Original Article

Determining the Sequence of Two Colibactin Genes, *clbN* and *clbB*, in Some Clinical Isolates of *E. coli*

Noor Ismeal Nasser, Mohammed Sabri Abdul-Razzag, Nadia Mudher AL-Hili[1]

Department of Medical Laboratory Techniques, Al-Kufa Technical institute, University Al-Furat Al-Awsat, Najaf, Iraq, [1]Department of Obsterics & Gynecology, College of Medicine, University of Babylon, Babylon, Iraq

Abstract

Background: *Escherichia coli* is a widespread human-infecting bacterium that can produce the genotoxin colibactin, which is encoded by the pks pathogenicity island and has been reported as a determinant of bacterial virulence. **Objective:** This study aimed to determine the presence of two colibactin markers in some isolates of *E. coli*. **Materials and Methods:** Seven of nine pk-positive *E. coli* samples were chosen to determine the genetic variants in the sequence of two specific polymerase chain reaction (PCR) fragments partially covering the *clbN* and clbB genes. **Results:** The results indicated the presence of only one nucleic acid variant in both clbN and clbB compared to their reference sequences (GenBank acc. no. CP109929.1). Concerning the *clbN* gene, the identified variant was 243C>T, which was suitable for position 1113 of the entire *clbN* gene-encoded colibactin nonribosomal peptide synthetase. In the *clbB* gene, the identified variant was 62C>T, which was located at position 1790 in the entire clbB-encoded colibactin hybrid nonribosomal peptide synthetase/type I polyketide synthase. Both variants were synonymous, namely, p.Thr1113= and p.F1790=. **Conclusion:** This study detected that variants in both genes did not affect colibactin production because they were silent mutations. These PCR products can be efficiently used to detect the biological diversity of a wider spectrum of *E. coli* sequences. Nevertheless, these PCR products can also be explored to discover further details of the identified bacterial sequences in various infections.

Keywords: *clb*B gene, *clb*N gene, colibactin, pks pathogenicity island

Introduction

Escherichia coli is an important pathogen that causes a wide range of diseases in the intestinal or extra-intestinal sites and displays high phenotypic and genetic diversity because of its ability to acquire different mobile genetic elements such as pks pathogenicity island that includes gene cluster (*clb A-clb S*) that acts together as an assembly line for the synthesis of colibactin, which has a genotoxic effect, and a virulence factor for pathogenic *E. coli*, which induces DNA interstrand cross-links and double-strand breaks in the host cell.[1] The cascade action of Clb proteins, which include polyketide synthases, nonribosomal peptide synthetases, hybrid polyketide synthases, and nonribosomal peptide synthetases, in addition to several accessory and maturation enzymes, are all involved in the production of colibactin. This genotoxin is primarily synthesized as a prodrug called precolibactin, which contains an *N*-myristoyl-D-asparagine side chain that is cleaved in the periplasm space to liberate the active genotoxin.[2] The pks island genes not only produce colibactin but also several secondary metabolites that have a variety of modes of action. such as analgesic lipopeptides (C12-Asn-GABA), which are capable of diffusing across the epithelial cell barrier and acting on sensory neurons to reduce visceral pain in the host.[1,3] The pks island also contributes to the production of siderophores (enterobactin, salmochelin, and yersiniabactin) via its promiscuous phosphopantetheine transferase *ClbA* and siderophore microcins via its *ClbP* peptidase.[4,5] Here, we investigated the genetic variants in the *clb B* and *clb N*

Access this article online

Quick Response Code:

Website: https://journals.lww.com/mjby

DOI: 10.4103/MJBL.MJBL_613_23

Address for correspondence: Noor Ismeal Nasser, Department of Microbiology, College of Medicine, University of Babylon, Iraq.
E-mail: noornasser1984@gmail.com

Submission: 24-May-2023 **Accepted:** 11-Aug-2023 **Published:** 23-Jan-2026

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License (CC BY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Nasser NI, Abdul-Razzag MS, AL-Hili NM. Determining the sequence of two colibactin genes, *clbN* and *clbB*, in some clinical isolates of *E. coli*. Med J Babylon 2025;22:1142-8.

genes of the colibactin island in clinical isolates of *E. coli* from women with vaginitis.

Material and Methods

Sample collection and bacterial isolates

Out of 150 high vaginal swaps, 27 isolates of *E. coli* were isolated and identified from women with vaginal infections. The diagnosis of the isolates was made by traditional methods that rely on the use of culture media (such as MacConky Agar and Eosin methylene blue). The DNA of the isolates was extracted using a commercial kit (Geneaid/UK) and according to the instructions of the producing company, and all isolates were subjected to an examination for the presence of collection genes *clb B* (579bp) and *clb N* (733bp) using the polymerase chain reaction (PCR) technique according to the study of Johnson *et al.*[6] using each of the following primers clbB forward (5-GATTTGGATACTGGCGATAACCG-3), clbB reverse (5-CCATTTCCCGTTTGAGCACAC-3), clbN forward (5-GTTTTGCTCGCCAGATAGTCATTC-3), and clbN(5-CAGTTCGGGTATGTGTGGAAG G-3).

Nucleic acids sequencing of PCR amplicons

The resolved PCR amplicons were commercially sequenced from both termini (forward and reverse directions), following the manufacturer's instructions (Macrogen Inc. Geumchen, Seoul, South Korea). Only clear chromatographs obtained from ABI (Applied Biosystems) sequence files were further analyzed, ensuring that the annotation and variations were not due to PCR or sequencing artifacts. Nucleic acid sequences of a local sample were compared with the retrieved nucleic acid sequences. The virtual positions and other details of the recovered PCR fragments were identified.

Interpretation of sequencing data

The sequencing results of the PCR products of the targeted samples were edited, aligned, and analyzed with the respective sequences in the reference database using BioEdit Sequence Alignment Editor Software Version 7.1 (DNASTAR, Madison, WI, USA). The observed variations in each sequenced sample were numbered in the PCR amplicons, as well as in their corresponding positions within the reference genome. Each detected variant within the bacterial sequences was annotated using SnapGene Viewer ver. 4.0.4 (https://www.snapgene.com/updates/snapgene-4-0-4-release-notes).

Comprehensive phylogenetic tree construction

A specific comprehensive tree was constructed in this study according to the neighbor-joining protocol described in the study of Sarhan *et al.*[7] The observed variants were compared with their neighboring homologous reference sequences using the NCBI BLASTn server.[8] A fully inclusive tree, including the observed variant, was then built using the neighbor-joining method and visualized as a circular cladogram using the iTOL suit.[9] The sequences of each incorporated species in the comprehensive tree are appropriately colored.

Ethical approval

The necessary ethical approval was obtained through verbal consent. This study was approved by the Committee of Publication Ethics at the College of Medicine, Babylon Province, Iraq, under reference No. BMS\0203\016.

Results

In this study, the results revealed that out of 150 high vaginal samples from patients suffering from vaginitis, only 27 (18%) of *E. coli* isolated have been identified, among them only seven isolates harboring pks island, two loci were selected for genetic screening, which belongs to the *clbN* and *clbB* genes, respectively. Seven samples were included for each case. These samples were screened to amplify the *clbN* and *clbB* loci sequences. Thus, the variation in the *clbN* and *clbB* loci genes can be used for *E. coli* characterization because of their ability to adapt to variable genetic diversities, as was seen in different bacterial organisms. The sequencing reaction results indicated an accurate identity after performing NCBI blasts for these PCR amplicons. Both of *clbN* and *clbB* amplicons, the NCBI BLASTn engine showed about 99% sequence similarities between the sequenced samples and the intended reference target sequences. By comparing the observed nucleic acid sequences of the investigated samples with the retrieved nucleic acid sequences of *E. coli* (GenBank accession nos. CP109929.1), accurate positions, and other details of the retrieved PCR fragments. The total length of the targeted locus was localized to the NCBI server, and the positions of the start and end of the targeted locus were confirmed within the most homologous bacterial target [Figure 1A and B].

The alignment results of the 733 bp samples revealed the presence of one nucleic acid variant represented by one nucleic acid substitution, 243C>T, compared with their reference nucleic acid sequences [Figure 2A]. The alignment results of the 579 bp samples revealed the presence of one nucleic acid variant represented by one nucleic acid substitution, namely 62C>T, compared with their reference nucleic acid sequences [Figure 2B].

All the investigated sequences were deposited in the NCBI web server, and unique accession numbers were obtained for the analyzed strains sequences concerning the *clbB* gene, The deposited sequences received the GenBank accession number OQ079983, OQ079984, OQ079985, OQ079986, OQ079987, OQ079988, and OQ079989 to represent the S1, S2, S3, S4, S5, S6, and S7 samples, respectively. With regard to the *clbN* gene, the deposited sequences received GenBank

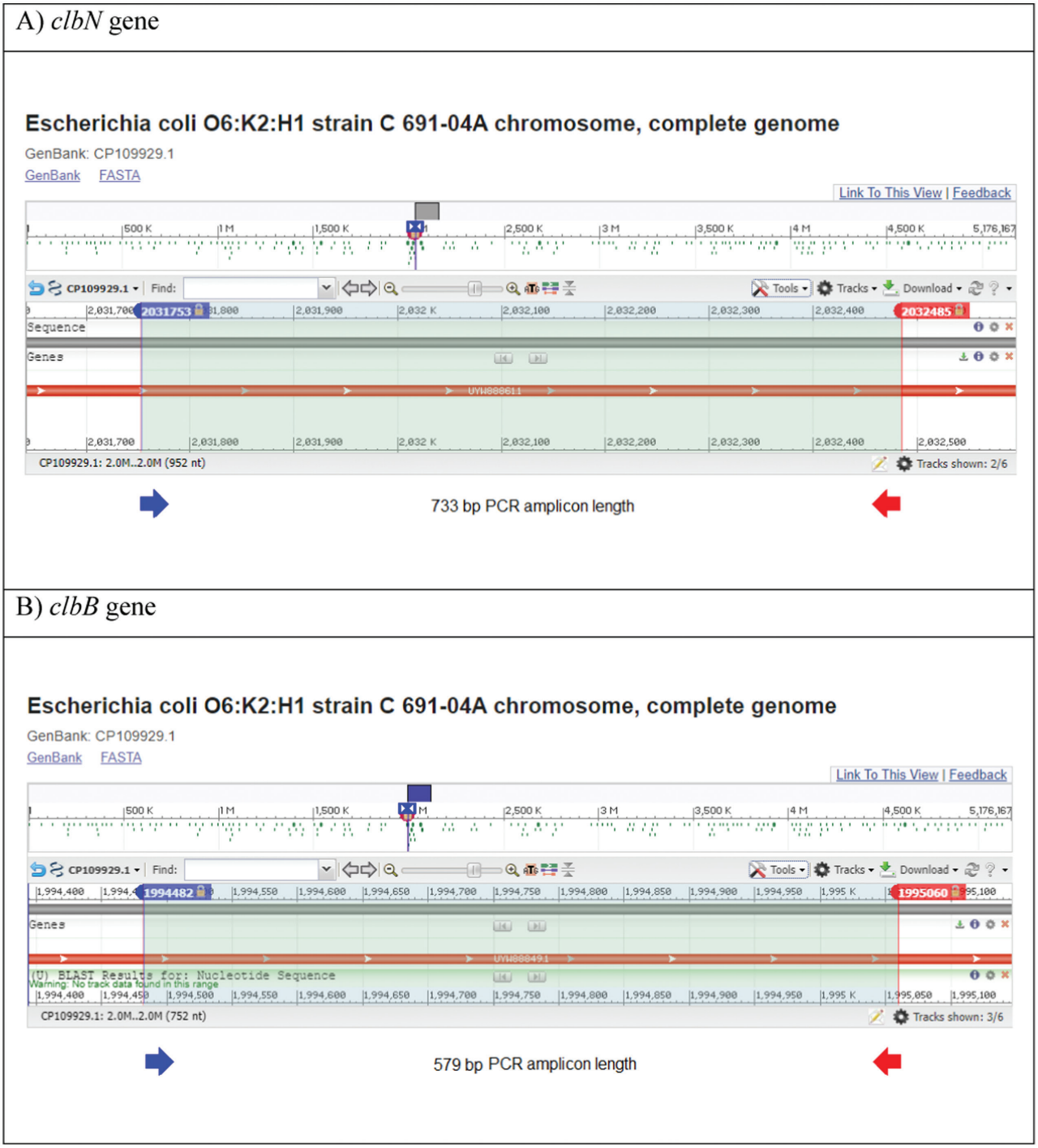

Figure 1: The exact position of the retrieved 733 bp and 579 bp amplicon partially covered a portion of the *clbN* gene (in branch A) and *clbB* gene (in branch B) within *E. coli* genomic sequences (GenBank acc. no. CP109929.1). The blue arrow refers to the starting point of this amplicon while the red arrow refers to its endpoint

accession numbers OQ079976, OQ079977, OQ079978, OQ079979, OQ079980, OQ079981, and OQ079982 to represent S1, S2, S3, S4, S5, S6, and S7, respectively. In the *clbN* amplicons, the observed nucleic acid sequences were further analyzed to identify the accurate positioning of the detected 243C>T variant in its corresponding location in both the encoded colibactin peptide synthetase *clbN* gene, which is located at position (1113) in the entire *clbN* gene, and located at position (2957062) in the whole genome of the reference strain. All nucleic acid sequences were translated to their corresponding amino acid sequences using the Expasy Translate Suite. The amino acid alignment of these amino acid sequences with their references showed the exact positions of the amplicons within the colibactin nonribosomal peptide synthetase *clbN* gene. Translation results showed that the amplified fragment occupied 243

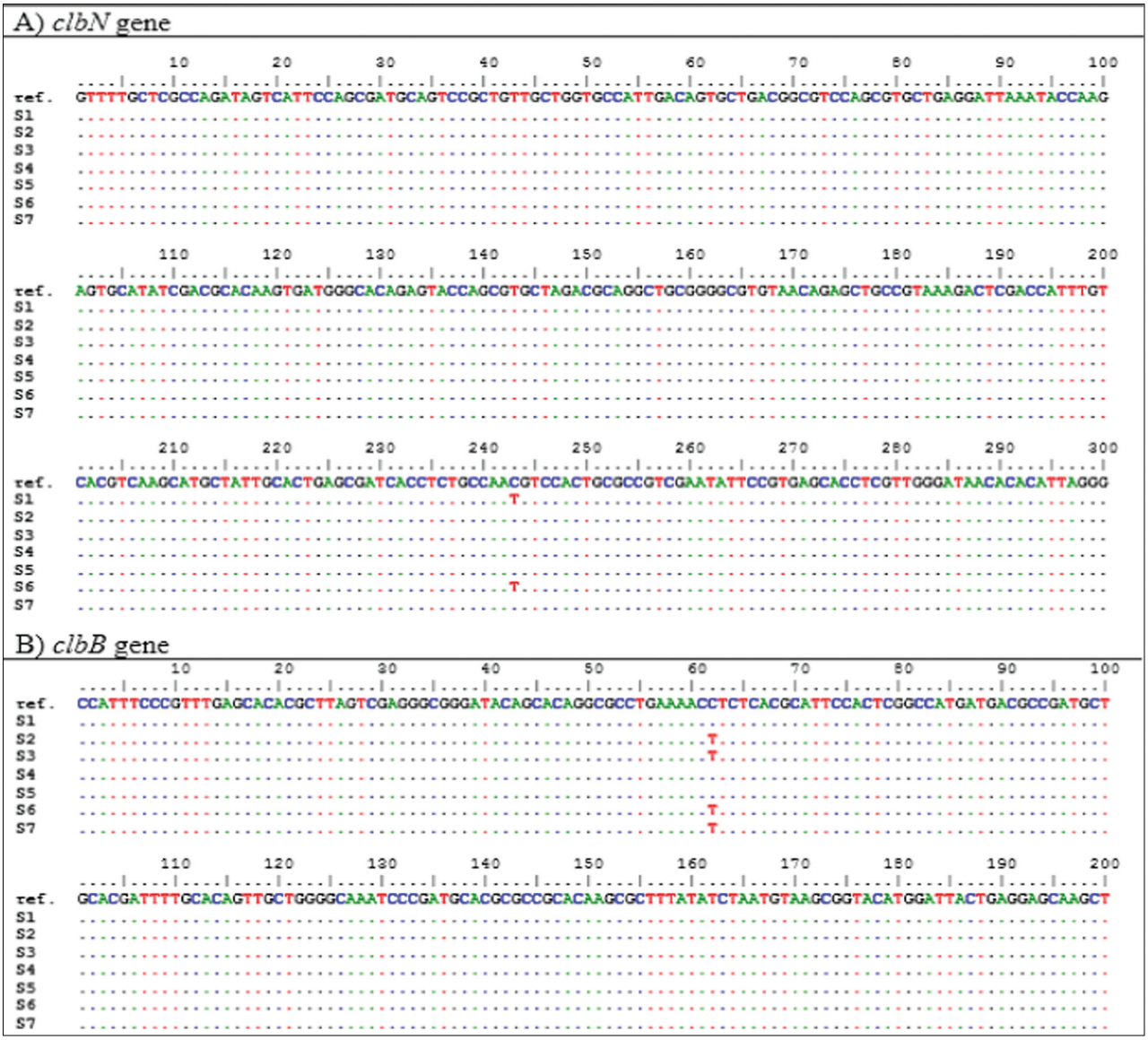

Figure 2: Nucleic acid sequences alignment of seven samples with their corresponding reference sequences of the 733 bp and 579 bp amplicons of the *clbN* gene (in branch A) and *clbB* gene (in branch B) sequences. The symbol "ref" refers to the NCBI referring sequence, letter "S," followed by a number refers to the sample number

amino acid residues of the entire protein, which was made of 1455 amino acid residues, and the translation of the *clbN* amplicons showed that the detected variant was positioned in the amino acid threonine situated at (1118) of the entire length of the protein. This variant did not cause amino acid substitutions in the encoded protein, p.thr1118=. In the *clbB* amplicons, the observed nucleic acid sequences were further analyzed to identify the accurate positioning of the detected 62C>T variant in its corresponding location in both the entire *clbB* gene-encoded colibactin hybrid nonribosomal peptide synthetase/type I polyketide synthase, which is located at (1790) position, and in the whole genome of the reference strain, which is located at (2994306) position. All the nucleic acid sequences were translated into their corresponding amino acid sequences. The amino acid alignment of these amino acid sequences with their references showed the exact positions of the amplicons within the colibactin *clbB* gene. Translation of these amplicons showed that the amplified fragment was made of 3206 amino acid residues, and translation of the *clbB* amplicons showed that the detected 62C>T variant was positioned in the amino acid phenylalanine, which is situated in 1942 of the entire length of the protein. This variant caused no amino acid substitutions in the encoded protein, namely p.F1942=. In this study, a comprehensive phylogenetic tree was generated based on the nucleic acid variations observed in the amplified 733 bp and 579 bp fragments of the *clbN* and *clbB* amplicons. This phylogenetic tree contained the investigated samples along with other relative nucleic acid sequences of *E. coli*. The investigated samples clustered into one phylogenetic clade, with close phylogenetic distances within the *E. coli* clade. Two types

of cladograms were generated to explain the different representations of the incorporated *E. coli* sequences: a rectangular cladogram [Figure 3A], and circular cladograms [Figure 3B]. In both constructed cladograms, the investigated samples were clustered into two main clades. The investigated sequences exhibited two patterns of phylogenetic distribution within the same clade. The S2, S3, S4, S5, and S7 samples were suited in one place in close vicinity to various strains of the *E. coli* sequences that have been deposited from China (GenBank acc. no. CP037449.1), USA (CP051688.1, CP051711.1, and CP051719.1). On the other side of this clade, a slight tilt was observed between samples S1 and S6 and related strains. For the *clbB* gene-based tree, the investigated sample was incorporated along

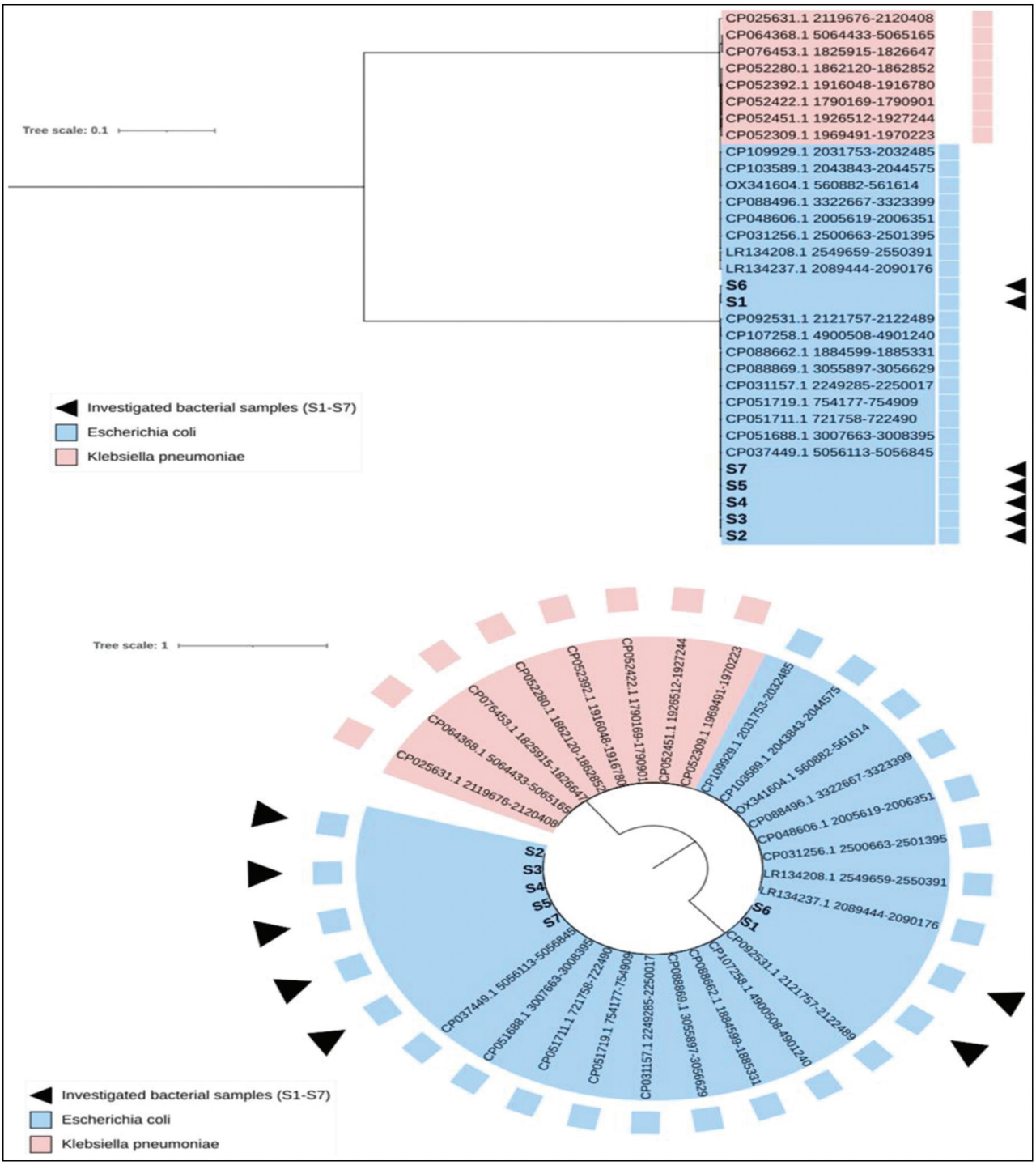

Figure 3: A comprehensive rectangular (in branch A) and circular (in branch B) cladogram phylogenetic tree of genetic variants of the *clbN* sequences of one *E. coli* sample. The black-colored triangle refers to the analyzed bacterial variants. All the mentioned numbers referred to GenBank accession number of each referring species. The number "0.1 to 1" at the left portions of the tree refers to the degree of scale range among the comprehensive tree-categorized organisms. The letter "S#" refers to the code of the investigated sample

with other relative sequences to constitute two closely related major clades of incorporated sequences within the cladogram. The investigated S1–S7 samples clustered into close phylogenetic positions with close phylogenetic distances within the *E. coli* clade. Two types of cladograms were generated to explain two different representations of the incorporated *E. coli* sequences, a rectangular cladogram [Figure 4A], and circular cladogram [Figure 4B]. In both constructed cladograms, the investigated samples clustered into one major clade within the *E. coli* sequences. The investigated sequences showed a phylogenetic distribution pattern within the same clade. However, the investigated samples were situated in the vicinity of various E. coli strains that have been deposited in the USA (GenBank acc. no.

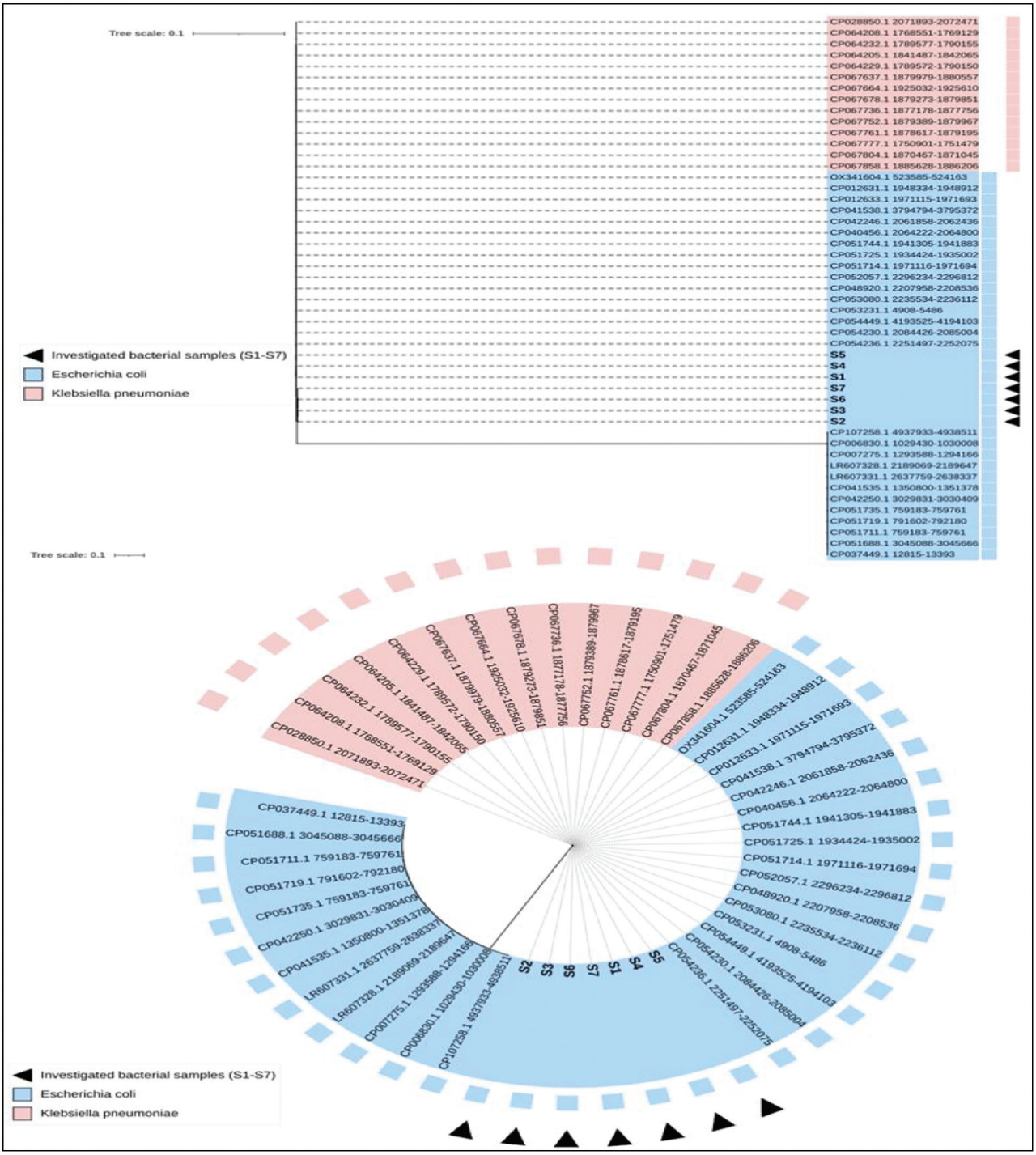

Figure 4: A comprehensive rectangular (in branch A) and circular (in branch B) cladogram phylogenetic tree of genetic variants of the *clbB* sequences of one *E. coli* sample. The black-colored triangle refers to the analyzed bacterial variants. All the mentioned numbers referred to GenBank accession number of each referring species. The number "0.1" at the left portions of the tree refers to the degree of scale range among the comprehensive tree-categorized organisms. The letter "S#" refers to the code of the investigated sample

CP054236.1, CP054230.1, CP054449.1, and CP053231.1). However, a slight tilt was observed between samples S2, S3, S6, and S7 and the related strains.

Discussion

The enzymes required for the manufacture of the genotoxin colibactin, which increases the virulence of *E. coli* strains, are encoded by the PKS island.[10,11] Of the 27 *E. coli* isolates, only nine were pk-positive, the sequencing results for seven isolates for both *clbB* and *clbN* indicated the presence of one silent mutation in *clbN* and one silent mutation in clbB. Neither mutation altered the amino acid sequence; therefore, it probably did not affect colibactin production, despite the high conservation of the pks island among *E. coli* strains. A recent study found that some *clbB* mutations are no longer genotoxic, although the majority of them still produce modest quantities of biologically active metabolites related to the PKS island,[1] and there is a high degree of conservation in the structure of the PKS island among the *E. coli* population, with more than 99% identity, suggesting that its integrity remains under strong structural and functional evolutionary constraints.[1,12] The pks island is under selective pressure to adapt *E. coli* to various ecological niches through the production of colibactin or other secondary metabolites. The *clbN* gene-based tree revealed that in comparison to the other studied wild-type sequences, the identified nucleic acid changes had a significant impact on shifting the evolutionary location of bacterial samples. The *clbB* gene-based tree revealed that, in comparison to the other studied wild-type sequences, the discovered nucleic acid change had a negligible impact on changing the evolutionary placement of bacterial samples.

These findings could be related to the type of mutation; in this case, substitutions of one nucleotide in the DNA sequence represent a type of mutation that can be directly observed in DNA sequence data.[13] Substitution mutations may reflect better or worse effects, depending on how natural selection has affected them. for example, certain substitution types (C to T transitions) appeared more frequently in the data of this study; this could indicate that such mutations occur more often or due to natural selection that prefers this mutation type once it appears. Synonymous mutations frequently evolve neutrally if there is no selection acting on an observed class of substitutions,[14] their patterns and rates are derivatives of mutation patterns and rates. However, it is uncommon to find cases in which DNA mutations are unaffected by selection. Thus, determining mutational patterns and rates is a difficult task that necessitates finding novel ways to eliminate or minimize the effects of natural selection on observed substitutions.[15-17]

Conclusion

This study found that variants in the *clbN* and *clbB* genes did not affect the production of colibactin because they are silent mutations. These amplicons can be efficiently used to detect the biological diversity of a broad spectrum of *E. coli* sequences. Nevertheless, these amplicons can be explored to discover further details of the identified bacterial sequences in various infections.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Auvray F, Perrat A, Arimizu Y, Chagneau CV, Bossuet-Greif N, Massip C, *et al*. Insights into the acquisition of the pks island and production of colibactin in the *Escherichia coli* population. Microb Genomics 2021;7:000579.
2. Tang JW, Liu X, Ye W, Li ZR, Qian PY. Biosynthesis and bioactivities of microbial genotoxin colibactins. Nat Prod Rep 2022;39:991-1014.
3. Rehm N, Wallenstein A, Keizers M, Homburg S, Magistro G, Chagneau CV, *et al*. Two polyketides intertwined in complex regulation: Posttranscriptional CsrA-mediated control of colibactin and yersiniabactin synthesis in *Escherichia coli*. Mbio 2022;13:e03814-21.
4. Chagneau CV, Massip C, Bossuet-Greif N, Fremez C, Motta JP, Shima A, *et al*. Uropathogenic *E. coli* induces DNA damage in the bladder. PLoS Pathog 2021;17:e1009310.
5. Dougherty MW, Jobin C. Shining a light on colibactin biology. Toxins 2021;13:346.
6. Johnson JR, Johnston B, Kuskowski MA, Nougayrede JP, Oswald E. Molecular epidemiology and phylogenetic distribution of the Escherichia coli pks genomic island. J Clin Microbiol 2008;46:3906-11.
7. Sarhan SR, Hashim HO, Al-Shuhaib MB. The Gly152Val mutation possibly confers resistance to beta-lactam antibiotics in ovine *Staphylococcus aureus* isolates. Open Vet J 2019;9:339-48.
8. Letunic I, Bork P. Interactive tree of life (iTOL) v4: Recent updates and new developments. Nucleic Acids Res 2019;47:W256-9.
9. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol 2000;7:203-14.
10. Jangir PK, Yang Q, Shaw LP, Caballero JD, Ogunlana L, Wheatley R, *et al*. Pre-existing chromosomal polymorphisms in pathogenic *E. coli* potentiate the evolution of resistance to a last-resort antibiotic. Elife 2022;11:e78834.
11. Guragain M, Schmidt JW, Kalchayanand N, Dickey AM, Bosilevac JM. Characterization of *Escherichia coli* harboring colibactin genes (clb) isolated from beef production and processing systems. Sci Rep 2022;12:5305.
12. Suresh A, Ranjan A, Jadhav S, Hussain A, Shaik S, Alam M, *et al*. Molecular genetic and functional analysis of pks-harboring, extra-intestinal pathogenic *Escherichia coli* from India. Front Microbiol 2018;9:2631.
13. Watford S, Warrington SJ. Bacterial DNA Mutations. [Updated 2023 Apr 10]. In: StatPearls [Internet]. Treasure Island, FL: StatPearls Publishing; 2024. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459274/.
14. Kivisaar M. Mutation and recombination rates vary across bacterial chromosome. Microorganisms 2019;8:25.
15. Chevallereau A, Meaden S, van Houte S, Westra ER, Rollie C. The effect of bacterial mutation rate on the evolution of CRISPR-Cas adaptive immunity. Philos Trans R Soc B 2019;374:20180094.
16. Kivisaar M. Mutation and recombination rates vary across bacterial chromosome. Microorganisms 2019;8:25.
17. Jebur HQ, Al-Charrakh AH. The effect of carbonyl cyanide m-chlorophenyl-hydrazine on antibiotic susceptibility in MDR Enterobacteriaceae isolates in Babylon, Iraq. Med J Babylon 2024;21:179-185. DOI: 10.4103/MJBL.MJBL_834_23.