

Phylogenetically stratified GWAS reveals insights into *Escherichia coli* specific virulence factors in adults and neonates

Riccardo Nodari^{1,2}, Lodovico Sterzi³, Diego Marco Minore³, Clara Bonaiti³, Laura Folgori⁴, Lucia Barcellini⁴, Simona Panelli³, Raffaele De Francesco^{1,2}, Francesco Comandatore³

¹ *INGM - Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, Italy*

² *Department of Pharmacological and Biomolecular Sciences (DiSFeB), University of Milan, Milan, Italy*

³ *Department of Biomedical and Clinical Sciences, Pediatric Clinical Research Center "Romeo and Enrica Invernizzi", University of Milan, Milan, Italy.*

⁴ *Department of Paediatrics, Vittore Buzzi Children's Hospital, Milan, Italy*

Escherichia coli is a highly diverse bacterial species capable of occupying a wide range of ecological niches, from harmless gut commensalism to severe extraintestinal infections. Identifying genetic determinants that underlie adaptation to specific host and environmental contexts remains challenging due to extensive horizontal gene transfer and strong phylogenetic structure. In this study, we performed a genome-wide association study (GWAS) on gene presence–absence across a large collection of *E. coli* genomes isolated from adult and neonatal bloodstream infections, urinary tract infections, healthy gut, hospital settings, and non-clinical environmental sources. By explicitly accounting for population structure, we identified lineage-independent, source-associated genetic signatures while minimizing confounding by vertically inherited genes.

Bloodstream and urinary isolates were significantly enriched in genes involved in metal acquisition (iron and manganese), adhesion, biofilm formation, and stress responses, highlighting shared adaptations for survival in hostile host environments. Neonatal bloodstream isolates showed a strong association with the yersiniabactin iron-scavenging system, whereas adult bloodstream isolates were enriched in acid resistance genes suggesting age-specific infection routes and virulence strategies. Urinary isolates exhibited marked metabolic flexibility, with enrichment of genes related to metal acquisition, sugar uptake, and secretion systems, consistent with adaptation to the nutrient-limited urinary tract.

Together, these findings demonstrate that *E. coli* adapts to distinct ecological and clinical niches through the acquisition and maintenance of functionally relevant genes independent of phylogenetic background. Source-specific genetic traits identified here provide valuable insights into *E. coli* pathogenesis, highlight age-dependent virulence mechanisms, and offer potential biomarkers for diagnostics, surveillance, and targeted intervention strategies, particularly in vulnerable populations such as neonates and hospitalized patients.