

RESULTS

1. Detection of *espK*, *espV*, and *espN* Among EHEC and EPEC Isolates

To characterize the distribution of *esp* family virulence genes across pathogenic *Escherichia coli*, we screened all genomes for the presence of ***espK***, ***espV***, and ***espN*** using a BLAST-based marker gene panel.

A presence/absence matrix was generated for all isolates and used to define two major groups:

- ***esp*-positive (*espPOS*):** isolates carrying at least one of *espK*, *espV*, or *espN*
- ***esp*-negative (*espNEG*):** isolates lacking all three genes

This classification allowed us to compare serotype enrichment between isolates harboring *esp* virulence determinants versus those lacking them.

Across the dataset, *esp* gene carriage showed strong lineage structure, with several EHEC-associated serotypes exhibiting consistently high frequencies of *esp*-positive genotypes, whereas many classical and atypical EPEC serotypes were dominated by *esp*-negative isolates.

2. Serotype Assignment of *espPOS* and *espNEG* Isolates

All *esp*-positive EHEC isolates and all *esp*-negative EPEC isolates were serotyped using ECTyper. Output summaries were merged with marker data to produce:

- **EHEC_espPOS_with_serotype.tsv**
- **EPEC_espPOS_with_serotype.tsv**
- **EPEC_espNEG_with_serotype.tsv**

For each serotype, counts and percentages of *espPOS* and *espNEG* isolates were calculated. These datasets enabled systematic comparison of serotype–*esp* linkage patterns.

3. Serotype Distribution Among *espPOS* EPEC Isolates

Horizontal stacked barplots revealed distinct serotype-associated patterns of *esp* gene carriage among EPEC isolates.

Several serotypes showed **strong enrichment for *espPOS* isolates**, including:

- **O157:H7**
- **O26:H11**

- **O108:H9**
- **O145:H34**

In these lineages, >60–100% of isolates carried at least one of espK, espV, or espN.

Notably, O157:H7 and O145:H34 isolates were uniformly esp-positive in this dataset, consistent with the presence of *esp*-associated pathogenicity island genes in these lineages.

In contrast, certain serotypes such as **O55:H7** exhibited mixed profiles, with 10–15% espPOS isolates and the majority being espNEG.

This variation underscores lineage-specific mosaicism in effector repertoires among EPEC populations.

4. Serotype Distribution Among espNEG EPEC Isolates

To investigate serotypes lacking *esp* genes entirely, separate analysis of espNEG isolates was performed.

The espNEG-only distribution showed that many “classic” or atypical EPEC serotypes are dominated by isolates without *esp* gene carriage, including:

- **O55:H7**
- **O071:H27**
- **O15:H2**
- **O109:H21**

These serotypes often exhibited 80–100% espNEG prevalence, indicating that absence of espK/V/N is characteristic for these lineages.

Plots revealed that some serotypes present in both espPOS and espNEG groups had markedly different proportions, suggesting ongoing acquisition or loss of *esp* gene clusters within specific clades.

5. Integrated Comparison of espPOS and espNEG Across EPEC Serotypes

A combined analysis of EPEC serotypes (Top 20 by total isolates) demonstrated clear shifts in *esp* gene carriage.

Key findings include:

- Serotypes **O157:H7**, **O108:H9**, and **O145:H34** were overwhelmingly espPOS ($\geq 80\text{--}100\%$).

- Serotypes **O55:H7**, **O71:H27**, and **O109:H21** were predominantly espNEG ($\geq 85\text{--}100\%$).
- A small subset of serotypes, such as **O26:H11** and **O88:H25**, exhibited intermediate patterns, suggesting more dynamic esp gene flow.

These results indicate that *espK/espV/espN* carriage is not random but instead strongly associated with specific serotypes and lineages within EPEC.

6. Serotype Distribution Among espPOS EHEC Isolates

Analysis of espPOS EHEC isolates showed a strikingly consistent pattern:

- Canonical EHEC serotypes—including **O157:H7**, **O26:H11**, **O111:H8**, and **O103:H2**—were nearly uniformly esp-positive.
- Most isolates in these serotypes carried multiple esp genes (espK+, espV+, espN+), reflecting their presence in EHEC-associated pathogenicity islands.

These findings reinforce the association of esp effector genes with the hallmark virulence architecture of shiga-toxigenic EHEC lineages.

7. Overall Patterns of esp Gene–Serotype Linkage

Across both EHEC and EPEC datasets, three major trends were observed:

1. Strong lineage specificity

esp-positive genotypes clustered tightly with characteristic EHEC and EPEC serotypes, with near-complete carriage in O157:H7 and O145:H34.

2. High espNEG prevalence in specific EPEC serotypes

Lineages such as O55:H7 and O109:H21 showed almost universal absence of espK/V/N.

3. Evidence of recombination or plasmid-mediated mobility

Intermediate categories (e.g., O26:H11, O88:H25) suggest ongoing acquisition or loss of esp effector genes within evolving lineages.