









RESEARCH ARTICLE

 OPEN ACCESS 

## Are *Escherichia coli* causing recurrent cystitis just ordinary uropathogenic *E. coli* (UPEC) strains?

Nicolas Vautrin <sup>a</sup>, Sandrine Dahyot <sup>b</sup>, Marie Leoz<sup>a</sup>, François Caron<sup>c</sup>, Maxime Grand<sup>a</sup>, Audrey Feldmann<sup>a</sup>, François Gravey <sup>d</sup>, Stéphanie Legris<sup>a</sup>, David Ribet <sup>e</sup>, Kévin Alexandre <sup>c</sup>, and Martine Pestel-Caron <sup>c</sup>

<sup>a</sup>Univ Rouen Normandie, Université de Caen Normandie, INSERM, Normandie Univ, DYNAMICURE UMR 1311, Rouen, France; <sup>b</sup>Department of Microbiology, Univ Rouen Normandie, Université de Caen Normandie, INSERM, Normandie Univ, DYNAMICURE UMR 1311, CHU Rouen, Rouen, France; <sup>c</sup>Department of Infectious Diseases, Univ Rouen Normandie, Université de Caen Normandie, INSERM, Normandie Univ, DYNAMICURE UMR 1311, CHU Rouen, Rouen, France; <sup>d</sup>Université de Caen Normandie, Univ Rouen Normandie, INSERM, Normandie Univ, DYNAMICURE UMR 1311, Caen, France; <sup>e</sup>Univ Rouen Normandie, INSERM, Normandie Univ, ADEN UMR 1073, Nutrition, Inflammation and Microbiota-Gut-Brain, Axis, France

### ABSTRACT

Specific determinants associated with Uropathogenic *Escherichia coli* (UPEC) causing recurrent cystitis are still poorly characterized. Using strains from a previous clinical study (Vitale study, [clinicaltrials.gov](https://clinicaltrials.gov), identifier NCT02292160) the aims of this study were (i) to describe genomic and phenotypic traits associated with recurrence using a large collection of recurrent and paired sporadic UPEC isolates and (ii) to explore within-host genomic adaptation associated with recurrence using series of 2 to 5 sequential UPEC isolates. Whole genome comparative analyses between 24 recurrent cystitis isolates (RCIs) and 24 phylogenetically paired sporadic cystitis isolates (SCIs) suggested a lower prevalence of putative mobile genetic elements (MGE) in RCIs, such as plasmids and prophages. The intra-patient evolution of the 24 RCI series over time was characterized by SNP occurrence in genes involved in metabolism or membrane transport and by plasmid loss in 5 out of the 24 RCI series. Genomic evolution occurred early in the course of recurrence, suggesting rapid adaptation to strong selection pressure in the urinary tract. However, RCIs did not exhibit specific virulence factor determinants and could not be distinguished from SCIs by their fitness, biofilm formation, or ability to invade HTB-9 bladder epithelial cells. Taken together, these results suggest a rapid but not convergent adaptation of RCIs that involves both strain- and host-specific characteristics.

### ARTICLE HISTORY

Received 6 May 2024  
Revised 27 August 2024  
Accepted 13 December 2024

### KEYWORDS

Uropathogenic *Escherichia coli*; recurrent cystitis; urinary tract infections; virulence factors; whole genome sequencing; genome-wide association studies

## Introduction

Urinary tract infections (UTIs) are very common bacterial infections in women: more than half will develop UTI(s) during their lifetime [1], and up to 25% of women with UTI will experience a second one within a year [2]. If a woman presents more than 2 episodes within a 6-month period, she will be considered as suffering from recurrent UTI (rUTI) [3]. rUTI is a public health concern, associated with an economic, societal, and personal burden. It represents 1% to 6% of all medical visits and is the second leading cause of antibiotic consumption in the United States [2]. Its annual cost is estimated at 1.6 billion US dollars [4].

rUTIs – whether relapses or reinfections – are mainly caused by uropathogenic *Escherichia coli* (UPEC) [2]. Relapse is defined by infection with the same strain, whereas reinfection corresponds to infection with a new strain [5]. It was previously estimated

that 47 to 81% of UPEC rUTIs were due to relapses [3], based on typing methods such as serotyping [6] and pulsed-field gel electrophoresis (PFGE) [5,7,8], which have limited discriminatory power compared to molecular-based methods [9]. Using next generation sequencing (NGS) and CH typing – a molecular typing method based on the polymorphism of internal fragments of two genes (*fumC* and *fimH*) [10] – we recently observed that relapse frequency was only 30.6% [11].

UPEC virulence relies on their ability to survive and grow in urine, as well as to adhere to and invade urothelial cells [12]. However, the physiopathology of recurrent cystitis remains poorly characterized [13]. UPEC persistence in the bladder as intracellular bacterial communities (IBC) and quiescent intracellular reservoir (QIR) may promote recurrent cystitis [14–16], but this phenomenon has not been explored using large cohorts of patients. Moreover, no genomic nor

**CONTACT** Nicolas Vautrin  [vautrin.nicolas@outlook.com](mailto:vautrin.nicolas@outlook.com)

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/21505594.2024.2444689>.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

phenotypic marker has been identified to distinguish recurrence-associated UPECs from those causing sporadic (or acute) cystitis [17–19].

Thus, using a clinically well-characterized collection, we compared 24 recurrent cystitis isolates (RCIs) and 24 phylogenetically paired sporadic cystitis isolates (SCIs) to identify genomic and/or phenotypic determinants associated with recurrence. Moreover, we described within-host microevolution of the 24 RCIs over time, by analysing the genomic and phenotypic changes in 24 series of 2 to 5 sequential isolates.

## Methods

### Ethics and statements

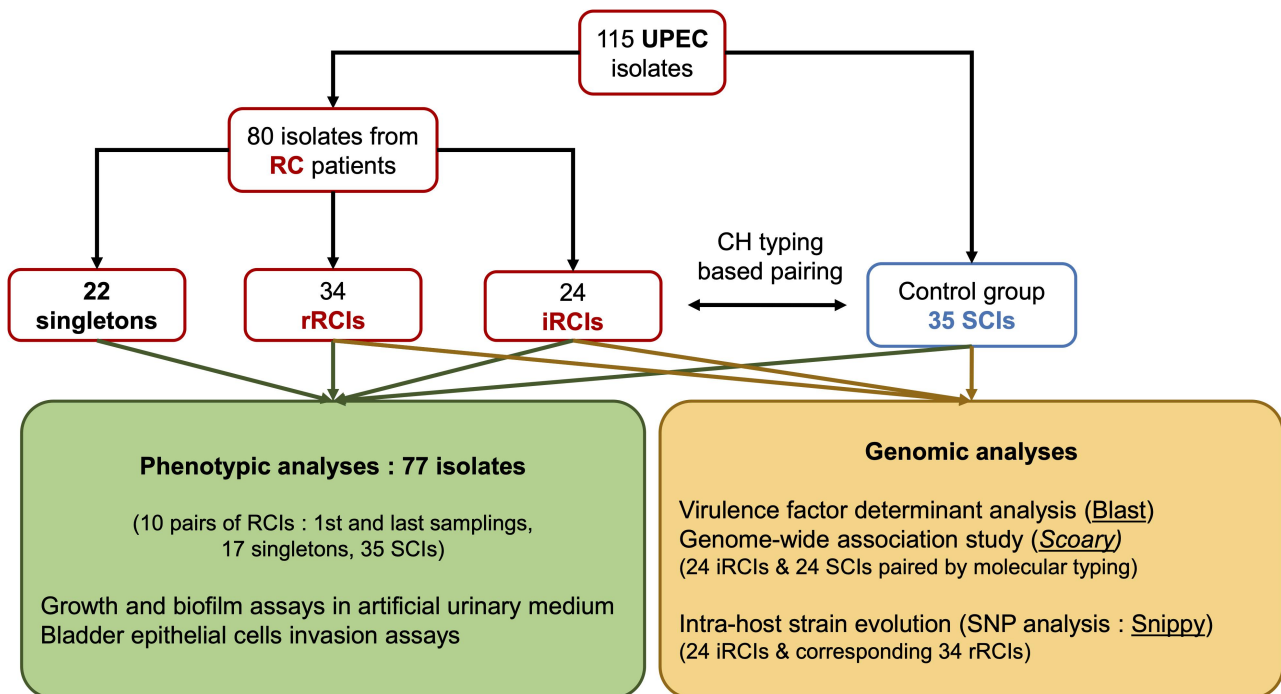
The UPEC collection derives from a prospective epidemiological study on community-acquired UTI over a 17-month period, funded by the French Ministry of Health and approved by the Medical Research Ethics Committee of the Rouen University hospital (VITALE study, Clinicaltrials.gov, identifier: NCT02292160). Patients received an information letter and provided written informed consent. The declaration of Helsinki for medical research was strictly followed.

### Bacterial isolates

We used 115 UPEC isolates from the VITALE collection (Figure 1), including 80 isolates from patients with recurrent cystitis ( $\geq 2$  episodes over 6 months). Of these, 22 were responsible for a single cystitis episode – defined as “singletons” – and 58 were involved in series of cystitis relapses (cystitis episodes caused by a single strain) – defined as “recurrent cystitis isolates (RCIs)” [11]. The 58 RCIs corresponded to 24 strains, each with an initial RCI (iRCIs) followed by up to 4 sequential relapse isolates per strain, for a total 34 relapse RCIs (rRCIs). The antibiotic susceptibility profiles of all isolates were previously determined using the Vitek 2 system (AST-N372 card) [11]. The results are available in the Table S1.

A control group included 35 UPEC “sporadic isolates” (SCIs) collected from patients experiencing only one acute cystitis episode over the VITALE study. They were selected based on their genetic proximity with the RCIs, according to CH typing data [11]. Proximity between CH types was determined using a phylogeny based on concatenated sequences of *fumC* and *fimH* [11].

Comparative genomic studies were performed on 48 isolates: the 24 iRCIs and 24 phylogenetically paired SCIs from the 35 SCIs control group. Intra-host evolution studies were based on the 58 RCIs (24 iRCIs and



**Figure 1.** Isolate and workflow description. RC patients = patients with recurrent cystitis, iRCI = initial recurrent cystitis isolate, rRCI = recurrent cystitis isolate associated with relapse(s), SCI = sporadic cystitis isolate.

34 rRCIs). Phenotypic studies were performed on 77 isolates (Figure 1).

*E. coli* str. UTI89 and *E. coli* str. K12 substr. MG1655 was used as control strain for phenotypic experiments.

### Whole genome sequencing and assembly

Short-read whole genome sequencing (WGS) and assembly were performed on SCIs and RCIs as previously described [11] (Table S2). Long-read WGS was also performed for each iRCI. Up to  $2.10^9$  cells were used for genomic DNA extraction using the DNeasy Blood and Tissue minikit (Qiagen) for isolating Gram-negative bacterial DNA. Alternatively, the Monarch<sup>®</sup> HMW DNA Extraction kit for Tissue (New England Biolabs) was used to improve DNA fragment length.

Up to 1  $\mu$ g genomic DNA was used for library preparation according to the « Native barcoding genomic DNA protocol with EXP-NBD104, EXP-NBD114 and SQK-LSK109 » (Oxford Nanopores Technologies). Libraries were sequenced using R9.4.1 flowcells on a minION Mk1B device. Real-time basecalling, demultiplexing, and filtering were performed using MinKNOW v21.11.8 (ONT, Oxford, UK) and Guppy v5.1.13 (Oxford Nanopores Technologies). Quality control was performed using Nanoplot v1.38.1 [20].

Short and long reads from Illumina and ONT sequencing, respectively, were used for hybrid assembly using Unicycler 0.4.9. Short-read assemblies from the previous study [11] and newly obtained hybrid assemblies were compared using Bandage v0.8.1 [21] for visualization and BUSCO v5.2.2 [22] with the enterobacterales\_odb10 lineage dataset for genome completeness evaluation. Sequencing metrics are available in Table S2. All genome assemblies were annotated using Prokka 1.14.5 [23].

### SCI and RCI comparative genomic study

The protein sequences of VFDs from Ejrnaes et al. [17] were retrieved from the annotated genomes of phylogenetically paired iRCIs and SCIs and compared using MEGA X [24]. The virulence score corresponded to the total number of VFDs for a given isolate. RCI and SCI short reads were also mapped to the sequence of pUTI89 (NC\_007941.1) using Snippy v4.3.6 ([github.com/tseemann/snippy](https://github.com/tseemann/snippy)). More than 75% coverage of pUTI89 sequence was interpreted as positive match.

Pan-genome analysis of the 48 RCIs and SCIs was performed using Roary 3.12.0 [25]. Neighbor-joining tree inference was performed using MEGA X (50) based on the translated core gene alignment produced

by Roary. The phylogenetic tree, the core genome allele profiles obtained from the raw short reads using cgMLSTFinder v1.1.5 [26,27] and the presence/absence table produced by Roary were visualized using Phandango v1.3.0 [28].

The presence/absence table produced by Roary was also used for GWAS with Scoary v1.6.16 [29], to identify recurrence-associated genes. The protein sequences deriving from these genes were submitted to BlastKOALA [30] and Kegg Mapper Reconstruct [31] to classify them according to functional pathways.

The 48 annotated draft genomes were visualized using Proksee [32] and searched for phage regions between the *cpx* operon and *fieF* gene using Phigaro 2.3.0 [33]. Reads from each RCI were mapped to its paired SCI draft genome using snippy v4.3.6 and visualized using Proksee to explore large sequences that were preferentially found in SCI genomes.

### Intra-patient analysis of RCI gene evolution

Short reads from each rRCI were mapped to their respective iRCI annotated hybrid genome assembly using Snippy v4.3.6 and visualized using Proksee to identify contig loss over time. Lost contigs were blasted (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to identify whether they were plasmidic, and their gene content was analysed.

Intra-patient SNP analysis was performed by mapping short reads from each rRCI to their respective iRCI annotated draft genome assembly using Snippy v4.3.6. SNPs identified when mapping reads from each iRCI to their own annotated draft genome assembly were considered as artifacts and removed from the series analysis. All the genes in which intra-patient SNPs were identified were listed and functionally classified using BlastKOALA [30] and Kegg Mapper Reconstruct [31].

Evolution rates were calculated based on the number of SNPs between each iRCI/rRCI pair divided by the time (days) elapsed between their samplings. Evolution rate plot was constructed using Microsoft Excel (Microsoft corporation Available from: <https://office.microsoft.com/excel>). Correlation between time and evolution rate was tested with Spearman's rank correlation test, using R software (V 4.1.3, R Foundation).

### Growth assays

The artificial urinary medium (AUM) was prepared based on Brooks and Keevil [34], with 100 mm 2-(N-morpholino) ethanesulfonic acid (MES) to improve buffer capacity and limit precipitate formation (Table S3). Isolates were grown overnight in LB or

AUM at 37°C with shaking (150 rpm). 200 µL of standardized suspension ( $OD_{600\text{ nm}} = 0.05$ ) was incubated in triplicate with shaking (108 rpm) at 37°C during 24 h in a microplate reader (Spark®, Tecan) and  $OD_{600\text{ nm}}$  was measured every 15 min. Each experiment was performed at least three times, with *E. coli* K12 and UTI89 strains as internal controls and fresh medium as negative control. Statistical analyses of the growth curves were performed using R software (v 4.2.1, <https://www.R-project.org/>).

### Biofilm assays

Isolates were grown to late stationary phase in LB or AUM; standardized overnight cultures were incubated in static conditions at 37°C for 24 h. Non-adherent bacteria were washed using H<sub>2</sub>O before adding crystal violet 0.5% and incubating 10 min under gentle shaking at room temperature. After washing 3 times with H<sub>2</sub>O and adding Ethanol 95%, the plate was incubated under gentle shaking at room temperature for 10 min and absorbance at 590 nm ( $A_{590\text{ nm}}$ ) was measured. Each experiment was performed at least three times, with *E. coli* str. K12as positive control and fresh medium as negative control. Data were expressed as percent of biofilm formation relative to the positive control.

### Gentamicin protection assay

Human bladder epithelial cell (BEC) line 5637 (ATCC HTB-9) was maintained at 37°C and 5% CO<sub>2</sub> in RPMI 1640 medium (ThermoFisher Scientific) supplemented with 10% fetal bovine serum (FBS) and 2 mM Glutamine.

Bacterial isolates were grown in LB at 37°C for 48 h in static conditions to promote type I pilus expression [35]. Bacteria were washed three times in phosphate buffer saline (PBS) and used as inoculum for BEC infections at multiplicity of infection of 15. BECs were seeded the day before at  $1.6 \times 10^5$  cells/cm<sup>2</sup>. Before infection, culture medium was replaced by RPMI with 5% FBS. After bacterial inoculation, plates were centrifuged at 750 rpm for 5 min at room temperature and incubated for 2 hours at 37°C. Cells were washed twice with PBS and incubated 1 hour at 37°C in culture medium supplemented with gentamicin (50 µg/mL). After washing three times with PBS, BECs were lysed in PBS-0.2% Triton-X-100 solution for 10 min at 37°C and lysates were plated on LB agar plates to obtain the

number of CFUs. Five replicates per isolates were performed. Results were expressed as invasion percentages:

$$\text{Invasion percentage}(\%) = \frac{\text{Intracellular bacteria}}{\text{Inoculated bacteria}} \times 100$$

### Fluorescence microscopy

BECs were seeded on glass slides and infected with four UPEC isolates (UTI89, rRCI\_2627, SCI\_2259 and SCI\_2263) as described above. After 3 hours of infection, cells were washed three times with PBS and fixed in PBS-4% paraformaldehyde. Intracellular bacteria were identified by differential staining [36] (Table S4) using antibodies in PBS-1% BSA. First round of staining consisted in a primary labelling for 1 hour at room temperature, cells were washed, labeled for 1 hour with the secondary antibody, washed with PBS and permeabilized in PBS-0.3% A second round of staining was performed (Table S4). Slides were mounted in Fluoromount reagent (Invitrogen), images were acquired with a Leica Thunder tissue 3D microscope and processed with ImageJ software [37].

### Statistical analyses

All statistical analyses were performed using R (V 4.1.3, R Foundation). Comparison of proportions was performed using Pearson's chi-squared test when applicable. Otherwise, we used Fisher's exact test for count data. Global means comparisons were performed using Kruskal-Wallis's rank sum test, and pairwise comparisons were performed using Wilcoxon, Mann-Whitney test.

## Results

### Virulence factor determinants did not distinguish RCIs from SCIs

To identify genomic traits associated with recurrence, we compared the presence and protein sequence of 26 virulence factor determinants (VFDs) (17) (Table 1) in the genomes of 24 initial RCI (iRCIs) with those of 24 SCIs that phylogenetically matched based on CH typing (Figure 1, Table S2).

The prevalence of the 26 VFDs ranged from 0.0% to 95.8% (Table 1). The global prevalence of toxin-encoding genes and of the *flu* gene associated with biofilm production was low (<25%), while the prevalence of genes encoding protectins was



**Table 1.** Prevalence of virulence factors determinants (VFDs) in the 24 initial recurrent cystitis isolates (iRCIs) and 24 sporadic cystitis isolates (SCIs.) NA: not applicable. PUT189: a plasmid associated with iron uptake (Cusumano et al., 2010).

	Total (n, %)	Total allele number	iRCIs (n = 24)	SCIs (n = 24)	<i>p. value</i> #1	<i>p. value</i> #2
Aggregate VFD score median (range)	10 (2–14)	NA	9 (3–13)	10.5 (2–14)	0.582	NA
Virulence factor determinants count (%)						
Adhesins						
<i>fimH</i>	46 (95.8%)	17	23 (95.8%)	23 (95.8%)	1 <sup>□</sup>	NA
<i>papG</i>	15 (31.3%)	8	7 (29.2%)	8 (33.3%)	1*	0.23*
<i>focD</i>	11 (22.9%)	10	6 (25.0%)	5 (20.8%)	1*	0.45*
<i>iha</i>	10 (20.8%)	4	6 (25.0%)	4 (16.6%)	0.722*	1*
<i>afaE1</i>	2 (4.2%)	1	1 (4.2%)	1 (4.2%)	1 <sup>□</sup>	NA
<i>bmaE</i>	2 (4.2%)	2	1 (4.2%)	1 (4.2%)	1 <sup>□</sup>	1*
<i>focG</i>	1 (2.1%)	1	1 (4.2%)	0 (0.0%)	1 <sup>□</sup>	NA
<i>afaE3</i>	0 (0.0%)	0	0 (0.0%)	0 (0.0%)	NA	NA
<i>draA</i>	0 (0.0%)	0	0 (0.0%)	0 (0.0%)	NA	NA
Biofilm related						
<i>flu</i>	11 (22.9%)	10	3 (12.5%)	8 (33.3%)	0.169*	0.56*
Iron uptake						
<i>fyuA</i>	41 (85.4%)	9	21 (87.5%)	20 (83.3%)	1 <sup>□</sup>	1*
<i>chuA</i>	40 (83.3%)	10	21 (87.5%)	19 (79.2%)	0.700 <sup>□</sup>	0.35*
<i>iroN</i>	28 (58.3%)	13	13 (54.2%)	15 (62.5)	0.770*	0.77*
<i>iutA</i>	23 (47.9%)	7	10 (41.7%)	13 (54.2%)	0.563*	0.94*
<i>ireA</i>	10 (20.8%)	4	6 (25.0%)	4 (16.6%)	0.722*	1*
pUTI89	8 (16.7%)	NA	3 (12.5%)	5 (20.8%)	0.700 <sup>□</sup>	NA
Protectins						
<i>iss</i>	42 (87.5%)	9	21 (87.5%)	21 (87.5%)	1 <sup>□</sup>	1*
<i>traT</i>	36 (75.0%)	11	16 (66.7%)	20 (83.3%)	0.317*	0.09*
<i>kpsM</i>	34 (70.8%)	7	18 (75.0%)	16 (66.7%)	0.751*	0.88*
Toxins						
<i>hlyA</i>	9 (18.8%)	6	3 (12.5%)	6 (25.0%)	0.461 <sup>□</sup>	0.11*
<i>sat</i>	9 (18.8%)	4	5 (20.8%)	4 (16.6%)	1 <sup>□</sup>	0.52*
<i>cnf1</i>	7 (14.6%)	5	5 (20.8%)	2 (8.3%)	0.416 <sup>□</sup>	1*
<i>cdtB</i>	0 (0.0%)	0	0 (0.0%)	0 (0.0%)	NA	NA
Miscellaneous						
<i>malX</i>	43 (89.6%)	18	20 (83.3%)	23 (95.8%)	0.347 <sup>□</sup>	0.71*
<i>ibeA</i>	12 (25.0%)	5	7 (29.2%)	5 (20.8%)	0.739*	0.67*
<i>usp</i>	10 (20.8%)	5	4 (16.6%)	6 (25.0%)	0.722*	0.62*

#1: H0 hypothesis: There is no difference of gene prevalence between iRCIs and SCIs.

#2: H0 hypothesis: Allele distribution among iRCIs and SCIs is identical.

\**p. value* determined using Fisher's exact test for count data.

<sup>□</sup>*p. value* determined using Pearson's Chi-squared test with Yates' continuity correction.

high (>70%). Iron-uptake and adhesin-encoding genes had very diverse prevalences. The great majority of the isolates (41/48, 85.4%) presented at least two iron uptake genes. Three isolates (one iRCI/SCI pair and one SCI) were negative for all iron uptake markers tested.

However, VFD prevalence and aggregate VFD score median [17] (9 for iRCIs, 10 for SCIs) were not significantly different between the two groups of isolates. Moreover, a principal component analysis based on the detection of the 26 VFDs revealed that iRCIs and SCIs did not cluster separately (Figure 2).

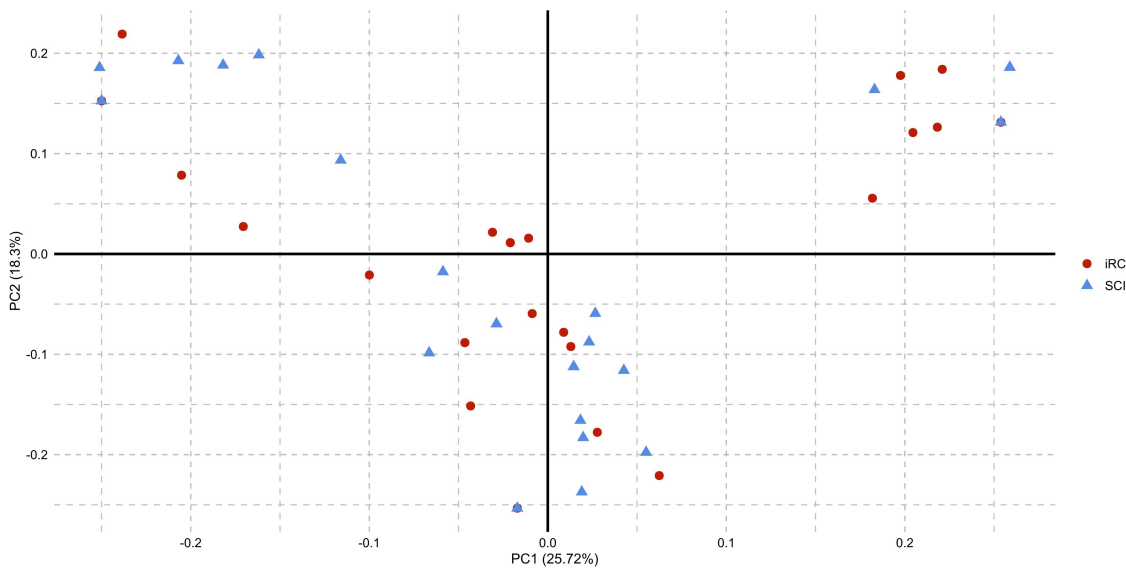
The level of polymorphism was high in some genes, with up to 18 alleles identified in *malX* (Table 1, Fig S1). Two of the four previously described *papG* variants [38,39] were identified (*papGII* and *papGIII*). However, no gene allele was significantly associated with recurrence.

### **iRCI genomes were significantly smaller than phylogenetically paired SCIs**

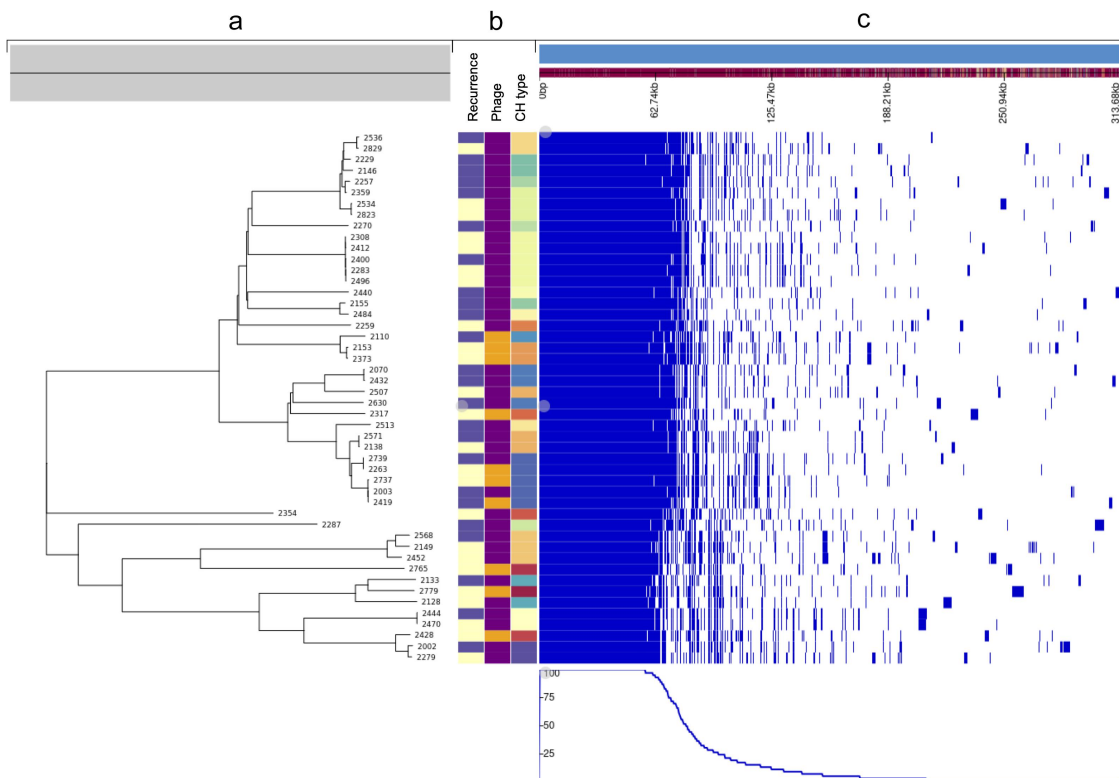
The mean genome size of iRCIs was significantly lower than that of SCIs (5.01 Mb vs. 5.11 Mb,

respectively,  $p < 0.05$ ). The pan-genome of the 48 assemblies contained 15,682 genes, of which 2,864 (18.3%) were considered core genes. Phylogenetic analysis of the core genes confirmed that isolates clustered based on CH type rather than on the sporadic/recurrent clinical context (Figure 3a), consistent with the SCI/iRCI CH type-based pairing strategy.

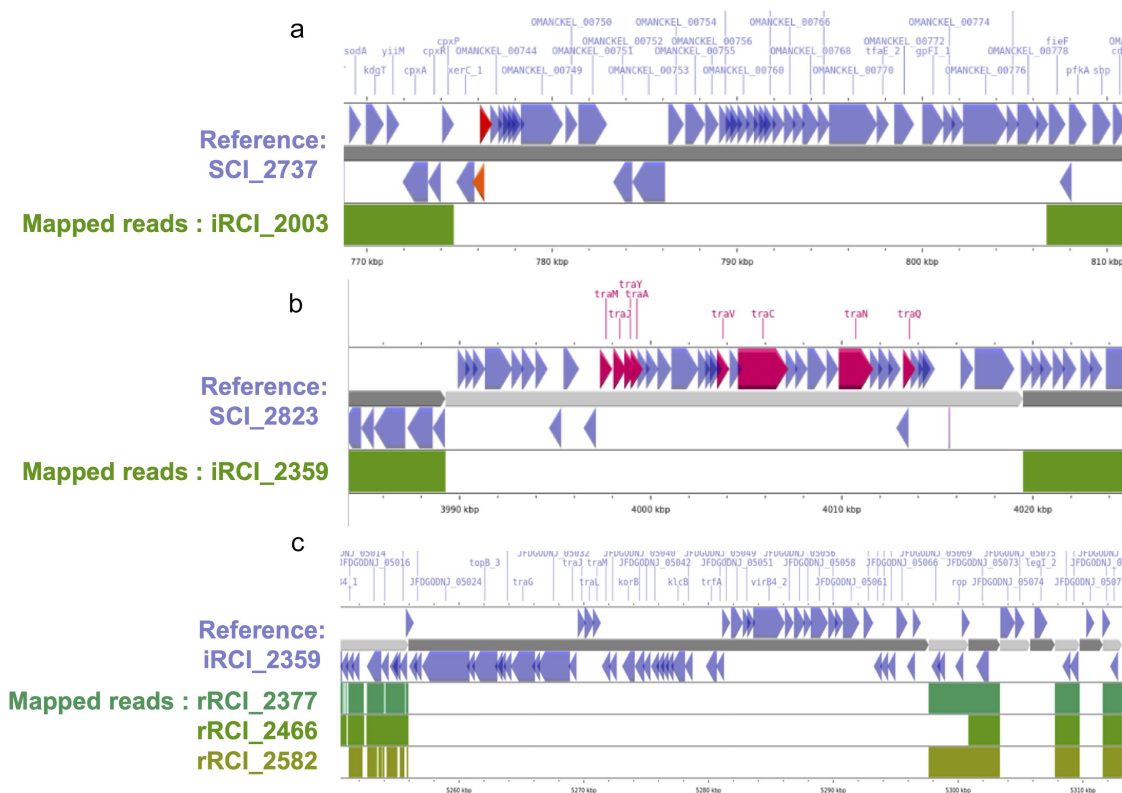
A genome-wide association study (GWAS) was performed based on the presence of accessory genes in iRCIs vs. SCIs (Figure 3c). Only two genes were significantly less frequent in RCI genomes (Sgene1,  $p = 0.009$  and Sgene2,  $p = 0.009$ ) and encoded phage-associated proteins according to InterProScan. Sgene1 encoded a DNA-binding protein (LocusTag OMANCKEL\_0742 in CS2737) and Sgene2 a regulatory protein from the Cox family (LocusTag OMANCKEL\_0743 in CS2737). In the 7 SCIs where Sgene1 and Sgene2 were identified, they were followed by up to 40 phage-associated proteins according to InterProScan. The prophage was located between the *cpX* operon and the *fieF* gene, encoding an envelope stress response-system and a ferrous iron efflux pump, respectively.



**Figure 2.** Principal component analysis based on the detection of virulence factor genes. Red dots represent iRCIs (initial recurrent cystitis isolates) and blue triangles represent SCIs (sporadic cystitis isolates).



**Figure 3.** Core genome analysis of the 24 iRCIs and their 24 CH type paired SCIs. iRCIs = initial recurrent cystitis isolates, SCIs = sporadic cystitis isolates. (a) Neighbor-joining phylogenetic tree based on the 48 strains' translated core gene alignment (2,755,382 amino acid positions). (b) Color-coded CH type, prophage presence (purple = absent, yellow = present), clinical context of recurrence (purple = iRCI, yellow = SCI) for the 48 strains. (c) Presence/absence (blue/white) of the 15,682 genes from the pan genome in each strain, sorted by gene frequency among the 48 strains. The first 2,864 genes on the left represent the core genome.



**Figure 4.** Mobile genetic elements found in SCIs but not in paired iRCIs (a and b) or lost in sequential relapses (c). grey arrows: contigs, iRCIs = initial recurrent cystitis isolates, shades of green = coverage by iRCI reads, rRCI = recurrent cystitis isolate associated with relapse(s), SCIs = sporadic cystitis isolates. Blue arrows = CDSs. A. The >30 kb genome region that is not covered by iRCI\_2003 reads includes 44 phage proteins similar to those from peduovirus P24B2 (AccNum NC\_049387). (b) Pink arrows = *tra* operon. The > 30 kb genome region not covered by iRCI\_2359 derives from a plasmid similar to the plasmid pHS13-1-IncF (AccNum CP026494) C. Contig #3 from hybrid genome assembly of iRCI\_2359 is indicated in dark grey. iRCI\_2359 contig #3 corresponds to a 42kb plasmid similar to the *E. coli* pRHB15-C18\_3, (AccMum CP057780.1) that is lost in the first to third relapses.

Systematic Phigaro analysis of the 48 draft genomes confirmed phage insertion between *cpx* and *fieF* in 8 SCIs, and only 2 of their paired iRCIs (Figure 3b).

iRCI reads were mapped to their paired SCI assembly to illustrate the phage presence/absence (Figure 4a). This analysis also highlighted that some very large SCI contigs (up to 40 kb) were absent in paired iRCI; some of these contigs contained genes from the *tra* operon, which suggested that they derived from plasmids (Figure 4b).

Taken together, these results suggested that iRCIs smaller genome size was linked to a lesser frequency of mobile genetic elements (MGEs).

### Longitudinally sampled RCIs lost plasmids over the course of relapses

Long-read whole genome sequencing and hybrid assemblies were performed on the 24 iRCIs in order to better characterize their genomes, and in particular, their plasmids. Hybrid assembly reduced by 19-fold the

mean number of contigs per genome and modestly increased the total genome size (mean of 5.10 Mb instead of 5.01 Mb), though mean estimated completeness remained 99.9% for hybrid as well as short-read assemblies according to BUSCO.

To investigate RCI evolution over time, short reads from each relapse-associated RCI (rRCI, one to four relapses per patient) were mapped to the hybrid assembly of their iRCI, used as an intra-host reference. Interestingly, 5 of the 24 RCIs (21%) lost up to 3 circular contigs over the course of relapses, identified as plasmids using blast. Of the 8 lost plasmids, 5 were small cryptical plasmids (<5kb) and 3 were large plasmids (>40kb) encoding conjugation system (*tra* operon), toxin-antitoxins systems, and a few proteins that could not be associated with urovirulence (Table S5). In 7 cases out of 8, plasmid loss occurred between the iRCI and the first rRCI (exemplified in Figure 4c).

**Table 2.** Functional classification of the 58 proteins encoded by genes in which non-synonymous SNPs have occurred overtime in the recurrent cystitis isolate (RCI) series. Annotation and corresponding biological process were provided by BlastKoala. Of note, a single gene can be involved in several biological processes.

Functionnal category	Number of genes
<b>Metabolism</b>	<b>(n = 22 genes)</b>
<b>Global and overview maps</b>	<b>17</b>
Diverses metabolic pathways	16
Biosynthesis of secondary metabolites	9
Microbial metabolism in diverse environments	4
Carbon metabolism	2
Biosynthesis of amino acids	1
Biosynthesis of nucleotide sugars	1
Biosynthesis of cofactors	3
Biosynthesis of aromatic compounds	1
<b>Carbohydrate metabolism</b>	<b>5</b>
Glycolysis/Gluconeogenesis	1
Citrate cycle (TCA cycle)	1
Pentose and glucuronate interconversions	1
Fructose and mannose metabolism	1
Amino sugar and nucleotide sugar metabolism	1
Pyruvate metabolism	2
Propanoate metabolism	2
Butanoate metabolism	1
C5-Branched dibasic acid metabolism	1
<b>Energy metabolism</b>	<b>2</b>
Carbon fixation pathways	2
Methane metabolism	1
<b>Lipid metabolism</b>	<b>1</b>
Fatty acid degradation	1
<b>Nucleotide metabolism</b>	<b>1</b>
Purine metabolism	1
<b>Amino acids metabolism</b>	<b>5</b>
Glycine, serine and threonine metabolism	1
Lysine degradation	1
Arginine and proline metabolism	1
Histidine metabolism	1
Tyrosine metabolism	1
<b>Metabolism of other amino acids</b>	<b>2</b>
Taurine and hypotaurine metabolism	1
Glutathione metabolism	1
<b>Glycan biosynthesis and metabolism</b>	<b>5</b>
Lipopolysaccharide biosynthesis	4
Exopolysaccharide biosynthesis	1
<b>Metabolism of cofactors and vitamins</b>	<b>3</b>
Biotin metabolism	1
Porphyrin metabolism	1
Ubiquinone and other terpenoid-quinones biosynthesis	1
<b>Metabolism of terpenoids and polyketides</b>	<b>1</b>
Biosynthesis of siderophore group nonribosomal peptides	1
<b>Biosynthesis of other secondary metabolites</b>	<b>1</b>
Tropane, piperidine and pyridine alkaloid biosynthesis	1
<b>Xenobiotics biodegradation and metabolism</b>	<b>4</b>
Chloroalkane and chloroalkene degradation	1
Naphthalene degradation	1
Metabolism of xenobiotics by cytochrome P450	1
Drug metabolism – cytochrome P450	1
Drug metabolism – other enzymes	1
<b>Environnemental information processing</b>	<b>(n = 15 genes)</b>
<b>Membrane transport</b>	<b>11</b>
ABC transporters	10
Bacterial secretion systems	1
<b>Signal transduction</b>	<b>3</b>
Two-component system	3
<b>Genetic information processing</b>	<b>(n = 6 genes)</b>
<b>Translation</b>	<b>2</b>
Ribosome	1
Aminoacyl-tRNA biosynthesis	1
<b>Folding, sorting and degradation</b>	<b>3</b>
RNA degradation	3
<b>Replication and repair</b>	<b>4</b>
DNA replication	2
Base excision repair	1

(Continued)

**Table 2.** (Continued).

Functionnal category	Number of genes
Mismatch repair	2
Homologous recombination	3
<b>Cellular processes</b>	<b>(n = 6 genes)</b>
<b>Cellular community – prokaryotes</b>	<b>5</b>
Quorum sensing	2
Biofilm formation – <i>Vibrio cholerae</i>	2
Biofilm formation – <i>Escherichia coli</i>	2
<b>Cell motility</b>	<b>1</b>
Flagellar assembly	1
<b>Human diseases</b>	<b>(n = 5 genes)</b>
<b>Cancer: overview</b>	<b>1</b>
Pathways in cancer	1
Chemical carcinogenesis – DNA adducts	1
Chemical carcinogenesis – receptor activation	1
Chemical carcinogenesis – reactive oxygen species	1
<b>Cancer: specific types</b>	<b>1</b>
Hepatocellular carcinoma	1
<b>Infectious disease: bacterial</b>	<b>1</b>
Shigellosis	1
Yersinia infection	1
Pertussis	2
Bacterial invasion of epithelial cells	1
<b>Infectious disease: parasitic</b>	<b>1</b>
Amoebiasis	1
<b>Cardiovascular disease</b>	<b>1</b>
Fluid shear and atherosclerosis	1
<b>Drug resistance: antineoplastic</b>	<b>1</b>
Platinum drug resistance	1
<b>Organismal systems</b>	<b>(n = 1 gene)</b>
<b>Aging</b>	<b>1</b>
Longevity regulating pathway – worm	1

### Longitudinally sampled RCIs acquired non-synonymous SNPs in metabolism and membrane transport-associated genes

To investigate RCI micro-evolution, we looked for SNP acquisition over time. In the 24 relapse series; a total of 666 SNPs were identified, including 255 (38.3%) non-synonymous SNPs (nsSNPs). Fifty-eight of the 160 proteins affected by nsSNPs were mapped in defined biological processes by blastKoala (Table 2); 37.9% of them were involved in diverse metabolic pathways, and 20.7% in membrane transport.

Notably, 10 out of the 24 RCIs acquired nsSNPs in diverse ABC membrane transporter encoding-genes, including 2 metal ion transporters (NikD in RCI12 and FecC in RCI20), and 1 lipopolysaccharide export protein (LptB) that acquired nsSNPs in two distinct relapse series RCI23 and RCI6). Of note, another nsSNP occurred in a gene involved in lipopolysaccharide biosynthesis (MsbA in RCI16).

Overtime-conserved SNPs were investigated in the 7 RCI series that contained more than 2 episodes. A total of 13 SNPs identified in the first rRCI were conserved in the following ones, 10 of which occurred in ORFs and 6 were non-synonymous. Two of the 6 genes that acquired conserved nsSNPs were involved in membrane transport, according to Prokka: an ABC transporter (YdcV in RCI2) and a permease (Dipeptide and



tripeptide permease A in RCI23). Other genes affected by overtime-conserved nsSNPs encoded ribonucleases and a sigma-E factor regulatory protein (Table S6).

### RCI genomic evolution rate tended to decrease overtime

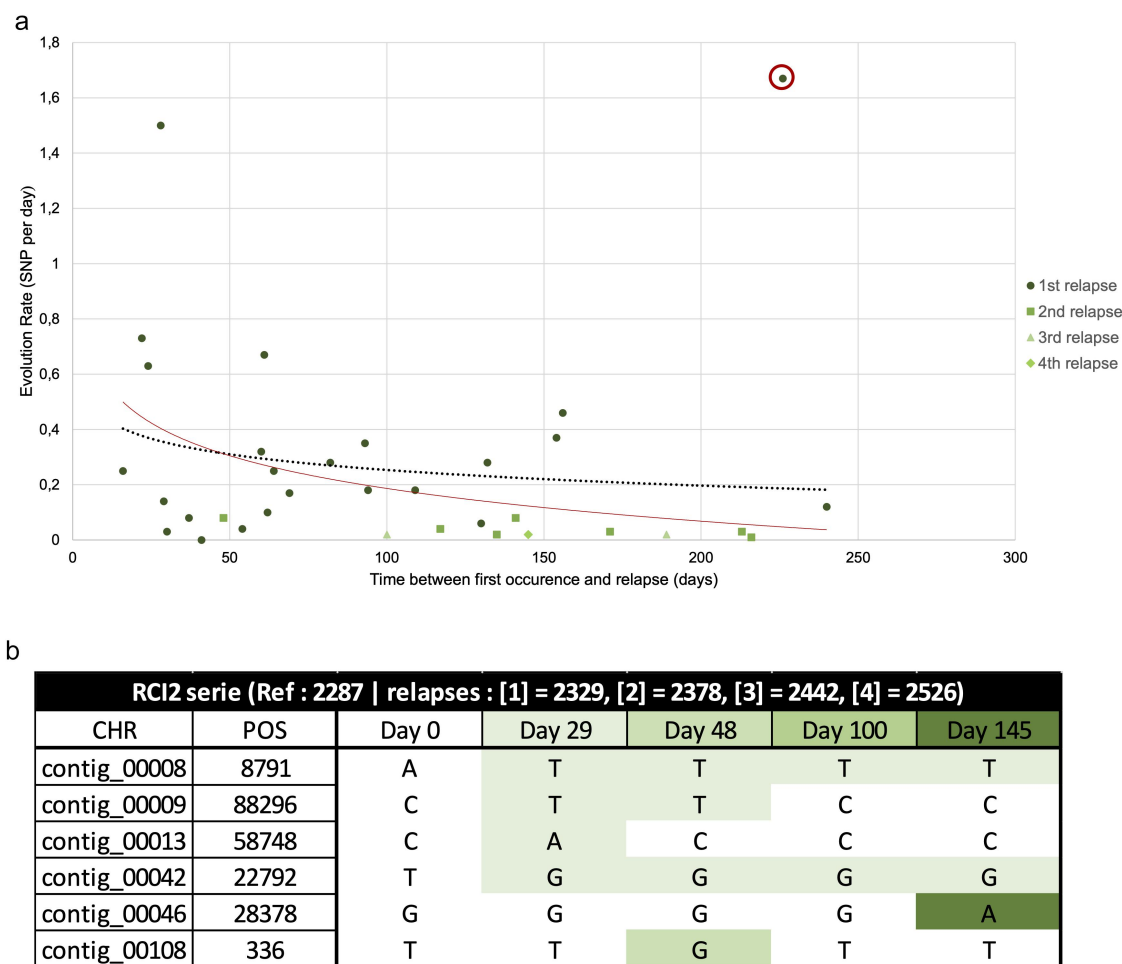
Evolution rates between each rRCI and its corresponding iRCI ranged from 0 to 1.67 SNPs/day (mean: 0.27 SNP/day). Evolution rates tended to decrease over time, but this trend was not significant ( $p = 0.124$ ) (Figure 5a) due to an outlier (circled in red on Figure 5a). An in-depth analysis suggested that 93% of the outlier SNPs were artefacts due to the comparison of similar regions that derived from different plasmids. When discarding this outlier, the overtime evolution rate decrease became significant ( $p = 0.028$ ).

This trend was also observed at the individual scale within the largest RCI serie (RCI2, four relapses): the

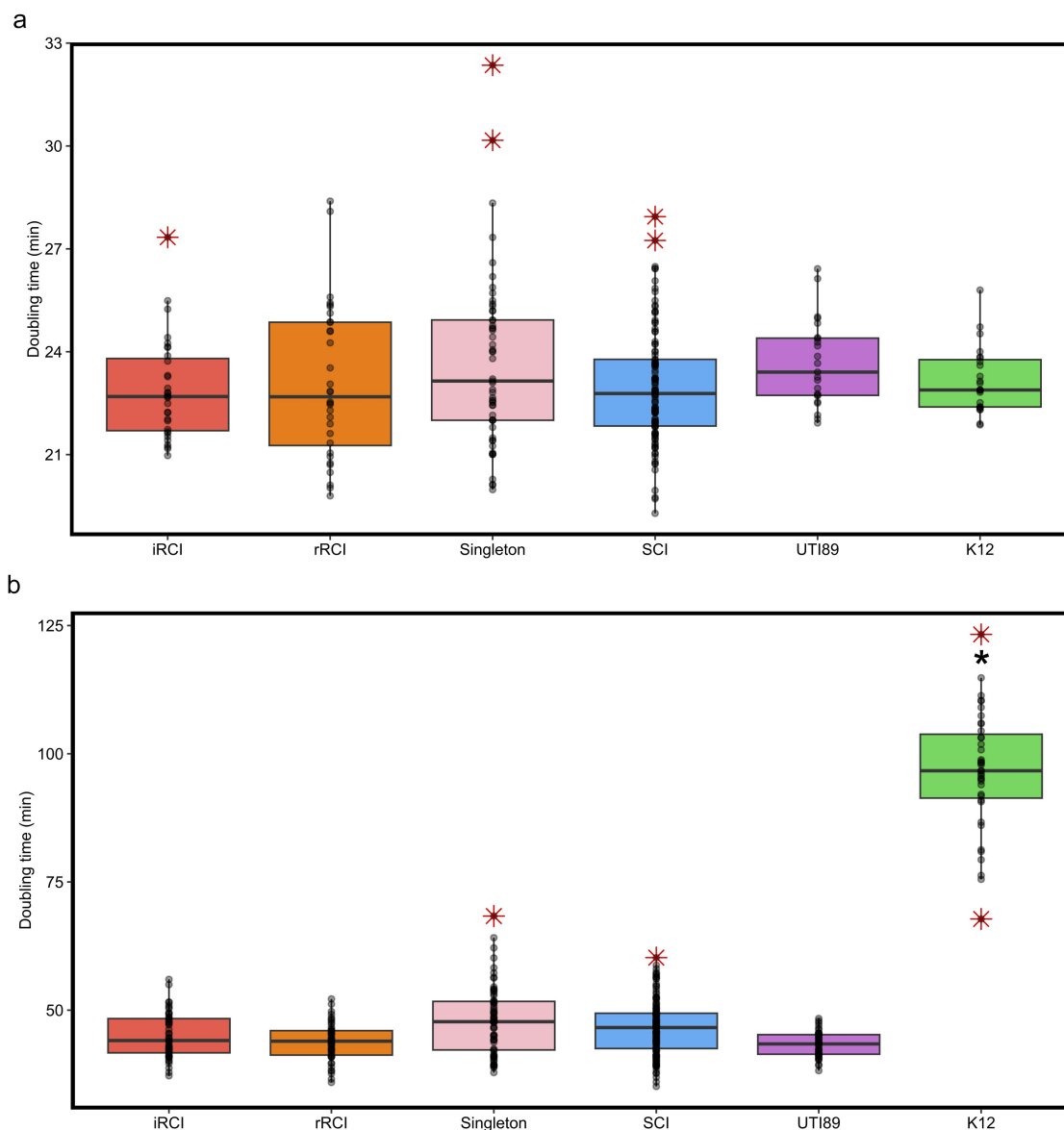
overtime-conserved SNPs occurred early, often between the iRCI and the first rRCI. Afterwards, a dynamic of SNP emergence and clearance was observed, as shown in Figure 5b.

### Growth in AUM did not discriminate RCIs from SCIs

The growth of 72 isolates was evaluated in rich medium (lysogenic broth [LB]) and in artificial urinary medium (AUM). These isolates included 10 RCI pairs (10 iRCIs and their corresponding 10 last rRCIs), 35 SCIs and 17 singletons isolates (isolates responsible for a single infection in patients with recurrent cystitis) (Figure 1). The mean doubling times of the four groups (iRCI, rRCI, SCI and singletons) were similar in LB ( $G_{iRCI} = 22.9$  min,  $G_{rRCI} = 23.2$  min,  $G_{SCI} = 23.0$  min and  $G_{Singleton} = 23.6$  min) and not statistically different from those of the reference strains UTI89 and K12 ( $G_{UTI89} = 23.7$  min and



**Figure 5.** Intra patient micro-evolution of RCIs by SNP acquisition. RCIs = recurrent cystitis isolates A. Global representation of the evolution rates (SNPs per day) from the 24 series of intra-patient RCIs depending on the time elapsed between the first occurrence and the relapse (days). An outlier is circled in red. The black dotted curve represents the non-significant trend when including all the dots ( $p = 0.124$ ). The red curve represents the significant trend when excluding the outlier ( $p = 0.028$ ). B. Individual representation of the core SNPs observed over time in the largest intra-patient RCI serie (RCI2). SNPs are color-coded depending of the time of emergence (the darker, the latter). CHR = contig on which the SNP was identified, POS = position of the SNP on the contig.



**Figure 6.** Boxplots representing the doubling time (in min) of each group of isolates in lysogenic broth (LB) (A) or in artificial urinary medium (AUM) (B). iRCIs = initial recurrent cystitis isolates, rRCI = recurrent cystitis isolate associated with relapse(s), SCIs = sporadic cystitis isolates, K12 = *E. coli str.* K12. Each dot represents a mean doubling time for one isolate in the corresponding group. Red asterisks represent extreme phenotypes for a given group (outliers). Significant differences are indicated by black asterisks.

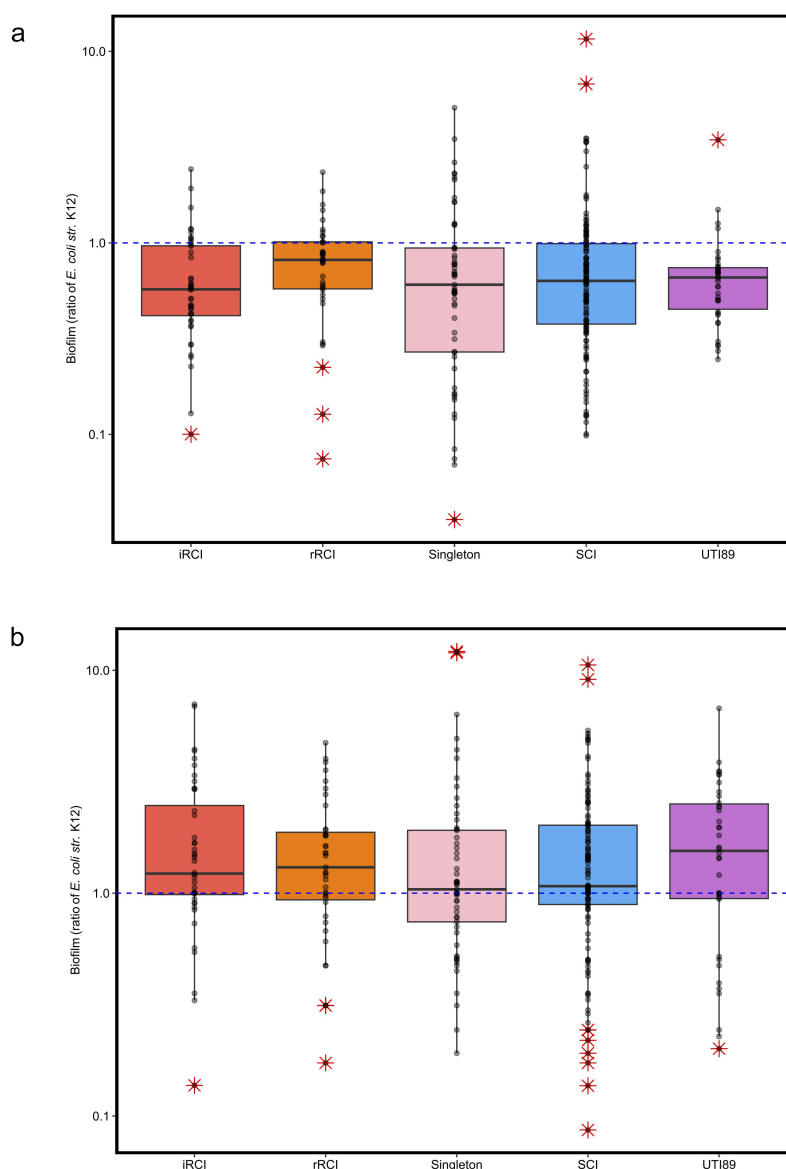
$G_{K12} = 23.2$  min) (Figure 6a). In AUM, all isolates grew significantly slower ( $p < 0.01$ ) than in LB ( $G_{iRCI} = 43.6$  min,  $G_{rRCI} = 43.6$  min,  $G_{SCI} = 46.2$  min and  $G_{Singleton} = 47.9$  min) (Figure 6b). No significant difference was observed between groups and *E. coli* UTI89 in AUM ( $G_{UTI89-AUM} = 43.3$  min), but *E. coli* K12 grew significantly slower than the other isolates ( $G_{K12-AUM} = 96.6$  min,  $p < 0.01$ ) (Figure 6b). We observed only one significant growth rate increase in RCI pairs (RCI2), which could not be linked to a particular genomic evolution (Fig S2).

Some isolates exhibited higher or lower growth rate than the rest of their group (represented as outliers on

Figure 6), with no association with specific genomic characteristics.

### **RCIs and SCIs exhibited similar biofilm formation capacity**

To evaluate whether biofilm production could be linked to recurrence, we studied the biofilm formation capacity of the same 72 isolates in LB and in AUM. All groups exhibited low levels of biofilm formation in LB ( $A_{590\text{ nm}} < 1$ ) and even lower in AUM ( $A_{590\text{ nm}} < 0.4$ ). No significant difference was



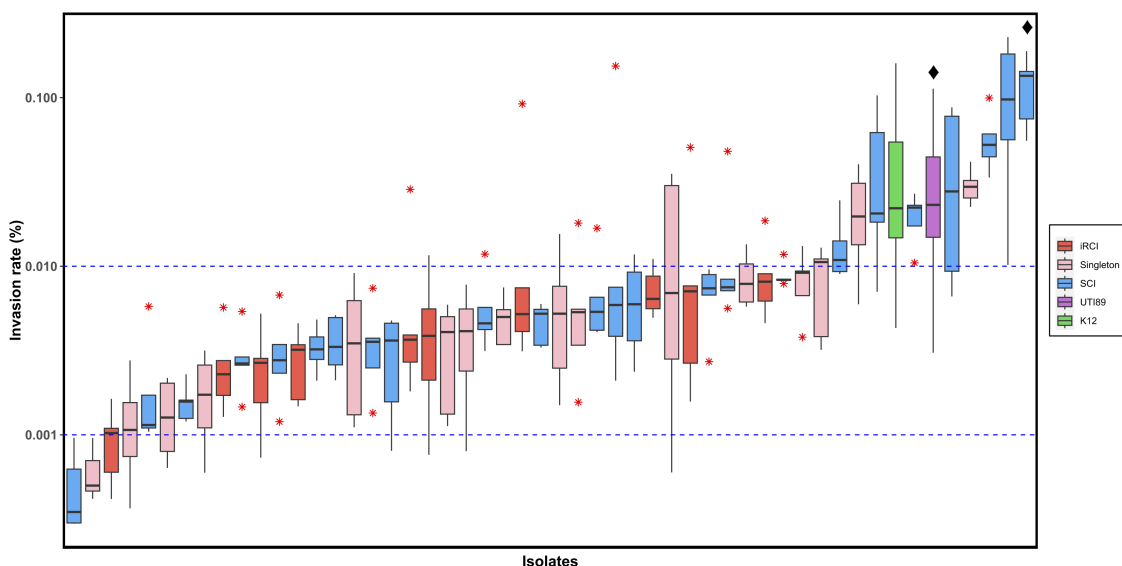
**Figure 7.** Biofilm formation in lysogenic broth (LB) (a) and in artificial urinary medium (AUM) (b) expressed as a ratio of *E. coli* str. K12 biofilm production. iRCIs = initial recurrent cystitis isolates, rRCI = recurrent cystitis isolate associated with relapse(s), SCIs = sporadic cystitis isolates. The blue dashed line represents the biofilm production level of *E. coli* str. K12. Each dot represents a mean ratio of biofilm production for one isolate in the corresponding group. Red asterisks represent extreme phenotypes for a given group (outliers).

observed between groups and the positive control *E. coli* K12, neither in LB nor in AUM (Figure 7a,b). No significant evolution in biofilm production was observed for RCI pairs, neither in LB nor in AUM (Fig S3).

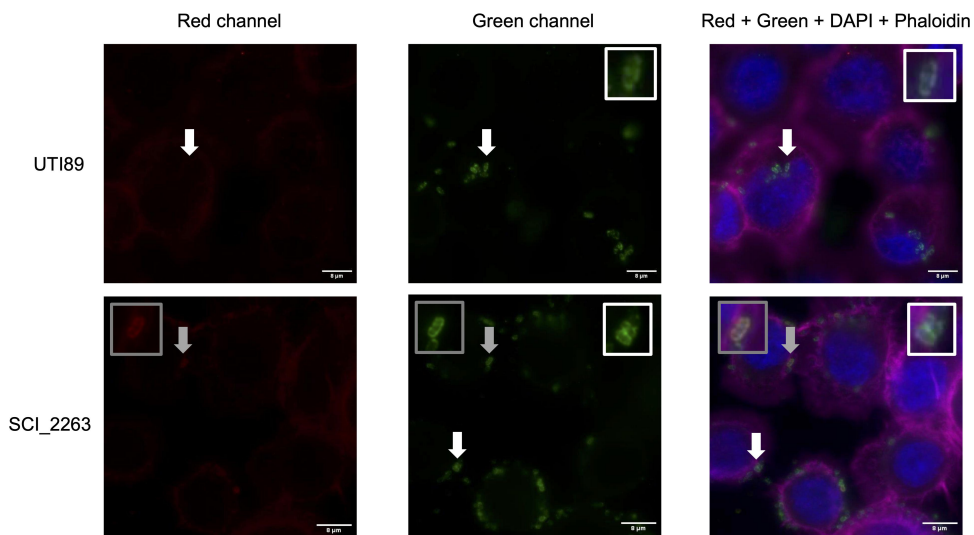
Of note, a missense mutation was identified in the PNAG biosynthesis associated poly-beta-1,6-N-acetyl-D-glucosamine synthase in rRCI\_2681 (from the RCI1 series [Table 2]) but did not lead to a significant difference in biofilm formation capacity compared to its corresponding iRCI\_2630.

### ***Invasion of bladder epithelial cells was similar between iRCI and SCI***

To determine whether bladder epithelial cell (BEC) invasion was a common feature among RCIs, invasion capacity of 60 isolates (10 iRCIs and their corresponding last rRCIs, 24 SCIs and 16 singletons) was evaluated by gentamicin protection assay (Figure 8). The invasion rates were low (between 0.0005 and 0.1194%), but 8 isolates (16.0%, 2 singletons and 6 SCIs) exhibited similar or greater invasion capacity than the positive control strain



**Figure 8.** Invasion capacity of UPEC isolates in HTB-9 bladder epithelial cells after gentamicin protection assay. Boxplots represent the individual invasion rates of 10 iRCIs (initial recurrent cystitis isolates, red), 16 singletons (pink), 24 SCIs (sporadic cystitis isolates, blue), *E. coli* str. K12 (K12, green) and positive control strain UTI89 (purple). Red asterisks represent outliers. Diamond indicates isolate SCI\_2263 for which microscopy results are presented in fig 9. Blue dashed lines represent the invasion rate interval described by Schwartz *et al.* (2011).



**Figure 9.** Observation of intracellular bacteria (SCI\_2263) in HTB-9 bladder epithelial cells (BEC) by fluorescence microscopy. Extracellular bacteria are labelled in red (staining without cell membrane permeabilization). Total (i.e. intra + extracellular) bacteria are labelled in green (staining with cell permeabilization). BEC nuclei are labelled with DAPI (in blue) and actin with phalloidin (in purple). Extracellular bacteria are labelled in red or yellow (green + red) and intracellular bacteria are labelled strictly in green. White arrows and insets are focusing on strictly intracellular bacteria while grey arrows and insets on strictly extracellular bacteria.

UTI89 (Figure 8). Surprisingly, *E. coli* str. K12 invasion rate was similar to that of *E. coli* UTI89. There was no correlation between invasion capacity and recurrence, though the invasion rate significantly increased between 2 iRCIs and their rRCIs (RCI5b and RCI9, Fig S4). The presence of intracellular bacteria was confirmed for 4 strains by fluorescence microscopy (Figure 9).

## Discussion

Our study firstly aimed to identify *E. coli* genomic and/or phenotypic characteristics associated with UTI recurrence, by comparing 24 RCIs to 24 phylogenetically paired SCIs. This pairing aimed to reduce bias due to genetic background and was based on CH typing,

a powerful and cost-effective tool to predict MLST results and investigate UPECs [10,11].

Previous studies could not identify a specific combination of *s* virulence factor determinants (VFDs) as predictive marker of rUTI [17,18]. However, Ejrnaes *et al.* compared 78 UPEC isolates causing persistence or relapse of UTI to 77 isolates followed by cure or reinfection and reported that rUTI-associated isolates exhibit a higher aggregate virulence score, associated with 12 significantly more prevalent VFDs [17]. In contrast, our study, based on less but paired isolates, showed no difference in virulence score or VFD prevalence between groups. We also studied VFD polymorphism and observed various levels of polymorphism without association with recurrence. The most polymorphic VFDs encoded adhesins and iron acquisition systems, which are key virulence factors for UTI establishment [12]. Such non-convergent VFD polymorphism could be due to host-specific adaptation: Zdziarski *et al.* showed that urinary tract inoculation of a single strain of *E. coli* in six volunteers led to unique adaptation patterns in each individual [40].

When comparing whole genomes, we found that SCIs' were smaller than RCIs,' mostly due to MGE abundance. There was no RCI-associated gene and only 2 SCI-associated genes, which derived from a prophage integrated between an operon encoding an envelope stress response system involved in UPEC virulence regulation (*cpx*) and an iron efflux system (*fiuF*) [41,42]. However, no effect of prophage presence was observed on growth or biofilm formation under iron limitation. Moreover, plasmid curation events were observed in intra-patient relapse series, confirming a result from Thänert *et al.* [19] profiling within-host UPEC adaptation in the gastrointestinal and urinary tracts. Indeed, they observed that rUTI-associated isolates exhibited lower MGE richness, suggesting that MGE loss constitutes a common UPEC adaptation to the urinary tract. By contrast, SCIs maintained their fitness in multiple habitats partly thanks to the conservation of their MGEs [19].

Interestingly, except for one patient, rRCIs were not associated with the emergence of antibiotic resistance. However, mutations occurring preferentially in genes involved in diverse metabolic pathways were observed, in line with the observations of Nielsen *et al.* [18]. Metabolism genes represent approximately 30% of the coding *E. coli* genome [43], and their representation among the mutated genes of our study (37.9%) was not significantly higher than expected ( $p = 0.2$ ). Interestingly, genes encoding ABC transporters were the second most represented functional category of mutated genes in our RCI series. They were mainly involved in transport of

important compounds for UPEC growth in urine: ionic compounds (including iron, nickel and phosphate), carbohydrates and peptides [44]. This should be taken carefully since the analysis was performed on the only few mutated proteins (58 out of 160) for which a functional pathway could be successfully attributed, but some of the mutated transporters were previously identified as critical factors for UPEC adaptation, such as *envZ* (response to osmolarity and pH variations) [45] or *fecC* (iron transport) [46]. Such mutations, combined with those affecting metabolism genes, might constitute a specific adaptation to human urine composition.

No pathoadaptive mutations [47] were identified in this study, probably because of the great inter-individual diversity of human urine composition [48] leading to host-specific isolate adaptation rather than convergent evolution. This host-shaped adaptation probably occurs early, as suggested by a significant decrease of RCI evolution rates over the course of relapses. Moreover, plasmid loss, as well as overtime-conserved SNPs – which possibly represent adaptive SNPs – mostly occurred between the iRCI and the first rRCI. Taken together, these results suggest that RCIs rapidly adapt to the urinary tract followed by a slower phase of micro-adaptation. Moreover, the main limitation of our approach was the lack of information concerning the UTI history of the patients before inclusion. Very early adaptation events may have occurred before iRCI sampling, leading to a possible underestimation of this trend.

AUM was designed to reproduce urine physico-chemical conditions [34] for growing various uropathogens, including *E. coli* [49], with better reproducibility and stability than pooled human urine [34,50]. The mean doubling time observed in AUM for our large collection of clinical isolates ( $G = 45.8$  min) is higher than that described for UPECs in pooled human urine (between 36.3 and 40.5 min) [50,51]. This can be explained by the greater complexity of human urine, which contains more usable substrates (carbohydrates, amino acids ...) than the AUM (only small amounts of peptone and yeast extract) [34,48].

No difference in mean doubling time was observed between RCIs and SCIs, but notable within-group diversity that cannot be explained by our genomic observations. However, if isolates specifically adapt to conditions encountered in their host, the use of AUM might not be able to reveal the consequent physiological changes. Moreover, such *in vitro* experiments do not mimic the immune pressure existing *in vivo*. Further studies should focus on characterizing RCI growth in their host's urine to determine whether genomic events identified in RCIs constitute a specific adaptation to the patient's urine composition.



Biofilm promotes persistent infection in patients [52], but its association with recurrent UPEC UTI is controversial [17,53]. Most studies described biofilm formation in minimal media, as nutrient stress appears to promote *E. coli* biofilm induction [54]. Soto *et al.* observed a positive correlation between biofilm production and recurrence in a minimal medium supplemented with LB [53], but Ejrnaes *et al.* found no correlation using AB medium [17]. The more intense starvation induced by AUM led to a reduced biomass production and might have prevented biofilm formation. This reduced biomass production also limited the comparison. It could be interesting to perform biofilm formation assays in dynamic models that better mimic the urine flux, since the combination of mechanical and biological stresses encountered in the urinary tract was shown to trigger biofilm formation [54,55].

BEC invasion by UPECs is considered as one of the main mechanisms explaining relapses [14–16]. This mechanism has been extensively studied in murine UTI models using the strain UTI89 [56–58], but data concerning clinical isolates were lacking. We found that most RCI and SCI invasion rates were in the range of published data [58]. Massive internalization was rare, with no significant correlation between relapse and invasion capacity, suggesting that this factor alone cannot explain the onset of recurrence. We used a monolayered cell infection experiment, which allows to study the formation of intracellular bacterial communities (IBCs) [35]. However, IBCs are only part of a complex mechanism, including quiescent intracellular reservoirs (QIRs) formation in transitional epithelium. More complex models, such as organoids or *in vivo* experiments, should thus be used to investigate the role of QIRs in recurrence. Moreover, UPEC ability to durably colonize the gastrointestinal tract and vagina may also promote persistence [19,59,60]. Future studies should therefore consider exploring these reservoirs, through longitudinal analysis of vaginal, intestinal, and bladder microbiomes of patients with rUTI.

In conclusion, rather than a common adaptation mechanism to the urinary tract, our results suggest a diversity of mechanisms leading to host-specific adaptation and thus to recurrence. Further studies exploring host–pathogen relationships and QIR formation with organoid models in rUTI pathogenesis should be performed to help translate these results into innovative treatments.

## Acknowledgment

We are indebted to the volunteers. We thank all the collaborators and colleagues who helped in the study. We are

grateful to the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing help and computing resources.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The VITALE study [NCT02292160] was funded by the French Ministry of Health (Programme Hospitalier de Recherche Clinique) We are grateful to Normandy Region and Rouen University for funding in part the cursus of Nicolas Vautrin.

## Author contributions

All authors have read and approved the final work.

Conceptualization: FC, MPC, KA, SD, NV; Methodology: NV, KA, DR, SL, ML, FG, MG; Validation: MPC, SD, KA, DR, NV, FG; Formal analysis: NV, ML, DR; Investigation: NV, MG, AF; Resources: MPC, SL, FG, DR; Data curation: NV, ML, MG; Writing – original draft preparation: NV, ML, MPC, SD, DR, FG; Writing – review and editing: NV, ML, MPC, SD, DR, FG; Visualization: NV, ML, DR, MG; Supervision: MPC, FC, SD, KA; Project administration: MPC, SD, FC, KA; Funding acquisition: MPC, FC

## Data availability statement

The data that support the findings of this study are openly available in figshare (<https://figshare.com>) at <https://doi.org/10.6084/m9.figshare.25651701>, reference number 25,651,701. The raw sequencing data and assembled genomes used in this study have been deposited in the Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra>) under NCBI BioProject accession no. PRJNA900024. SRA and assembly accession numbers are listed in Table S2 in the supplemental material.

## Preprint publication

A preprint version of this manuscript was submitted to bioRxiv and is available at the following link <https://doi.org/10.1101/2023.11.08.566351>.

## ORCID

Nicolas Vautrin  <http://orcid.org/0000-0003-1271-5436>  
 Sandrine Dahyot  <http://orcid.org/0000-0002-4230-1886>  
 François Gravey  <http://orcid.org/0000-0001-8672-8286>  
 David Ribet  <http://orcid.org/0000-0003-1175-6471>  
 Kévin Alexandre  <http://orcid.org/0000-0001-7773-5695>  
 Martine Pestel-Caron  <http://orcid.org/0000-0002-9888-7601>

## References

- [1] Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Disease-A-Month*. 2003;49(2):53–70. doi: [10.1067/mda.2003.7](https://doi.org/10.1067/mda.2003.7)
- [2] Foxman B. The epidemiology of urinary tract infection. *Nat Rev Urol*. 2010;7(12):653–660. doi: [10.1038/nrurol.2010.190](https://doi.org/10.1038/nrurol.2010.190)
- [3] Glover M, Moreira CG, Sperandio V, et al. Recurrent urinary tract infections in healthy and nonpregnant women. *Urol Sci*. 2014;25(1):1–8. doi: [10.1016/j.urols.2013.11.007](https://doi.org/10.1016/j.urols.2013.11.007)
- [4] Medina M, Castillo-Pino E. An introduction to the epidemiology and burden of urinary tract infections. *Ther Adv Urol*. 2019;11:175628721983217. doi: [10.1177/1756287219832172](https://doi.org/10.1177/1756287219832172)
- [5] Ejrnaes K, Sandvang D, Lundgren B, et al. Pulsed-field gel electrophoresis typing of *Escherichia coli* strains from samples collected before and after pivmecillinam or placebo treatment of uncomplicated community-acquired urinary tract infection in women. *J Clin Microbiol*. 2006;44(5):1776–1781. doi: [10.1128/JCM.44.5.1776-1781.2006](https://doi.org/10.1128/JCM.44.5.1776-1781.2006)
- [6] Vosti KL. A prospective, longitudinal study of the behavior of serologically classified isolates of *Escherichia coli* in women with recurrent urinary tract infections. *J Infect*. 2007;55(1):8–18. doi: [10.1016/j.jinf.2007.01.006](https://doi.org/10.1016/j.jinf.2007.01.006)
- [7] Beerepoot MAJ, Den Heijer CDJ, Penders J, et al. Predictive value of *Escherichia coli* susceptibility in strains causing asymptomatic bacteriuria for women with recurrent symptomatic urinary tract infections receiving prophylaxis. *Clin Microbiol Infect*. 2012;18(4):E84–E90. doi: [10.1111/j.1469-0691.2012.03773.x](https://doi.org/10.1111/j.1469-0691.2012.03773.x)
- [8] Skjøt-Rasmussen L, Hammerum AM, Jakobsen L, et al. Persisting clones of *Escherichia coli* isolates from recurrent urinary tract infection in men and women. *J Med Microbiol*. 2011;60(4):550–554. doi: [10.1099/jmm.0.026963-0](https://doi.org/10.1099/jmm.0.026963-0)
- [9] Nemoy LL, Kotetishvili M, Tigno J, et al. Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum beta-lactamase-producing *Escherichia coli* isolates. *J Clin Microbiol*. 2005;43(4):1776–1781. doi: [10.1128/JCM.43.4.1776-1781.2005](https://doi.org/10.1128/JCM.43.4.1776-1781.2005)
- [10] Weissman SJ, Johnson JR, Tchesnokova V, et al. High-resolution two-locus clonal typing of extraintestinal pathogenic *Escherichia coli*. *Appl Environ Microbiol*. 2012;78(5):1353–1360. doi: [10.1128/AEM.06663-11](https://doi.org/10.1128/AEM.06663-11)
- [11] Vautrin N, Alexandre K, Pestel-Caron M, et al. Contribution of antibiotic susceptibility testing and CH typing compared to next-generation sequencing for the diagnosis of recurrent urinary tract infections due to genetically identical *Escherichia coli* isolates: a prospective cohort study of cystitis in women. In: Taneja N, editor. *Microbiol spectr*. Vol. 11. 2023. p. e02785–22. doi: [10.1128/spectrum.02785-22](https://doi.org/10.1128/spectrum.02785-22)
- [12] Lüthje P, Brauner A. Virulence factors of uropathogenic *E. coli* and their interaction with the Host. *Adv Microb Physiol [Internet]*. Elsevier; 2014 [cited 2023 Aug 4];337–372. doi: [10.1016/bs.ampbs.2014.08.006](https://doi.org/10.1016/bs.ampbs.2014.08.006). <https://linkinghub.elsevier.com/retrieve/pii/S0065291114000071>
- [13] Murray BO, Flores C, Williams C, et al. Recurrent urinary tract infection: a mystery in search of better model systems. *Front Cell Infect Microbiol*. 2021;11:691210. doi: [10.3389/fcimb.2021.691210](https://doi.org/10.3389/fcimb.2021.691210)
- [14] Liu S, Han X, Shi M, et al. Persistence of uropathogenic *Escherichia coli* in the bladders of female patients with sterile urine after antibiotic therapies. *J Huazhong Univ Sci Technol [Med Sci]*. 2016;36(5):710–715. doi: [10.1007/s11596-016-1649-9](https://doi.org/10.1007/s11596-016-1649-9)
- [15] Robino L, Scavone P, Araujo L, et al. Detection of intracellular bacterial communities in a child with *Escherichia coli* recurrent urinary tract infections. *Pathog Disease*. 2013;68(3):78–81. doi: [10.1111/2049-632X.12047](https://doi.org/10.1111/2049-632X.12047)
- [16] Robino L, Scavone P, Araujo L, et al. Intracellular bacteria in the pathogenesis of *Escherichia coli* urinary tract infection in children. *Clin Infect Dis*. 2014;59(11):e158–e164. doi: [10.1093/cid/ciu634](https://doi.org/10.1093/cid/ciu634)
- [17] Ejrnæs K, Stegger M, Reisner A, et al. Characteristics of *Escherichia coli* causing persistence or relapse of urinary tract infections: phylogenetic groups, virulence factors and biofilm formation. *Virulence*. 2011;2(6):528–537. doi: [10.4161/viru.2.6.18189](https://doi.org/10.4161/viru.2.6.18189)
- [18] Nielsen KL, Stegger M, Kiil K, et al. *Escherichia coli* causing recurrent urinary tract infections: comparison to non-recurrent isolates and genomic adaptation in recurrent infections. *Microorganisms*. 2021;9(7):1416. doi: [10.3390/microorganisms9071416](https://doi.org/10.3390/microorganisms9071416)
- [19] Thänert R, Choi J, Reske KA, et al. Persisting uropathogenic *Escherichia coli* lineages show signatures of niche-specific within-host adaptation mediated by mobile genetic elements. *Cell Host Microbe*. 2022;30:1034–1047.e6. doi: [10.1016/j.chom.2022.04.008](https://doi.org/10.1016/j.chom.2022.04.008)
- [20] De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. In: Alkan C, editor. *Bioinformatics*. Vol. 39. 2023. p. btad311. doi: [10.1093/bioinformatics/btad311](https://doi.org/10.1093/bioinformatics/btad311)
- [21] Wick RR, Schultz MB, Zobel J, et al. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics*. 2015;31(20):3350–3352. doi: [10.1093/bioinformatics/btv383](https://doi.org/10.1093/bioinformatics/btv383)
- [22] Manni M, Berkeley MR, Seppely M, et al. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. In: Kelley J, editor. *Molecular biology and evolution*. Vol. 38. 2021. p. 4647–4654. doi: [10.1093/molbev/msab199](https://doi.org/10.1093/molbev/msab199)
- [23] Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–2069. doi: [10.1093/bioinformatics/btu153](https://doi.org/10.1093/bioinformatics/btu153)
- [24] Kumar S, Stecher G, Li M, et al. MEGA X: molecular evolutionary genetics analysis across computing platforms. In: Battistuzzi F, editor. *Molecular biology and evolution*. Vol. 35. 2018. p. 1547–1549. doi: [10.1093/molbev/msy096](https://doi.org/10.1093/molbev/msy096)
- [25] Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015;31(22):3691–3693. doi: [10.1093/bioinformatics/btv421](https://doi.org/10.1093/bioinformatics/btv421)
- [26] Jolley KA, Maiden MC. Bigsdb: scalable analysis of bacterial genome variation at the population level.

- BMC Bioinformatics. 2010;11(1):595. doi: 10.1186/1471-2105-11-595
- [27] Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics. 2018;19(1):307. doi: 10.1186/s12859-018-2336-6
- [28] Hadfield J, Croucher NJ, Goater RJ, et al. Phandango: an interactive viewer for bacterial population genomics. In: Kelso J, editor. Bioinformatics. Vol. 34. 2018. p. 292–293. doi: 10.1093/bioinformatics/btx610
- [29] Brynildsrud O, Bohlin J, Scheffer L, et al. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 2016;17(1):238. doi: 10.1186/s13059-016-1108-8
- [30] Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428(4):726–731. doi: 10.1016/j.jmb.2015.11.006
- [31] Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features in biological data. Protein Sci. 2022;31(1):47–53. doi: 10.1002/pro.4172
- [32] Grant JR, Enns E, Marinier E, et al. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Res. 2023;51(W1):W484–W492. doi: 10.1093/nar/gkad326
- [33] Starikova EV, Tikhonova PO, Prianichnikov NA, et al. Phigaro: high-throughput prophage sequence annotation. In: Valencia A, editor. Bioinformatics. Vol. 36. 2020. p. 3882–3884. doi: 10.1093/bioinformatics/btaa250
- [34] Brooks T, Keevil CW. A simple artificial urine for the growth of urinary pathogens. Lett Appl Microbiol. 1997;24(3):203–206. doi: 10.1046/j.1472-765X.1997.00378.x
- [35] Blango MG, Ott EM, Erman A, et al. Forced resurgence and targeting of intracellular uropathogenic *Escherichia coli* reservoirs. In: Beloin C, editor. PLOS ONE. Vol. 9. 2014. p. e93327. doi: 10.1371/journal.pone.0093327
- [36] Kühbacher A, Cossart P, Pizarro-Cerdá J. Internalization assays for *listeria monocytogenes*. Methods Mol Biol. 2014;1157:167–178. doi: 10.1007/978-1-4939-0703-8\_14
- [37] Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–675. doi: 10.1038/nmeth.2089
- [38] Johnson JR, Stell AL, Kaster N, et al. Novel molecular variants of allele I of the *Escherichia coli* P fimbrial adhesin gene *papG*. In: O'Brien A, editor. Infect Immun. Vol. 69. 2001. p. 2318–2327. doi: 10.1128/IAI.69.4.2318-2327.2001
- [39] Manning SD, Zhang L, Foxman B, et al. Prevalence of known P-Fimbrial G alleles in *Escherichia coli* and identification of a new adhesin class. Clin Diagn Lab Immunol. 2001;8(3):637–640. doi: 10.1128/CDLI.8.3.637-640.2001
- [40] Zdziarski J, Brzuszkiewicz E, Wullt B, et al. Host imprints on bacterial genomes—rapid, divergent evolution in individual patients. In: Guttman D, editor. PLOS Pathog. Vol. 6. 2010. p. e1001078. doi: 10.1371/journal.ppat.1001078
- [41] Debnath I, Norton JP, Barber AE, et al. The Cpx stress response system potentiates the fitness and virulence of uropathogenic *Escherichia coli*. Infect Immun. 2013;81(5):1450–1459. doi: 10.1128/IAI.01213-12
- [42] Grass G, Otto M, Fricke B, et al. FieF (YiiP) from *Escherichia coli* mediates decreased cellular accumulation of iron and relieves iron stress. Arch Microbiol. 2005;183(1):9–18. doi: 10.1007/s00203-004-0739-4
- [43] Wagner A. Metabolic networks and their evolution. In: Soyer O, editor. Evolutionary systems biology [internet]. New York (NY): Springer New York; 2012 [cited 2023 Aug 29]. p. 29–52. doi: 10.1007/978-1-4614-3567-9\_2.
- [44] Reitzer L, Zimmern P. Rapid growth and metabolism of uropathogenic *Escherichia coli* in relation to urine composition. Clin Microbiol Rev. 2019;33(1):e00101–19. doi: 10.1128/CMR.00101-19
- [45] Schwan WR. Survival of uropathogenic *Escherichia coli* in the murine urinary tract is dependent on OmpR. Microbiology. 2009;155(6):1832–1839. doi: 10.1099/mic.0.026187-0
- [46] Frick-Cheng AE, Sintsova A, Smith SN, et al. Ferric citrate uptake is a virulence factor in uropathogenic *Escherichia coli*. In: Comstock L, editor. mBio. Vol. 13. 2022. p. e01035–22. doi: 10.1128/mbio.01035-22
- [47] Sokurenko E, Mulvey MA, Stapleton AE, et al. Pathoadaptive mutations in uropathogenic *Escherichia coli*. Microbiol Spectr. 2016;4(2). doi: 10.1128/microbiolspec.UTI-0020-2015
- [48] Bouatra S, Aziat F, Mandal R, et al. The human urine metabolome. In: Dzeja P, editor. PLoS ONE. Vol. 8. 2013. p. e73076. doi: 10.1371/journal.pone.0073076
- [49] Ipe DS, Horton E, Ulett GC. The basics of bacteriuria: strategies of microbes for persistence in urine. Front Cell Infect Microbiol [Internet]. 2016 [cited 2023 Aug 29];6. doi: 10.3389/fcimb.2016.00014. <http://journal.frontiersin.org/Article/10.3389/fcimb.2016.00014/abstract>
- [50] Hogins J, Fan E, Seyan Z, et al. Bacterial growth of uropathogenic *Escherichia coli* in pooled urine is much higher than predicted from the average growth in individual urine samples. In: Prokesh B, editor. Microbiol spectr. Vol. 10. 2022. p. e02016–22. doi: 10.1128/spectrum.02016-22
- [51] Forsyth VS, Armbruster CE, Smith SN, et al. Rapid growth of uropathogenic *Escherichia coli* during human urinary tract infection. In: Miller J, editor. mBio. Vol. 9. 2018. p. e00186–18. doi: 10.1128/mBio.00186-18
- [52] Donlan RM. Biofilm formation: a clinically relevant microbiological process. Clin Infect Dis. 2001;33(8):1387–1392. doi: 10.1086/322972
- [53] Soto SM, Smithson A, Horcajada JP, et al. Implication of biofilm formation in the persistence of urinary tract infection caused by uropathogenic *Escherichia coli*. Clin Microbiol Infect. 2006;12(10):1034–1036. doi: 10.1111/j.1469-0691.2006.01543.x
- [54] Chu EK, Kilic O, Cho H, et al. Self-induced mechanical stress can trigger biofilm formation in uropathogenic *Escherichia coli*. Nat Commun. 2018;9(1):4087. doi: 10.1038/s41467-018-06552-z
- [55] Eberly A, Floyd K, Beebout C, et al. Biofilm formation by uropathogenic *Escherichia coli* is favored under oxygen conditions that mimic the bladder

- environment. *IJMS*. 2017;18(10):2077. doi: [10.3390/ijms18102077](https://doi.org/10.3390/ijms18102077)
- [56] Mulvey MA, Schilling JD, Hultgren SJ. Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection. In: O'Brien A, editor. *Infect immun*. Vol. 69. 2001. p. 4572–4579. doi: [10.1128/IAI.69.7.4572-4579.2001](https://doi.org/10.1128/IAI.69.7.4572-4579.2001)
- [57] Eto DS, Sundsbak JL, Mulvey MA. Actin-gated intracellular growth and resurgence of uropathogenic *Escherichia coli*. *Cell Microbiol*. 2006;8(4):704–717. doi: [10.1111/j.1462-5822.2006.00691.x](https://doi.org/10.1111/j.1462-5822.2006.00691.x)
- [58] Schwartz DJ, Chen SL, Hultgren SJ, et al. Population Dynamics and niche distribution of uropathogenic *Escherichia coli* during acute and chronic urinary tract infection. In: Payne S, editor. *Infect immun*. Vol. 79. 2011. p. 4250–4259. doi: [10.1128/IAI.05339-11](https://doi.org/10.1128/IAI.05339-11)
- [59] Salazar AM, Neugent ML, De Nisco NJ, et al. Gut-bladder axis enters the stage: implication for recurrent urinary tract infections. *Cell Host Microbe*. 2022;30:1066–1069. doi: [10.1016/j.chom.2022.07.008](https://doi.org/10.1016/j.chom.2022.07.008)
- [60] Worby CJ, Schreiber HL, Straub TJ, et al. Longitudinal multi-omics analyses link gut microbiome dysbiosis with recurrent urinary tract infections in women. *Nat Microbiol*. 2022;7(5):630–639. doi: [10.1038/s41564-022-01107-x](https://doi.org/10.1038/s41564-022-01107-x)