Data Article

Data highlighting phenotypic diversity of urine-associated *Escherichia coli* isolates

Allison R. Eberly¹,a, Connor J. Bebou², Ching Man Carmen Tong³, Gerald T. Van Horn¹, Hamilton D. Green¹, Madison J. Fitzgerald¹,b, Shuvro De³, Emily K. Apple¹,c, Alexandra C. Schrimpe-Rutledge⁴, Simona G. Codreanu⁴, Stacy D. Sherrod⁴, John A. McLean⁴, Douglass B. Clayton³, Charles W. Stratton¹,²,⁵, Jonathan E. Schmitz¹,²,⁵,**, Maria Hadjifrangiskou¹,³,⁵,*

¹ Division of Molecular Pathogenesis and
² Division of Laboratory Medicine, Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232
³ Division of Pediatric Urology, Department of Urology, Vanderbilt University Medical Center, Nashville, TN 37232
⁴ Center for Innovative Technologies, Department of Chemistry, Vanderbilt University, Nashville, TN 37232
⁵ Vanderbilt Institute for Infection, Immunology & Inflammation (V4), Vanderbilt University Medical Center, Nashville, TN 37232

** Corresponding Author: Maria Hadjifrangiskou, MCN AA-4210B, 1161 21st Ave South, Nashville TN 37212, 615-322-4851
* Co-corresponding Author: Jonathan E. Schmitz, 4519 The Vanderbilt Clinic, 1211 Medical Center Drive, Nashville TN 37212, 615-343-1289
E-mail addresses: jonathan.e.schmitz@vumc.org (J.E. Schmitz), maria.hadjifrangiskou@vumc.org (M. Hadjifrangiskou).

A R T I C L E   I N F O

Article history:
Received 1 May 2020
Revised 26 May 2020
Accepted 27 May 2020
Available online 3 June 2020

A B S T R A C T

This article provides a reusable dataset describing detailed phenotypic and associated clinical parameters in n=303 clinical isolates of urinary *Escherichia coli* collected at Vanderbilt University Medical Center. De-identified clinical data collected with each isolate are detailed here and correlated to biofilm abundance and metabolomics data. Biofilm-abundance data were collected for each isolate under different in vitro conditions along with datasets quantifying
Keywords: Asymptomatic bacteriuria Biofilm Cystitis Metabolomics Urinary tract infection Uropathogenic Escherichia coli

biofilm abundance of each isolate under different conditions. Metabolomics data were collected from a subset of bacterial strains isolated from uncomplicated cases of cystitis or cases with no apparent symptoms accompanying colonization. For more insight, please see “Defining a Molecular Signature for Uropathogenic versus Urocolonizing Escherichia coli: The Status of the Field and New Clinical Opportunities” [1]. © 2020 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications Table

| Subject | Microbiology |
|---------|--------------|
| Specific subject area | Clinical microbiology, Diagnostics, urocolonizer vs uropathogen, Escherichia coli |
| Type of data | Table Figure |
| How data were acquired | Clinical isolates were collected via semi-quantitative aerobic culture by the Vanderbilt University Medical Center Clinical Microbiology Laboratory. Corresponding patient data from the electronic health record were reviewed by the Clinical Laboratory Medical Director. Isolates and patient data were collected in accordance with IRB #151465. Spectrophotometric data of crystal violet-stained biofilms were collected at A570nm optical density using a SpectraMax i3 plate reader (Molecular Devices). Colony biofilms were imaged after 11 days of incubation for phenotypic categorization based on colony morphology. Mass spectrometry analyses were performed on a Q-Exactive HF hybrid mass spectrometer (Thermo Fisher Scientific) equipped with a Vanquish UHPLC binast system and autosampler (Thermo Fisher Scientific). Metabolite data processing and analysis were performed using Progenesis QI v.2.1 (Non-linear Dynamics) and Metaboanalyst 4.0. |
| Data format | Raw Analyzed Filtered |
| Parameters for data collection | No urine specimens were collected for the sole purpose of this project; rather isolates were isolated from urine specimens as part of routine patient care. Strains were designated for inclusion if only E. coli (no additional species isolated) was identified at a burden greater than 100,000 colony forming units per milliliter. Targeted data were collected from the electronic health record and reviewed by the Clinical Laboratory Medical Director. |
| Description of data collection | Three hundred and three urinary E. coli isolates and corresponding patient data were collected and reviewed by the Clinical Laboratory Medical Director (targeted data recorded in Fig. 1). De-identified patient information was recorded into Data Sheet 1 and researchers performing biofilm analyses remained blinded to the patient data until all data were collected. Surface-associated and colony biofilms were assessed for each isolate in Data Sheets 2 and 3. For metabolomics analyses, the asymptomatic bacteriuria or cystitis isolates were pooled from patients without any documented complicating predispositions or risk factors (Data Sheets 4 and 5) and randomly re-pooled for validation (Data 6 and 7). |
| Data source location | Vanderbilt University Medical Center, Nashville, TN, USA |
| Data accessibility | All data used and generated is included in this article or within the related research article. |
| Related research article | Eberly AR, Beebout CJ, Carmen Tong CM, Van Horn GT, Green HD, Fitzgerald MJ, et al. Defining a Molecular Signature for Uropathogenic versus Urocolonizing Escherichia coli: The Status of the Field and New Clinical Opportunities. J Mol Biol. 2020;432:786-804. |

Value of the Data

• There are no empiric tools to date to differentiate asymptomatic bacteriuria (ASB) causing from cystitis causing E. coli isolates.
• Current clinical diagnostics rely on patient symptomatology when urinary bacterial burden is above 100,000 colony forming units per milliliter.
• Clinicians and basic scientists can benefit from these data that draw attention to a long-standing issue that phenotypic assays cannot be used to predict the disease state of E. coli urinary tract isolates.
• These data can be: used as a tool to distil a metabolic signature to differentiate ASB- and cystitis-associated E. coli strains; apply additional algorithms to test for phenotype-clinical outcome associations that may not have been incorporated here.

1. Data

There are currently no clinical tests that differentiate asymptomatic bacteriuria from symptom-associated urinary E. coli isolates. To begin to address this gap in the field, a retrospective analysis was conducted. Three hundred and three urinary Escherichia coli strains were isolated from patient urine (derivatives of clinical care) at Vanderbilt University Medical Center (VUMC) Clinical Microbiology Laboratory. Targeted clinical data were extrapolated from source patient charts by the Clinical Laboratory Medical Director and de-identified prior to downstream analyses. Fig. 1 shows a mock-up of the worksheet used to record the de-identified patient information from each isolate. The information collected from the electronic health record (EHR) included the sex, age range, collection setting, associated infections, recent urinary catheter, pregnancy, diabetes, structural/functional urinary tract abnormalities, asymptomatic bacteriuria, immune status, and associated infection. Data Sheet 1 includes the de-identified clinical data for each of the isolates.

The formation of biofilms is one strategy that bacteria utilize to persist in the environment. We first assessed liquid-growth biofilm abundance under 21% (atmospheric level) and 4% (represents hypoxic bladder environment) oxygen, utilizing the crystal violet method of O’Toole [2]. The isolates are listed from highest to lowest crystal violet absorbance in Data Sheet 2. Colony morphology on nutrient agar was also assessed for each of the isolates in which overnight culture is spotted onto Yeast Extract/Casamino acids agar supplemented with Congo red dye (YESCA CR) and incubated for 11 days at room temperature (Data Sheet 3). Congo red uptake (indicative of the production of cellulose or curli fibers) and colony rugosity were evaluated on day 11, as described previously [3, 4].

For metabolomics analyses, the isolates from patients with no other co-morbidities were selected (Data Sheet 4). Strains were grown individually and pooled in groups of 8 for metabolomic analyses performed in collaboration with the Center for Innovative Technology (CIT) at Vanderbilt University (Data Sheet 4). The supernatant fractions – representing the extracellular milieu of the pooled isolates – were subjected to a global metabolomics workflow that utilized reverse phase liquid chromatography tandem mass spectrometry (LC-MS/MS). A total of 6,711 compounds were detected, of which 582 showed significant differences in abundance after filtering by a p-value <0.05 and a q-value of <0.1. Using ProGenesis 2.3 software, the detected compounds were searched against multiple metabolite libraries and assigned candidate annotations with confidence levels (detailed in Experimental Design, Materials, and Methods and [5]). A total of 429 significant compounds had a database hit, as detailed in Data Sheet 5. A second metabolomics analysis was performed to validate the initial findings. In the second analysis, isolate groups were scrambled for statistical robustness. The groups are detailed in Data Sheet 6, while Data Sheet 7 details the statistically significant compounds in the re-pooled samples.

2. Experimental Design, Materials, and Methods

2.1. Urine-associated E. coli collection

This project was completed in accordance with Vanderbilt IRB #151465. Urine-associated E. coli isolates were initially isolated by the VUMC Clinical Microbiology Laboratory (a CLIA/CAP
Fig. 1. Retrospective analysis: de-identified patient information. Form used by clinician to obtain de-identified clinical information from source patient charts corresponding to each Vanderbilt Urinary Tract Isolate (VUTI). Panel represents patient parameters collected that pertain to urinary tract infections. These include: sex, age, collection setting, presence of a urinary catheter, pregnancy, diabetes, structural or functional abnormality, if the patient was immunocompromised, and if the patient had an associated UTI infection.

certified facility) via guideline-recommended semi-quantitative aerobic culture [6]. They represent the routine derivatives of patient care. No urine specimens were collected for the express purpose of this project. Over the course of a month, clinical urine cultures were designated for inclusion in this study cohort if *E. coli* was observed in purity (no co-isolated species) and at a high burden (>100,000 CFU/mL). For each isolate, prior to permanent de-identification, the EHR of the source-patient (in Epic Hyperspace) was reviewed by the Clinical Laboratory Medical Director to record the targeted clinical information. These data were compiled in a database that was linked to the corresponding bacterial isolate (by the anonymized VUTI numbers), but also then permanently de-identified from the source. To avoid bias, downstream experimental analyses were conducted by individuals who were blinded to associated clinical parameters of the strains with which they were working. Overall, 303 isolates were collected in this study. Eighteen isolates were excluded from downstream analyses, as 4 were collected in error and 14 had unclear etiologies, leading to 285 isolates broken down into two groups for downstream analyses: UTI-associated or ASB-associated.
2.2. Biofilm Assays

For liquid-growth biofilms at the surface-interface, the crystal violet method of O’Toole was performed [2]. This assay quantifies total abundance of biofilm, which comprises the bacteria themselves and associated extracellular matrix. Overnight cultures were grown in Lysogeny Broth (LB) (Fisher) pH 7.4 at 37°C with shaking for all strains. These overnight cultures were diluted OD$_{600}=0.05$. A total volume of 100 μL was added to a 96 well PVC plate and each plate was incubated statically at room temperature for 48 hours. The 96 well PVC plates were rinsed in sterile water, stained with 0.5% crystal violet, and disaggregated with 35% acetic acid. Using a SpectraMax i3 plate reader (Molecular Devices), absorbance of the disaggregated biofilm was measured at 570 nm. At least 3 biological replicates with at least 8 technical replicates were recorded for each isolate tested. For colony morphology evaluation, 10 μL of overnight culture from each tested strain was spotted onto 1.2x YECA CR agar and allowed to grow at room temperature and ambient atmosphere for a period of 11 days. Colony biofilms were imaged after 11 days of incubation. Images represent at least two biological replicates. Description of CR uptake and rugosity was qualitatively recorded as described previously [3,4].

2.3. Combined human urine

In accordance with approved protocols, urine was collected from healthy human volunteers and filtered through a 0.22 μm filter prior to use. In order to create a standard ex vivo growth media for all of the metabolomics analyses, equal volumes of male and female urine were combined from multiple volunteers. Healthy volunteers were classified as individuals who are urologically asymptomatic and not menstruating, and who have not taken antibiotics in the last 90 days.

2.4. Growth conditions for metabolomics analysis

For each isolate, a single colony from an agar dish was inoculated in 5 mL LB and shaken overnight at 37°C under ambient atmospheric conditions. Cultures were then diluted 1:1000 in the combined human urine and grown for 6 hours to mid-log phase (37°C, shaking), under 4% oxygen to emulate the bladder environment. After 7 hours, OD$_{600}$ of each isolate was measured – and cultures were normalized by volume to yield equal number of organisms from each strain – prior to pooling into groups of isolates (Supplementary Data – worksheets 4-5). CFUs were enumerated for each pool to confirm bacterial density (~$10^9$ total E. coli per pool). Each pool was then centrifuged to separate the cellular (pellet) and supernatant fraction. Pellets and supernatants were flash frozen and stored at -80°C until for metabolomic analysis. An uninoculated urine sample was included for analyses.

2.5. LC-MS Sample Preparation

Global untargeted metabolomic analyses were performed on the supernatant-fraction of three ASB and nine cystitis pools or five ASB and five cystitis pools, for preliminary and validation analyses, respectively, prepared as described above. Aliquots of each pool (200μL) were added to individual Eppendorf tubes containing 200μL ice cold lysis buffer (1:1:2, acetonitrile (ACN):methanol (MeOH):ammonium bicarbonate (0.1M, pH 8.0) (LC-MS grade). Labeled creatinine-D3 and lysine-D4 were added to each sample to assess the metabolite extraction (sample preparation) step. Samples were first subjected to protein precipitation by addition of 800μL of ice-cold methanol (4x by volume), then incubated at -80°C overnight. centrifuged (10,000 rpm, 10 min) to pellet precipitated proteins; the metabolite-containing supernatant was
transferred to a clean Eppendorf tube, dried in vacuo, and stored at -80°C until further LC-MS analysis.

The pellet-fraction of each sample pool prepared as described above was lysed using 400μL ice cold lysis buffer (1:1:2, ACN:MeOH:ammonium bicarbonate (0.1M, pH 8.0) (LC-MS grade), followed by sonication in an ice bath for 10 min. Sample volume for each pool was adjusted such that all samples have the same cell number in each vial. Labeled creatinine-D3 and lysine-D4 were added to each sample to assess the metabolite extraction (sample preparation) step. Samples were first subjected to protein precipitation by addition of 1000μL of ice-cold methanol (4x by volume), then incubated at -80°C overnight. Following incubation, samples were centrifuged (10,000 rpm, 10 min) to pellet precipitated proteins; the metabolite-containing extract was transferred to a clean Eppendorf tube, dried in vacuo and stored at -80°C until further LC-MS analysis.

2.6. Global, Untargeted UPLC-MS/MS Analysis

Dried extracts were reconstituted in 100μL of 98:2 H2O:ACN with 0.1% formic acid and centrifuged for 5 min at 15,000 rpm to remove insoluble material. Labeled valine-D8 was added to each sample for downstream MS quality assessment. A quality control (QC) sample pool was prepared by combining equal volumes from each experimental sample. Full MS (FMS) data was acquired for this QC pool to use as a retention time alignment reference within Progenesis QI (version 2.1) for subsequent normalization and data quantitation. MS/MS (data dependent (DD)) acquisitions for the pooled QC sample pool were used for compound annotation. Quality Control Reference Material sample (LC-MS Reference Mix) injections bracketing entire experimental run sequence (before, during, and after experimental samples) were used as an additional measure of quality control.

MS analyses were performed on a Q-Exactive HF hybrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a Vanquish UHPLC binary system and autosampler (Thermo Fisher Scientific, Germany). Extracts (5μL injection volume) were separated on a Hypersil Gold, 1.9 μm, 2.1mm x 100 mm column (Thermo Fisher) held at 40°C. Liquid chromatography was performed at a 250 μL min⁻¹ using solvent A (0.1% formic acid (FA) in water) and solvent B (0.1% FA in acetonitrile (ACN)) with the following gradient: 5% B for 1 min, 5-50% B over 9 min, 50-70% B over 5 min, 70-95% B over 5 min, 95% B held 2 min, and 95-5% B over 3 min, 5% B held 5 min (gradient length 30 min). MS analyses were acquired over a mass range of m/z 70-1050 under an ESI positive profile mode. Full mass scan was used at a resolution of 120,000 with a scan rate at ~3.5 Hz. The automatic gain control (AGC) target was set at 1 x 10⁶ ions, and maximum ion injection time (IT) was at 100 ms. Source ionization parameters were optimized with the spray voltage at 3.0 kV, and other parameters were as follows: transfer temperature at 280°C; S-Lens level at 40; heater temperature at 325°C; Sheath gas at 40, Aux gas at 10, and sweep gas flow at 1. Data dependent (DD) MS/MS spectra were acquired using a data dependent scanning mode in which one full MS scan (m/z 70-1050) was followed by two MS/MS scans. MS/MS scans were acquired in profile mode using an isolation width of 1.3 m/z, stepped collision energy (NCE 20, 40), and a dynamic exclusion of 4 s. MS/MS spectra were collected at a resolution of 15,000 with an AGC target set at 2 x 10⁵ ions, and IT of 100 ms.

2.7. Metabolite data processing and analysis

UPLC-MS/MS raw data were imported, processed, normalized, and reviewed using Progenesis QI v2.1 (Non-linear Dynamics, Newcastle, UK). All FMS sample runs were aligned against a FMS QC pool reference, with alignment to the reference being ≥ 96%, demonstrating the reproducibility of the RPLC separation method. Peak picking, with a minimum threshold of 500,000 ion intensity, was performed for individual aligned runs based on an aggregate run (representative of all ion peaks detected in all samples). Unique features (retention time and m/z pairs)
were grouped (a sum of the abundancies of unique ions) using both adduct and isotope deconvolutions to generate unique “compounds” (retention time and m/z pairs) representative of unannotated metabolites. Data were normalized to all compounds. Based on the aggregate matrix, all runs have a measurement for every compound ion, therefore a ratio can be calculated for the compound ion abundance in a particular run relative to the value in the normalization reference. A Log_{10} transformation to the ratio was performed to yield a normal distribution on all ratio data within each run for all samples, and scalar estimations shift the Log_{10} distributions onto that of the normalization reference. Subsequent FMS data was utilized for relative quantification. Compounds with <20% coefficient of variance (%CV) (calculated within all FMS QC pool injections) were retained for further analysis. A one-way analysis of variance (ANOVA) test was used to assess significance between ASB and Cystitis groups and returned a p-value for each feature (retention time_m/z descriptor), with a nominal p-value ≤0.05 and q-value ≤0.10 required for significance. Tentative and putative annotations were determined using experimental accurate mass measurements (≤5 ppm error), isotope distribution similarity, and fragmentation spectrum matching (when available) by searching an in-house library, the Human Metabolome Database (HMDB), METLIN, MassBank, Waters Metabolic Profiling CCS Library, LipidBlast, LIPID MAPS, and NIST 14 databases. Confidence levels of candidate metabolites were assigned as previously described [5]. Chemometrics and hierarchical cluster analyses were performed on peak intensity values using the MetaboAnalyst 4.0 to generate principal component analysis (PCA) plots and heat maps, respectively, for log transformed, Pareto scaled data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank members of the Hadjifrangiskou and Schmitz laboratories and Division of Pediatric Urology for their critical feedback throughout this work. The Vanderbilt Initiative for Personalized Microbial Discovery and Innovation (supported by the Vanderbilt Trans-Institutional Programs and the Vanderbilt Institute for Infection, Immunology & Inflammation) provided the clinical isolates for this project as well as financial support. The Center for Innovative Technology (CIT) at Vanderbilt University provided resources that supported this work. The VUMC Department of Pathology, Microbiology and Immunology to JES; K08 grant 5K08DK106472 to DBC; and R01 AI107052 to MH provided additional financial support.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.105811.

References

[1] AR Eberly, CJ Beebout, CM Carmen Tong, GT Van Horn, HD Green, MJ Fitzgerald, et al., Defining a Molecular Signature for Uropathogenic versus Urocolonizing Escherichia coli: The Status of the Field and New Clinical Opportunities. J Mol Biol 432 (2020) 786–804.
[2] GA O’Toole, LA Pratt, PI Watnick, DK Newman, VB Weaver, R Kolter, Genetic approaches to study of biofilms, Methods Enzymol 310 (1999) 91–109.
[3] L Cegelski, JS Pinkner, ND Hammer, CK Cusumano, CS Hung, E Chorell, et al., Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation, Nat Chem Biol 5 (2009) 913–919.
[4] C Hung, Y Zhou, JS Pinkner, KW Dodson, JR Crowley, J Heuser, et al., Escherichia coli biofilms have an organized and complex extracellular matrix structure, MBio 4 (2013) e00645-13.

[5] AC Schrimpe-Rutledge, SC Codreanu, SD Sherrod, JA McLean, Untargeted Metabolomics Strategies—Challenges and Emerging Directions, J Am Soc Mass Spectrom 27 (2016) 1897–1905.

[6] McCarter YS BE, Hall GS, Zervos M. Cumitech 2C, Laboratory diagnosis of urinary tract infections: ASM Press, Washington, DC; 2009.