Draft genome sequences of semi-constitutive rdar biofilm forming *Escherichia coli* isolates

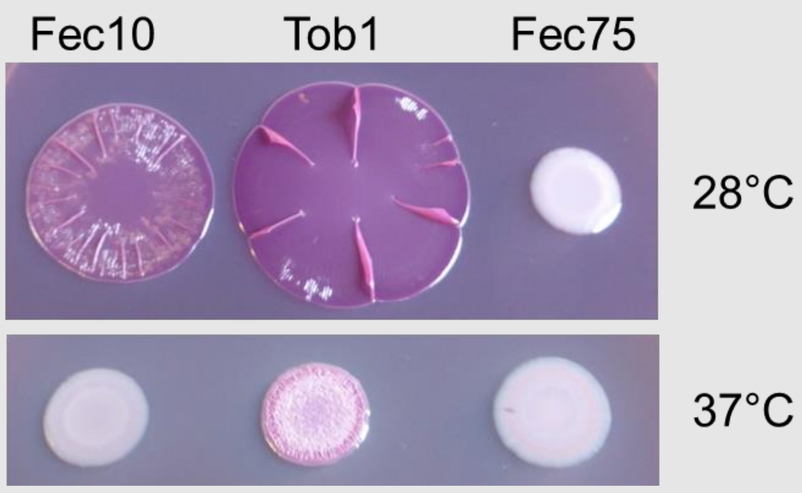
Annika Cimdins[1,2], Petra Lüthje[1,3], Roger Simm[4], Fengyang Li[1], Kaisa Thorell[1], Irfan Ahmad[1,5], Åsa Sjöling[1], Annelie Brauner[1], and Ute Römling[1]

1 Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
2 present address: Institute of Hygiene, University Hospital (UKM) and University of Muenster, Germany
3 present address: Department of Laboratory Medicine, Division of Clinical Microbiology, Karolinska Institutet and Karolinska University Hospital Huddinge, Stockholm, Sweden
4 Norwegian Veterinary Institute, Section of Bacteriology, Oslo, Norway
5 present address: Department of Molecular Biology, Umeå University, Sweden

Introduction

Escherichia coli is a well-studied model organism concerning various aspects of bacterial physiology and behavior. Biofilm formation is defined as multicellular microbial communities surrounded by a self-produced matrix of extracellular polymeric and non-polymeric substances adherent to each other, to interfaces and to biotic or abiotic surfaces [1]. Strains of *E. coli* were proven to exhibit diverse biofilm phenotypes. A well-studied biofilm type characterized by the formation of an extracellular matrix consisting mainly of amyloid curli fimbriae and the exopolysaccharide cellulose is the red, dry, and rough (rdar) morphotype. Whereas *E. coli* K-12 model strains express at most a rudimentary rdar morphotype at temperature below 30° C only [2, 3], clinical isolates have been shown to frequently form the rdar morphotype semi-constitutively also at body temperature [3-6]. In this study, the genomes of eight semi-constitutive rdar biofilm forming *E. coli* strains were sequenced with Illumina Miseq or PacBio sequencing.

Rdar morphotype formation



Examples of differential rdar morphotype expression of commensal fecal isolates [5]:
Fec10 ($rdar_{28°C}$/$saw_{37°C}$), Tob1 ($rdar_{28°C}$/$rdar_{37°C}$), Fec75 ($saw_{28°C}$/$saw_{37°C}$)
Strains were grown on LB without salt agar plates supplemented with 40 µg/ml Congo Red and 20 µg/ml Coomassie Brilliant Blue for 72h. Binding of the dye to curli fibers and cellulose results in staining of the colony. Non biofilm forming strains are smooth and white (saw).

Experimental procedures

Genomic DNA from Tob1 and ECOR 31 was sequenced with the PacBio RS II system (Pacific Biosciences; NGI Uppsala, SciLifeLab (Science For Life Laboratory), Uppsala). The assembly was done on SMRT portal version 2.3, using HGAP3 with default settings. The other strains were sequenced using an Illumina MiSeq v3 platform with read length up to 2x300bp (NGI Stockholm, SciLifeLab (Science For Life Laboratory), Solna). *De novo* assembly was performed using SPAdes (http://bioinf.spbau.ru/spades). Calculation of coverage has been done according to Lander and Waterman [8]. Annotation was performed using the RAST - Rapid Annotation using Subsystem Technology (version 2.0) server (http://rast.nmpdr.org/rast.cgi) and the NCBI annotation pipeline PGAP.

References

1. Costerton et al., Annu Rev Microbiol, 1995. 49: p. 711-45.
2. Hammar et al., Mol Microbiol, 1995. 18(4): p. 661-70.
3. Römling et al., Mol Microbiol, 1998. 28(2): p. 249-64.
4. Bian et al., 2000. 181(2): p. 602-12.
5. Bokranz et al., J Med Microbiol, 2005. 54(Pt 12): p. 1171-82.
6. Kai-Larsen et al., PLoS Pathog, 2010. 6(7): p. e1001010.
7. Ochman, H. and Selander, J Bacteriol, 1984. 157(2): p. 690-3.
8. Lander and Watermann, Genomics 1998 2, p.231-239.
9. Cimdins et al., Genome Announc. 2017 Jan 26;5(4).
10. Clermont et al., Environ Microbiol Rep, 2013 Feb;5(1):p. 58-65.

Strains used in the study

Strain	Isolation source
Tob1	Commensal human fecal isolate, Germany
Fec67	Commensal human fecal isolate, Germany
Fec101	Commensal human fecal isolate, Germany
ECOR31	Commensal leopard fecal isolate, USA
No. 12	Human pyelonephritis isolate, Slovakia
B-11870	Human urosepsis isolate, Sweden
80//6	Human UTI isolate, Estonia
B-8638	Human urosepsis isolate, Sweden

Three of the strains represent minimally passaged strains isolated from human feces [5]. Four additional minimally passaged uropathogenic/urosepsis strains originated from urine and human blood samples, respectively, obtained in Sweden, Slovakia, and Estonia ([6],[9]). Strain ECOR31 from leopard feces was obtained from the *E. coli* reference strain collection [7] and included as an historical strain.

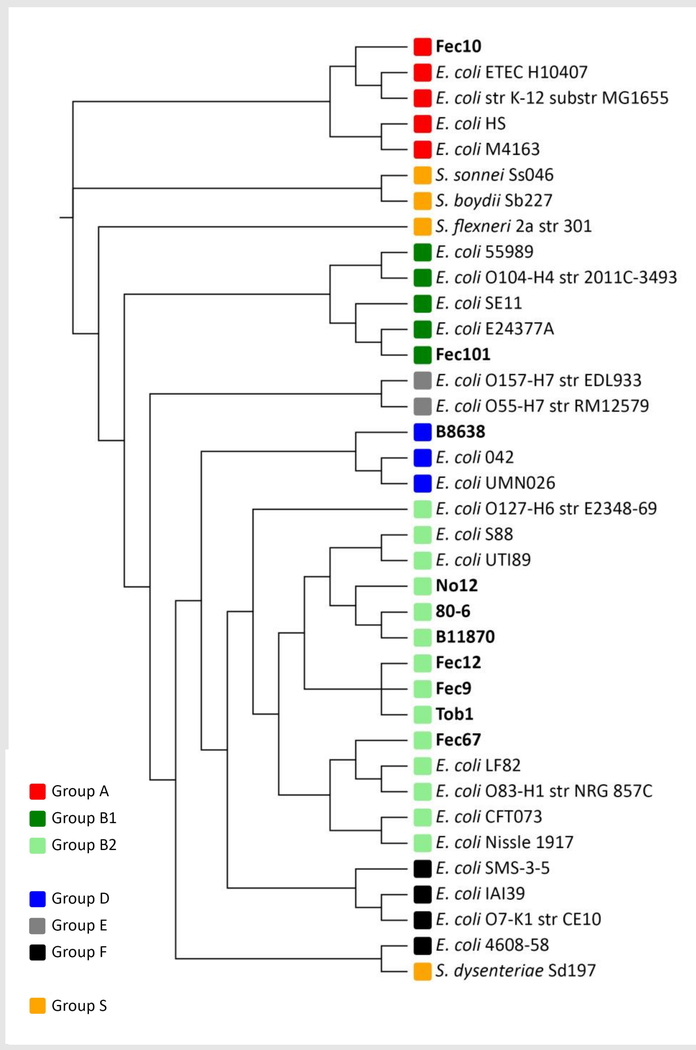
Sequence features

Strain	Size [Mbp][a]	G+C [%][a]	Genes (total/coding)[b]	tRNAs[b]
Tob1	5.19	50.9	5,147/4,882	90
Fec67	5.20	50.6	5,115/4,918	79
Fec101	4.97	50.7	4,942/4,751	79
ECOR31	5.44	50.6	5,328/5,095	90
No. 12	5.10	50.5	5,009/4,807	78
B-11870	5.65	50.5	5,188/4,970	77
80//6	4.98	50.6	4,886/4,697	79
B-8638	5.21	50.7	5,143/4,917	77

a= RAST server, b= NCBI Prokaryotic Genome Annotation Pipeline

PacBio sequencing gave rise to 4 contigs for Tob1 (one plasmid) and 6 contigs for ECOR31 (three plasmids). The Illumina sequencing followed by SPAdes assembly resulted in scaffold numbers from 90 – 140. All sequences have been published and deposited at DDBJ/ENA/GenBank [9].

Phylogenetic distribution compared to reference *E. coli* strains



Core genomes were compared using MAFFT (http://mafft.cbrc.jp/alignment/software/) and the tree was computed by FastTree (http://meta.microbesonline.org/fasttree/). Analysis of the phylo-groups of the isolates was done by *in silico* testing for the presence of reference genes according to Clermont [10]. Exemplary strains for the different *E. coli* clades and pathovars as well as *Shigella* have been included for comparison.

Summary and Conclusion

The genomes of eight semi-constitutive rdar biofilm forming *E. coli* strains, representing both commensal and uropathogenic strains from different phylogroups were sequenced with Illumina Miseq or PacBio sequencing and analyzed. As published genomes lack coupling of the sequencing to a distinct biofilm phenotype, the sequencing will contribute to shed light on the variability of the *E. coli* genome, in particular on the regulatory mechanisms of rdar biofilm formation.

AC receives funding from the German Research Foundation (DFG), CI 239/1-1 and CI 239/2-1. The authors would like to thank the Bioinformatics Infrastructure for Life Sciences (BILS) and the National Genomics Infrastructure (NGI), Science for Life Laboratory, Uppsala Genome Center and UPPMAX for support. Work performed at NGI/Uppsala Genome Center NGI and BILS is funded by the Swedish Research Council.

Karolinska Institutet
Ute Römling
Department of Microbiology, Tumor and Cellbiology
17177 Stockholm
Sweden
mail: ute.romling@ki.se
phone: 0046-8-524 87319