

# Microarray-based studies on gene content of *Escherichia coli* O157

Outi Leppäranta<sup>1</sup>, Panu Somervuo<sup>2</sup>, Suvi Nykäsenoja<sup>1</sup>, Leila Rantala<sup>1</sup>, Petri Auvinen<sup>2</sup>, and Vesa Myllys<sup>1</sup>

<sup>1</sup> Finnish Food Safety Authority Evira, Mustialankatu 3, FI-00790 Helsinki, Finland • <sup>2</sup> Institute of Biotechnology, Viikinkaari 4, FI-00014 University of Helsinki, Finland

## Introduction

Enterohaemorrhagic *Escherichia coli* (EHEC) is a food and waterborne pathogen with a very low infective dose. An infection poses a risk of serious complications. Most EHEC outbreaks have been caused by *E. coli* O157 strains. Production of Shiga toxins (*stx*) and adherence-mediating factor intimin (*eaeA*) are counted as important factors in the pathogenesis. However, the virulence of *E. coli* O157 is still not fully understood.

## Aims of the study

- to compare the gene content of pathogenic and non-pathogenic *E. coli* strains using DNA microarrays
- to use statistical methods to find differences between the studied strains

## Materials and methods

- 21 *E. coli* strains, 19 of them *E. coli* O157
  - clinical isolates along with isolates from various other sources, e.g. bovine faecal samples, environmental samples from barn surfaces, and bovine meat
- Prior to microarray experiments, the strains were divided into groups
  - I six strains associated with human disease
  - II six strains with *eae* and *stx* genes
  - III five strains with *eae* but no *stx* genes
  - IV four non-EHEC strains
- Genomic DNA was fluorescently labelled with Cy3 / Cy5 and hybridized with the probes of OciChip *E. coli* O157 Arrays (Ocimum Biosolutions)
- Data was normalized within each array using iterative linear regression method
- Microarray profile-based similarity between the strains was compared with the classification of strains described above

## Results

- Analysis of all ca. 6000 probes on the array shows that the gene contents of strains in group IV differ from strains in groups I-III.
  - The difference is explained by absence of ca. 100 genes from strains in group IV (Fig. 1)

- Statistical analyses - Wilcoxon rank sum test - revealed a set of genes best explaining the differences between pathogenic (group I) and possibly pathogenic (group III) strains (Fig.2)
  - 692 genes,  $P < 0.05$
- Based on the 692 genes mentioned above, the six strains of group II were computed against the strains of groups I and III, in order to categorize strains of group II as pathogenic or possibly pathogenic
  - the gene content of four strains resembles that of the strains of group I
  - the gene content of one strain resembles that of the strains of group III
  - Figure 3 shows a dendrogram based on the selected 692 genes

## Conclusions

- A set of genes best explaining the differences between the studied pathogenic and possibly pathogenic *E. coli* strains was revealed by statistical analyses
- Preliminary results indicate that microarray profiles and statistical analyses might be used for classification of *E. coli* strains as pathogenic or possibly pathogenic. However, further studies are still required.

## Acknowledgements

This study has been a part of a research consortium that develops microarray-based diagnostics to bacterial pathogens in plants, food or animals in Viikki Campus, Helsinki, Finland.

The study was supported by Finnish Funding Agency for Technology and Innovation (TEKES) and Ministry of Agriculture and Forestry, Finland.

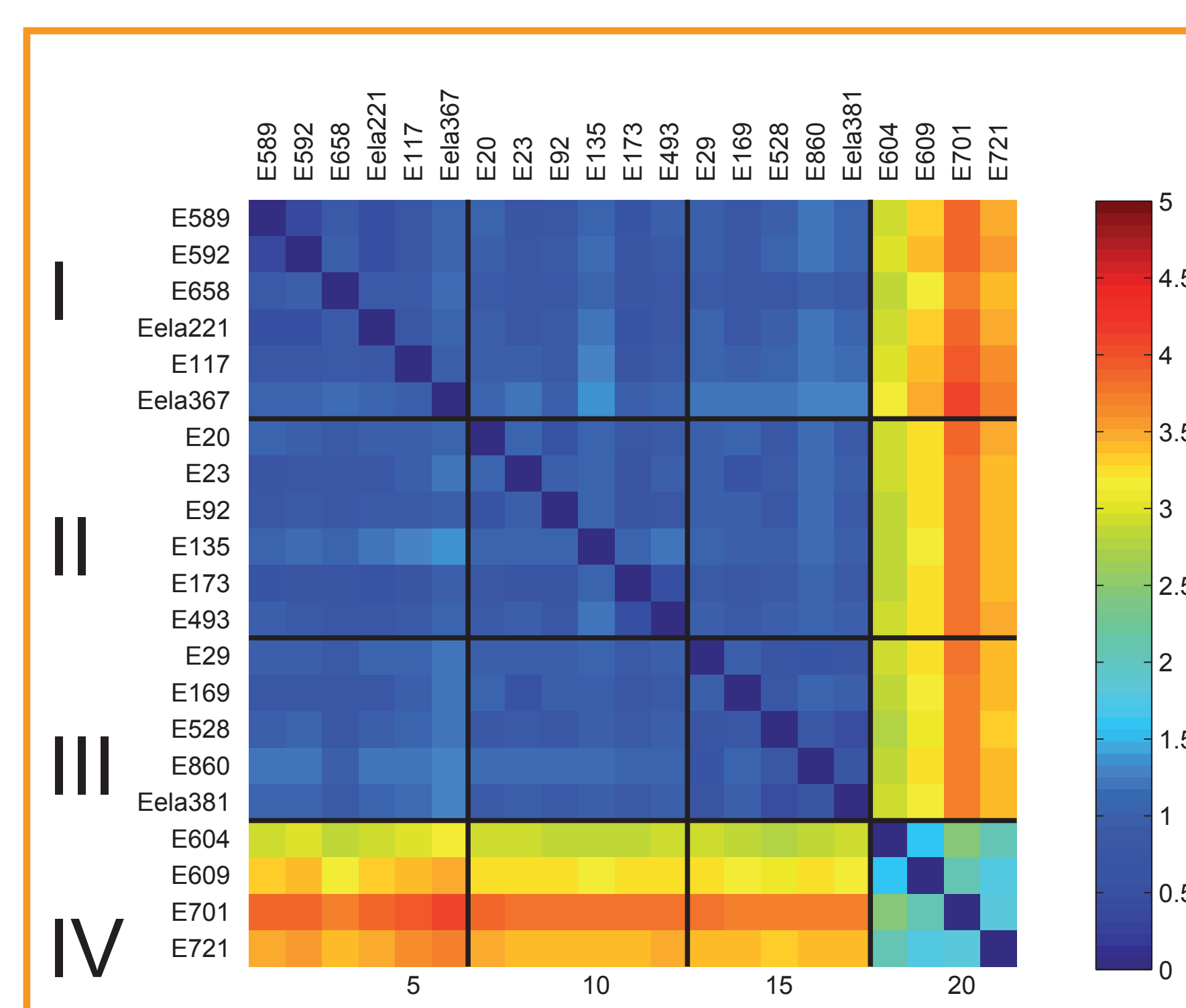


Figure 1. Similarity of strains (groups I-IV) when all probes on the array are included in the analysis. The gene contents of strains in group IV differ from strains in groups I-III.

The strains are arranged in the same order on both axes. The microarray profile of each strain was compared to profiles of all other strains. The colour of each intersection describes similarity between the compared strains, with dark blue (0) being the most similar. Thus, a comparison of each strain with itself creates the dark blue diagonal line.

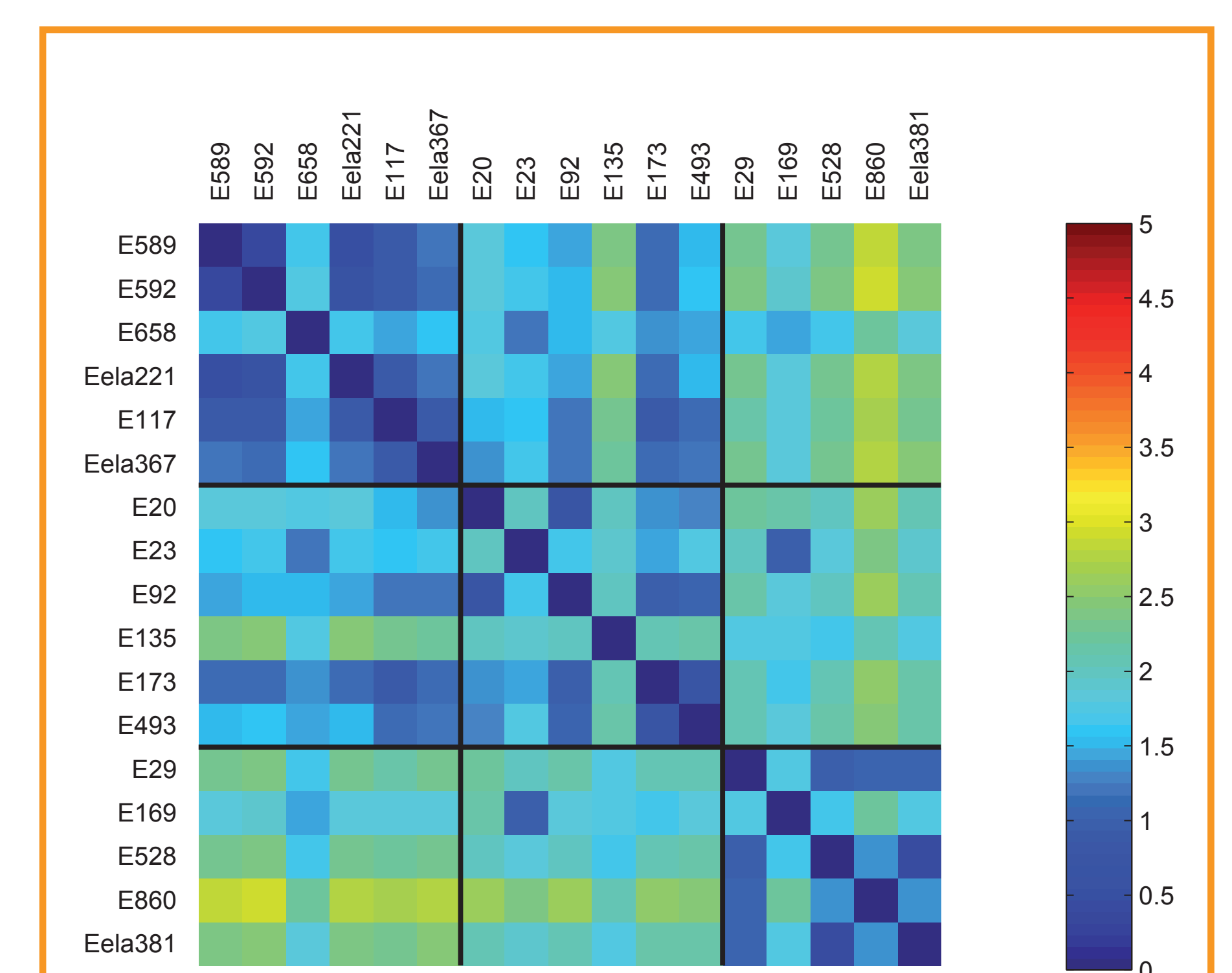


Figure 2. Similarity of strains (groups I-III) when only a selected set of genes (692 genes,  $P < 0.05$ ) are included in the analysis. The strains are arranged in the same order on both axes. The microarray profile of each strain was compared to profiles of all other strains. The colour of each intersection describes similarity between the compared strains, with dark blue (0) being the most similar. Thus, a comparison of each strain with itself creates the dark blue diagonal line.

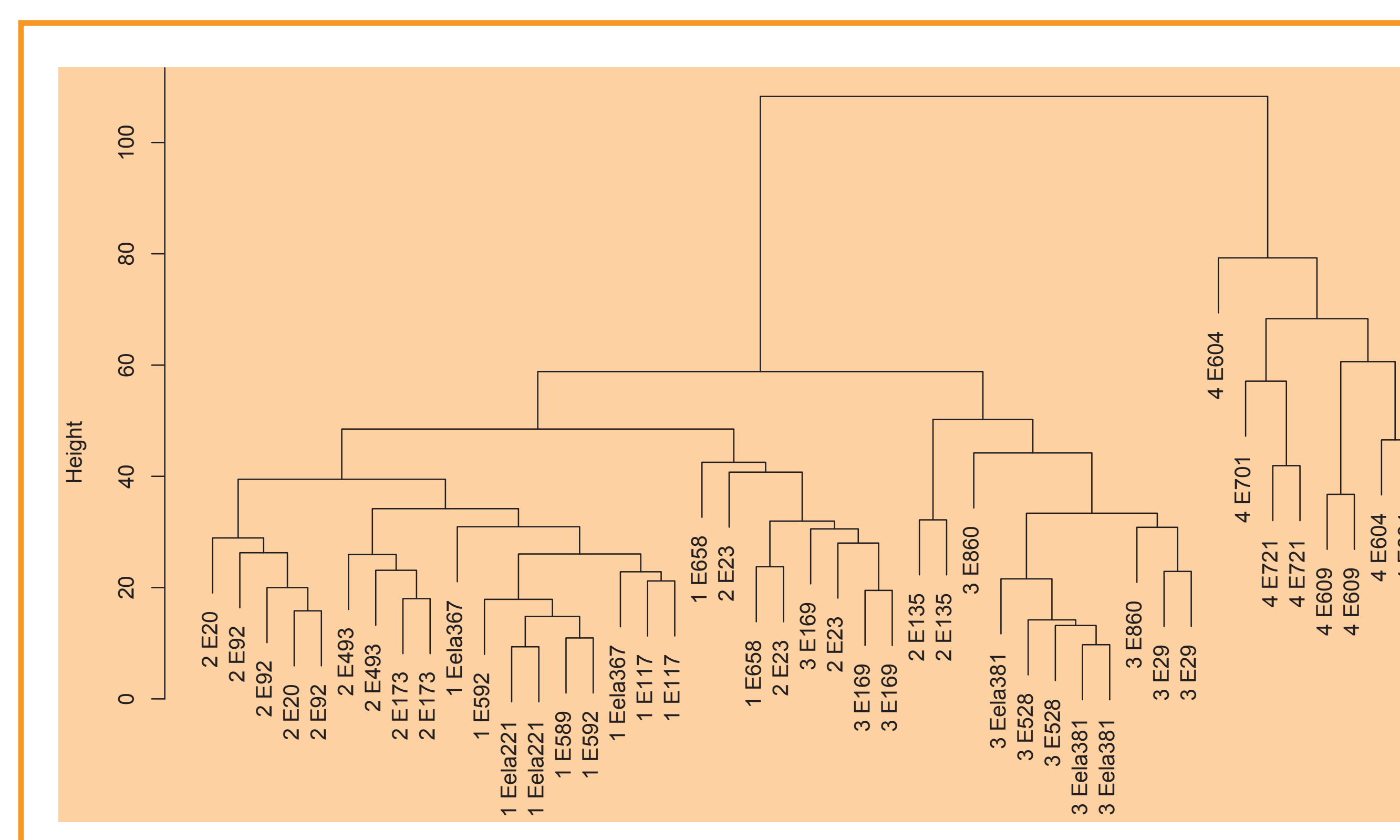


Figure 3. A dendrogram showing relations between the studied *E. coli* strains is based on the selected set of genes (692 genes,  $P < 0.05$ ). The first number of each sample refers to the original classification of strains in four groups (I-IV).

On the right side, there is a separate branch formed by the strains of group IV. The strains of groups I and III are mainly in their own branches, and the strains of group II are placed in these branches. In the middle there is a branch consisting of three strains from three groups (cf. Figure 2; strains E658, E23, and E169).