Supplementary Materials for

Whole-genome sequencing of uropathogenic *Escherichia coli* reveals long evolutionary history of diversity and virulence

Yancy Lo[1†], Lixin Zhang[2] , Betsy Foxman[3,*], Sebastian Zöllner[1,4,*]

[1] Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48109

[2] Department of Epidemiology and Biostatistics, Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, 48824

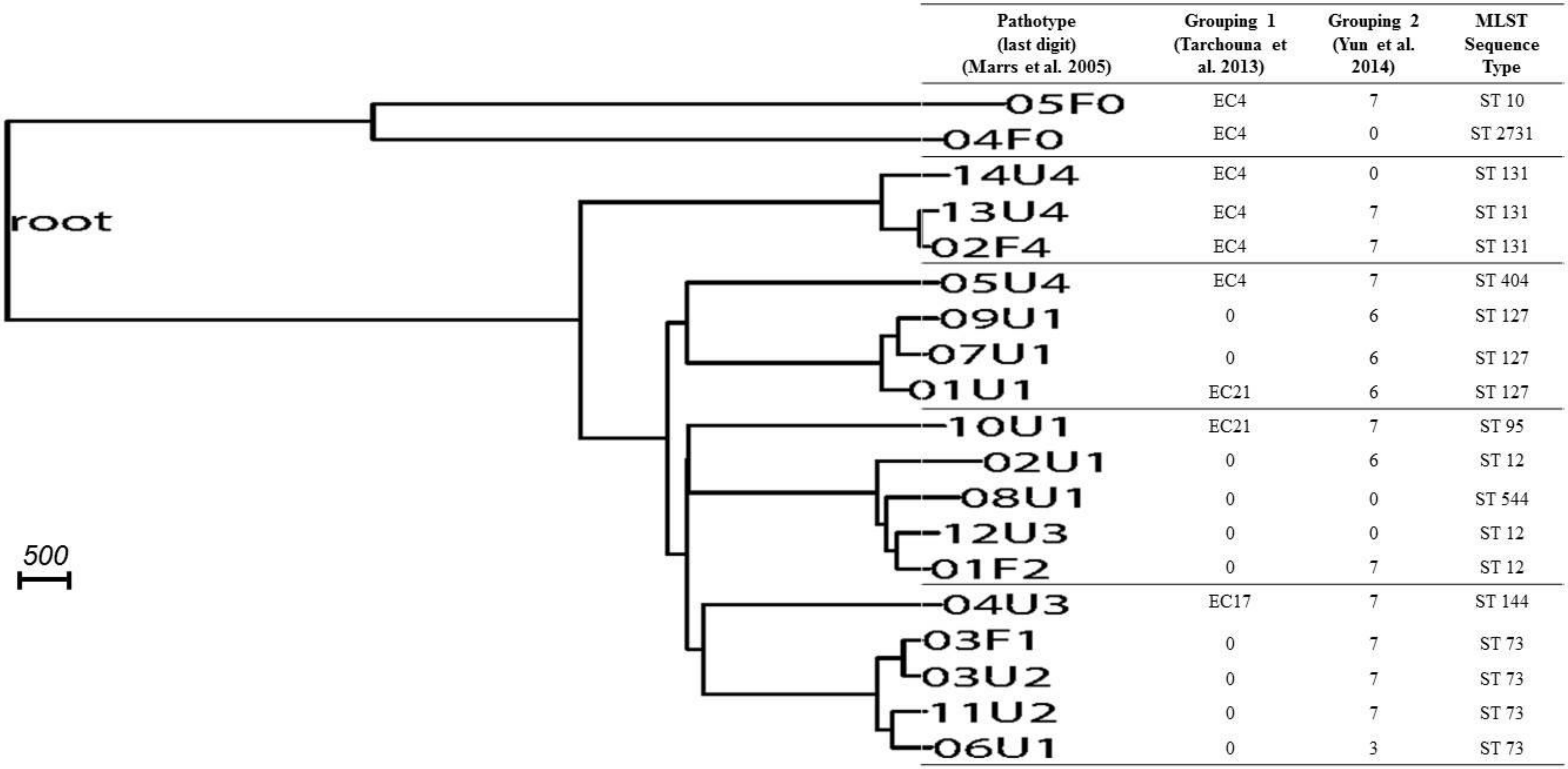
[3] Department of Epidemiology, University of Michigan, Ann Arbor, MI, 48109, USA

[4] Department of Psychiatry, University of Michigan, Ann Arbor, MI, 48109, USA

This PDF file includes:

Supplementary Figure 1

Supplementary Table 1



Pathotype (last digit) (Marrs et al. 2005)	Grouping 1 (Tarchouna et al. 2013)	Grouping 2 (Yun et al. 2014)	MLST Sequence Type
05F0	EC4	7	ST 10
04F0	EC4	0	ST 2731
14U4	EC4	0	ST 131
13U4	EC4	7	ST 131
02F4	EC4	7	ST 131
05U4	EC4	7	ST 404
09U1	0	6	ST 127
07U1	0	6	ST 127
01U1	EC21	6	ST 127
10U1	EC21	7	ST 95
02U1	0	6	ST 12
08U1	0	0	ST 544
12U3	0	0	ST 12
01F2	0	7	ST 12
04U3	EC17	7	ST 144
03F1	0	7	ST 73
03U2	0	7	ST 73
11U2	0	7	ST 73
06U1	0	3	ST 73

Supplementary Figure 1. Classifications based on presence and absence of several virulence factors, described in Marrs et al. (2005), Tarchouna et al. (2013) and Yun et al. (2014). We also listed the multilocus sequence typing (MLST) of our sample based on presence and absence of seven housekeeping genes. For Groupings 1 and 2, type "0" refers to the combinations of virulence factors in our sampled strains that were not found in the original studies. MLST types correspond well to the whole-genome phylogeny, while none of the groupings based on uropathogenic virulence factors were consistent with the whole-genome clustering. Three pathotype 4 strains formed a significant subclade, but the remaining pathotype 4 strain had a split with this subclade over 10 million generations ago. Similarly, using Grouping 2, three type 6 strains formed a significant subclade, while the remaining type 6 strain showed a deep split with this subclade over 7 million generations ago.

Supplementary Table 1. List of virulence factors by sampled strain. We annotated our sampled strains with the UTI89 reference. Based on the virulence factors compiled from Spurbeck et al. 2012 and Yun et al. 2014 and conditioning on their presence in UTI89, we listed the presence and absence of virulence factors in our sample. *E coli* ID identifies each strain from the original collection of Marrs et al. 2005. Sample code of each strain refers to: Individual host ID (first two digits), Urine/ Fecal sample (U/F in the third digit), and pathotype (last digit).

E coli ID	223F64	Y32458-61	438F4	105F2	Y32458-11	69F2	389F2	438F62	129F2	Y34511-011	54F2	54F66	223F2	347F1	361F2	302F3	302F61	30F2
Sample Code	05F0	01F2	02U1	12U3	07U1	09U1	02F4	13U4	14U4	04U3	04F0	05U4	08U1	06U1	03U2	03F1	11U2	10U1
MLST Type	ST 10	ST 12	ST 12	ST 12	ST 127	ST 127	ST 131	ST 131	ST 131	ST 144	ST 2731	ST 404	ST 544	ST 73	ST 73	ST 73	ST 73	ST 95
aer	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
afa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
aufA	-	+	+	+	+	+	-	-	-	+	-	+	+	+	+	+	+	+
chuA	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+
cnf1	-	+	+	-	+	+	-	-	-	-	-	-	+	+	+	+	+	+
feoB	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
fimH	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
fyuA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
hlyA	-	+	+	+	+	+	-	-	-	+	-	-	+	+	+	+	+	+
iroN	-	+	+	+	-	-	-	-	-	+	+	-	+	+	+	+	+	+
iutA	+	+	-	+	-	-	+	+	+	+	+	+	-	-	+	+	+	-
kpsMT	+	+	+	+	+	+	+	+	-	+	-	+	-	+	+	+	+	+
ompT	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
papA	-	+	+	-	-	-	-	-	-	+	-	-	-	-	+	+	-	+
papC	-	+	+	+	+	+	-	-	-	+	-	-	+	+	+	+	+	+
papG	-	-	+	-	+	+	-	-	-	-	-	-	+	+	+	+	-	+
sfaS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
vat	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
yadN	-	+	+	+	+	+	-	-	-	+	-	+	+	+	+	+	+	+
yejE	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
yfcV	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+
ygiL	-	+	-	+	-	-	+	+	+	+	-	-	+	+	+	+	+	+
yniA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+