En conclusión, los aislados de *E. coli* diarreogénica de Lima, Perú, poseen genes de resistencia a múltiples fármacos, se ha detectado patotipos, ST y filogrupos con diversidad molecular y filogenética. Se sugiere continuar vigilando la resistencia de estas bacterias a los antimicrobianos con el secuenciamiento de nueva generación, puesto que esta información es de utilidad para la salud pública local.



EHEC: *E. coli* enterohemorrágica, EIEC: *E. coli* enteroinvasiva, EPEC: *E. coli* enteropatogénica, ETEC: *E. coli* enterotoxigénica

**Figura 3.** Genómica comparativa de patotipos de *Escherichia coli*. La intersección central indica los genes compartidos por los cuatro patotipos (genoma central).

**Agradecimientos:** A los miembros del Laboratorio Nacional de Referencia de Enteropatógenos/INS, especialmente a María Luz Zamudio, Gustavo Bellido y Ana María Meza por su asistencia técnica.

**Contribución de los autores:** WQ y RG participaron en la concepción, delineación de hipótesis y diseño del estudio. OM, JC y CH participaron en el análisis, interpretación de datos y redacción del artículo. WQ y RG participaron en la revisión crítica del artículo.

**Financiamiento:** La investigación fue financiada por el Instituto Nacional de Salud en el marco del proyecto «Desarrollo de plataformas para el diagnóstico y vigilancia molecular de infecciones relacionadas con síndrome febril y síndrome diarreico agudo», aprobado con RD 2012016-OGITT-OPE/INS.

**Conflictos de interés:** Todos los autores declaran que no tienen ningún conflicto de interés en relación con esta publicación.

## REFERENCIAS BIBLIOGRÁFICAS

- Marston HD, Dixon DM, Knisely JM, Palmore TN, Fauci AS. Antimicrobial Resistance. *JAMA*. 2016;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 *Escherichia coli* in Infants in Peru. *Am J Trop Med Hyg*. 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, Rapheny AR, Alcock B, Wagelchner 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  $\beta$ -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.