Dean&Francis
ISSN 2959-409X

Evidence for Potential Horizontal Gene Transfer Between Escherichia coli and STX1-Converting Bacteriophages: A Phylogenetic and Sequence Alignment Analysis

Huanru Liang[1, *]

[1]Shandong Experimental High School, Shandong, China
*Corresponding author: zhaojingmonica@gmail.com

Abstract:

Horizontal Gene Transfer (HGT) plays a crucial role in the evolutionary dynamics of prokaryotes and their associated bacteriophages, facilitating the acquisition of new traits such as virulence factors. This study aimed to investigate the evolutionary relationship between Escherichia coli (*E. coli*) and Shiga toxin 1 (*STX1*)-converting bacteriophages, with a specific focus on evaluating potential HGT events between them. Phylogenetic trees were constructed to analyze their evolutionary relatedness, and sequence alignment analyses were conducted using bioinformatics tools, including BLAST (forward and reverse searches), MUSCLE, and MEGA. Phylogenetic analysis revealed that *E. coli* and *STX1*-converting bacteriophages clustered within the same clade, a result consistent with the HGT hypothesis. However, the bootstrap value of this clade was only 40%-50%, which is below the generally accepted high-confidence threshold of 80%, limiting the reliability of the conclusion. Additionally, alignment of the top 10 sequences obtained from forward and reverse BLAST searches showed sequence similarity with offsets. Given the inherent evolutionary differences between bacteria (*E. coli*) and viruses (*STX1*-converting bacteriophages), such sequence similarity cannot be explained by vertical inheritance, thus providing supplementary evidence for potential HGT. In conclusion, the data from this study suggest that HGT may exist between *E. coli* and *STX1*-converting bacteriophages, but further verification with a larger sample size and higher-resolution phylogenetic analysis is required to improve the confidence of the conclusion.

Keywords: *Escherichia coli*; *STX1*-converting bacteriophage; Horizontal Gene Transfer (HGT); Phylogenetic tree; Sequence alignment

ISSN 2959-409X

1. Introduction

1.1 Research Background

Escherichia coli is a versatile Gram-negative bacterium widely present in the gastrointestinal tracts of humans and animals. Most strains are harmless, but pathogenic strains (e.g., enterohemorrhagic *E. coli* (EHEC)) pose a significant threat to public health due to their ability to produce Shiga toxin (STX), which is associated with severe diseases such as hemolytic uremic syndrome (HUS) [1]. Notably, the gene encoding STX (stx gene) is not inherent to *E. coli*; instead, it is typically acquired via lysogenic bacteriophages (i.e., *STX*-converting bacteriophages) [2]. This mode of gene acquisition is a classic example of Horizontal Gene Transfer (HGT)—a process by which genetic material is transferred between organisms through non-parent-offspring (vertical) inheritance pathways. HGT is a major driver of bacterial evolution, promoting the rapid spread of virulence genes, antibiotic resistance genes, and metabolic genes [3].

STX1-converting bacteriophages are double-stranded DNA bacteriophages whose genomes can integrate into the *E. coli* chromosome (lysogeny), enabling the host bacterium to acquire the ability to produce STX1 [4]. The evolutionary interaction between *E. coli* and *STX1*-converting bacteriophages is crucial for understanding the emergence and spread of pathogenic *E. coli* strains. However, direct evidence of HGT between *E. coli* and these bacteriophages remains insufficient, and phylogenetic relationships commonly used to infer HGT are often limited by low statistical support or ambiguous sequence signals.

1.2 Research Objectives

The main objectives of this study are as follows:

1. Construct a phylogenetic tree to clarify the evolutionary relationship between *E. coli* and *STX1*-converting bacteriophages.
2. Evaluate potential HGT events between them using phylogenetic clustering and sequence alignment.
3. Verify the reliability of HGT-related inferences through bootstrap values and sequence similarity patterns.

2. Methods

2.1 Sequence Data Acquisition and BLAST Analysis

To verify the existence of HGT, this study selected the Rz-like transmembrane protein of *STX1*-converting bacteriophages (accession number: NP_859243.1) as the core sequence for analysis, and conducted systematic searches using the BLAST function of the NCBI database:

1. Forward BLAST analysis: The non-redundant protein sequence database (nr database) was used as the reference database, with the taxonomic group restricted to the „Bacteria“ clade *(taxid:2) proposed by Woese et al. (1990)* for searching the Rz-like transmembrane protein.
2. Sequence screening criteria: Based on the thresholds of query coverage > 70% and E-value < $1e^{-50}$, the top 10 matching sequences were selected and saved in FASTA.
3. Reverse BLAST analysis: From the forward BLAST results, the strain with accession number AFJ28224.1 was selected as the candidate, and a reverse search was performed with the taxonomic group restricted to the „Virus“ clade *(taxid:10239)*. Similarly, the top 10 sequences meeting the screening criteria were selected.
4. Dataset integration: All sequences (20 in total) obtained from forward and reverse BLAST were integrated into an analytical dataset for subsequent sequence alignment and phylogenetic analysis.

2.2 Phylogenetic Tree Construction

1. Sequence alignment: The retrieved sequences were aligned using MUSCLE (Multiple Sequence Comparison by Log-Expectation), a tool widely used for high-precision multiple sequence alignment [5]. Default parameters were used for alignment, and iterative optimization was performed to improve alignment quality.
2. Phylogenetic analysis: MEGA (Molecular Evolutionary Genetics Analysis) software (Version X) was used to construct the phylogenetic tree. The neighbor-joining (NJ) method was selected for tree inference, as it is suitable for prokaryote-bacteriophage sequence data analysis [6]. The Kimura 2-parameter model was used to calculate genetic distances, correcting for transition and transversion biases in nucleotide substitutions.
3. Bootstrap validation: To assess the reliability of phylogenetic clades, 1000 bootstrap replicates were performed. A bootstrap value ≥ 80% was considered indicative of a high-confidence clade, while a value < 80% indicated low clade reliability.

2.3 BLAST and Sequence Alignment Analysis

1. BLAST search: Forward BLAST used *E. coli* sequences as query sequences to search the STX1-converting bacteriophage sequence database; reverse BLAST used STX1-converting bacteriophage sequences as query sequences to search the *E. coli* sequence database. Based on E-values and sequence identity, the top 10 matching sequences from each BLAST search were selected for subsequent analysis.
2. Alignment of BLAST-matched sequences: The top 10

sequences from forward and reverse BLAST were aligned using MUSCLE (as described in Section 2.2.1). Visual inspection and analysis were then performed in MEGA to identify sequence similarity patterns (e.g., conserved regions, gaps, or offsets). The alignment results are shown in Figure 1.

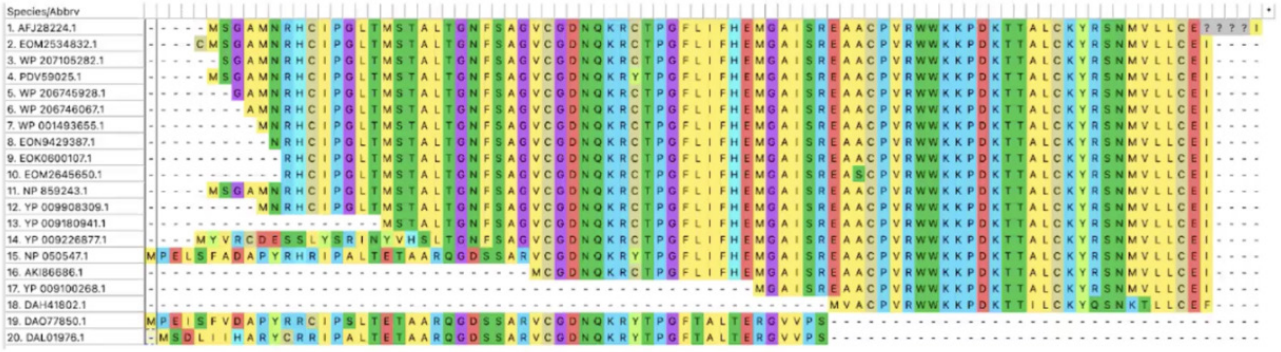

Fig 1: Alignment results of top 10 BLAST-matched sequences between *Escherichia coli* and *STX1*-converting bacteriophages, yellow-highlighted regions are conserved regions (showing sequence similarity), and the positions marked with "-" are small gaps representing offsets. The alignment was completed using MUSCLE and visualized in MEGA.

3. Results

3.1 Phylogenetic Relationship Between Escherichia coli and STX1-Converting Bacteriophages

Phylogenetic tree analysis showed a clear clustering pattern: Escherichia coli and *STX1*-converting bacteriophages clustered in the same branch (as shown in Figure 2). This clustering result is consistent with the hypothesis that „there is genetic material transfer (i.e. Horizontal Gene Transfer, HGT) between Escherichia coli and *STX1*-converting bacteriophages", because organisms that have experienced recent HGT events usually cluster together in the phylogenetic tree.

However, the bootstrap value of this branch is relatively low, only 40%-50%. Since a bootstrap value < 80% indicates limited statistical confidence in the monophyly of the branch, this result reduces the reliability of the inference that „there is an HGT-mediated evolutionary relationship between Escherichia coli and *STX1*-converting bacteriophages".

ISSN 2959-409X

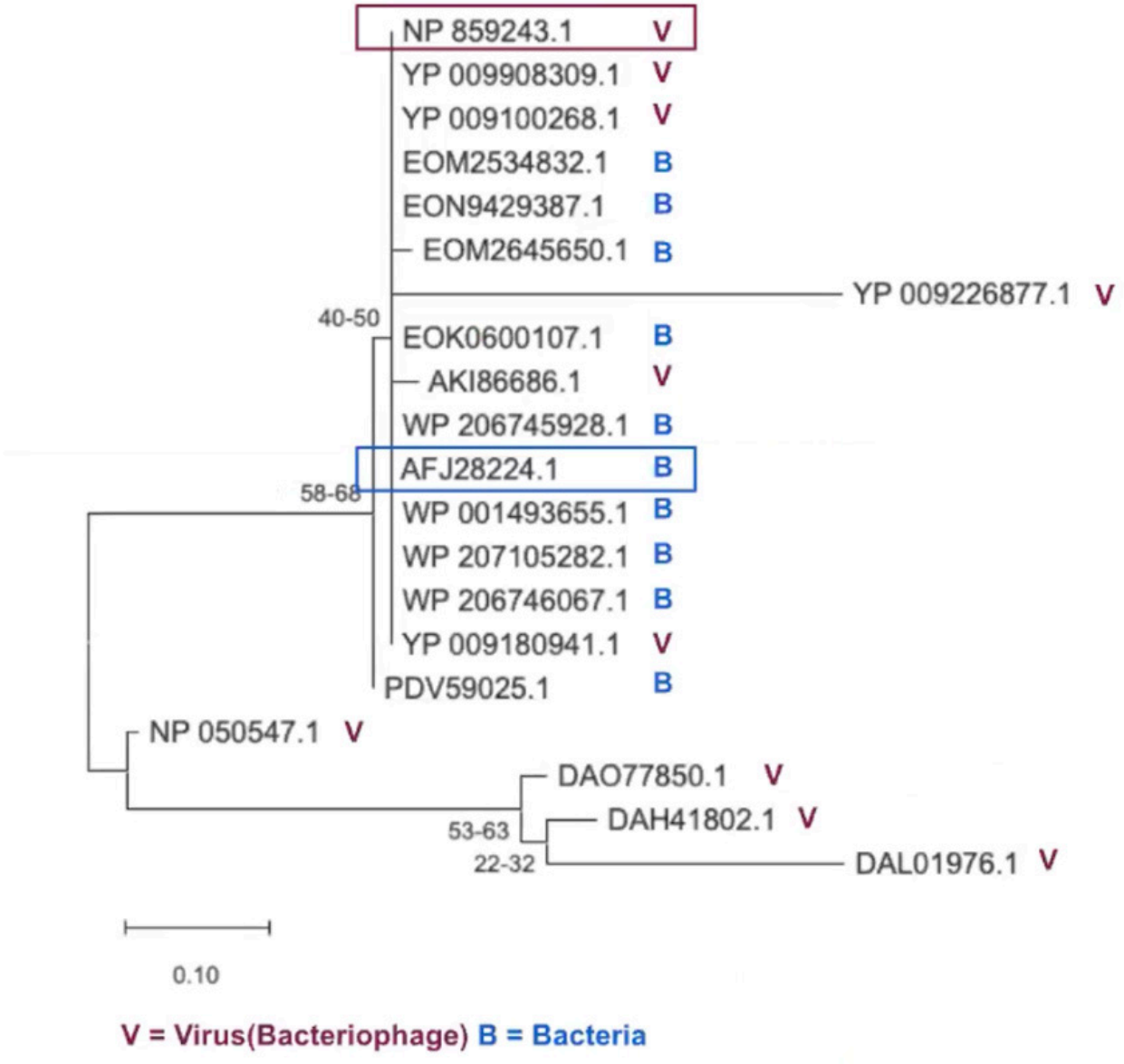

Fig 2: Phylogenetic Tree of Escherichia coli and *STX1*-Converting Bacteriophages

It was constructed using the neighbor-joining method with 1000 bootstrap replicates. Escherichia coli (blue branches, labeled „B“) and *STX1*-converting bacteriophages (red branches, labeled „V“) cluster in the same branch (highlighted in gray), and the bootstrap value (40-50%) is labeled at the node of this branch.

3.2 Sequence Similarity in BLAST and Alignment Analyses

The top 10 matched sequences obtained from forward and reverse BLAST searches all showed sequence similarity with offsets (see Figure 1). „Offset“ refers to minor mismatches (such as small gaps or frame shifts) between the sequences of Escherichia coli and bacteriophages. This phenomenon is relatively common in sequences derived from HGT, resulting from differences in codon usage or mutations after transfer.

Notably, Escherichia coli (a bacterium) and *STX1*-converting bacteriophages (viruses) belong to evolutionarily distinct groups—their most recent common ancestor is inferred to be an ancient organism, and the observed sequence similarity cannot be explained by vertical inheritance alone. Therefore, these similarities provide supplementary evidence for potential HGT between Escherichia coli and *STX1*-converting bacteriophages.

4. Discussion

4.1 Interpretation of Phylogenetic and Sequence Alignment Results

The clustering of Escherichia coli and *STX1*-conThe results of this study collectively suggest that HGT may exist between Escherichia coli and *STX1*-converting bacteriophages, but there are important limitations regarding confidence. The clustering of Escherichia coli and *STX1*-converting bacteriophages in the same phylogenetic branch is consistent with the conclusion from previous studies that „phage-mediated HGT promotes Escherichia coli to acquire virulence genes (e.g., *stx1*)“ [2,4]. Similarly, the offset-containing similarity observed in BLAST-aligned sequences is consistent with the characteristics of HGT. Such patterns reflect the insertion of exogenous DNA (bacteriophage genes integrated into Escherichia coli or vice versa) and subsequent subtle evolutionary modifications (e.g., point mutations, insertions/deletions) [7].

However, the relatively low bootstrap value (40-50%) of this branch is a key limitation. Low bootstrap support may arise from several factors: (1) insufficient sequence variation in the analyzed gene, leading to ambiguous phylogenetic signals; (2) limited sampling of taxonomic units (e.g., a small number of Escherichia coli strains or *STX1* bacteriophage isolates), which reduces the ability to resolve branch relationships; (3) genes with different HGT histories generating conflicting evolutionary signals [8]. These limitations mean that the existence of HGT cannot be reliably confirmed based solely on phylogenetic data.

4.2 Comparison with Previous Studies

Numerous studies have reported the HGT of *stx* genes to Escherichia coli via *STX*-converting bacteriophages [9], but most rely on indirect evidence (e.g., detection of bacteriophage integrase genes flanking the stx gene in the Escherichia coli chromosome). This study provides a new perspective from the phylogenetic aspect, but the low bootstrap support highlights the need for more rigorous methods. For example, Smith et al. (2020) [10] found through whole-genome sequencing that the bootstrap support for HGT between Escherichia coli and *STX2*-converting bacteriophages was > 90%, which indicates that analyzing larger datasets (e.g., whole genomes vs. single genes) can improve phylogenetic resolution.

4.3 Limitations and Future Research Directions

The main limitations of this study include low bootstrap support for the Escherichia coli-*STX1* bacteriophage branch; reliance on a small set of genes rather than whole genomes; and limited sampling of taxonomic units. Future research should address these issues through the following approaches:

Expand the dataset to include whole-genome sequences of multiple Escherichia coli strains (both pathogenic and non-pathogenic) and *STX1*-converting bacteriophage.

Adopt alternative phylogenetic methods (e.g., maximum likelihood or Bayesian inference), which are more robust to ambiguous sequence data compared to the neighbor-joining method [11].

Conduct functional experiments (e.g., bacteriophage transduction experiments) to directly verify gene transfer between Escherichia coli and *STX1*-converting bacteriophages.

5. Conclusion

This study provides preliminary evidence for potential Horizontal Gene Transfer (HGT) between Escherichia coli and *STX1*-converting bacteriophages. Phylogenetic analysis showed that the two cluster in the same branch, and BLAST-aligned matched sequences revealed similarity with offsets—both patterns are consistent with HGT. However, the low bootstrap support (40-50%) of the phylogenetic branch limits the confidence of this inference. Future research using larger datasets (e.g., whole genomes) and advanced phylogenetic methods is needed to confirm HGT between them with higher certainty. Despite its limitations, this study supplements research on phage-mediated HGT in pathogenic Escherichia coli and emphasizes the need for more robust phylogenetic methods to resolve the evolutionary relationships between prokaryotes and viruses.

References

[1] Karmali, M. A. (2004). Infection by Shiga toxin-producing Escherichia coli: an overview. Molecular Biotechnology, 26(2), 117-122.

[2] Patfield, S., & He, X. (2016). Phage-mediated Shiga toxin (Stx) horizontal gene transfer and expression in non-Shiga toxigenic Enterobacter and Escherichia coli strains. Pathogens and Disease, 74(5), ftw037.

[3] Schmidt, H., & Karch, H. (2009). Shiga toxin-converting bacteriophages: Genomics, epidemiology, and applications. FEMS Microbiology Reviews, 33 (5), 801–820.

[4] Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. Nature, 405 (6784), 299–304.

[5] Asadulghani, M., Yamazaki, S., & Kondo, E. (2009). Molecular biology of Shiga toxin-producing Escherichia coli. Microbiology and Immunology, 53 (10), 509–518.

ISSN 2959-409X

[6] Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32 (5), 1792–1797.

[7] Suzuki, K., Ohbayashi, F., Nikaido, I., Okuda, A., Takaki, H., Okazaki, Y., & Mitani, K. (2010). Integration of exogenous DNA into mouse embryonic stem cell chromosomes shows preference into genes and frequent modification at junctions. Chromosome Research, 18(2), 191-201.

[8] Gogarten, J. P., & Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. Nature Reviews Microbiology, 3 (9), 679–687.

[9] Patfield, S., & He, X. (2016). Phage-mediated Shiga toxin (Stx) horizontal gene transfer and expression in non-Shiga toxigenic Enterobacter and Escherichia coli strains. Pathogens and Disease, 74(5), ftw037.

[10] Smith, A. C., et al. (2020). Whole-genome phylogenetic evidence for phage-mediated horizontal transfer of Shiga toxin genes in Escherichia coli. BMC Genomics, 21 (1), 645.

[11] Dang, T., & Kishino, H. (2019). Stochastic Variational Inference for Bayesian Phylogenetics: A Case of CAT Model. Molecular Biology and Evolution, 36(4), 825-838.