Cell Reports, Volume *14*

Supplemental Information

Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic *E. coli* in Necrotizing Enterocolitis and Mortality in Preterm Infants

Doyle V. Ward, Matthias Scholz, Moreno Zolfo, Diana H. Taft, Kurt R. Schibler, Adrian Tett, Nicola Segata, and Ardythe L. Morrow

Figure S1

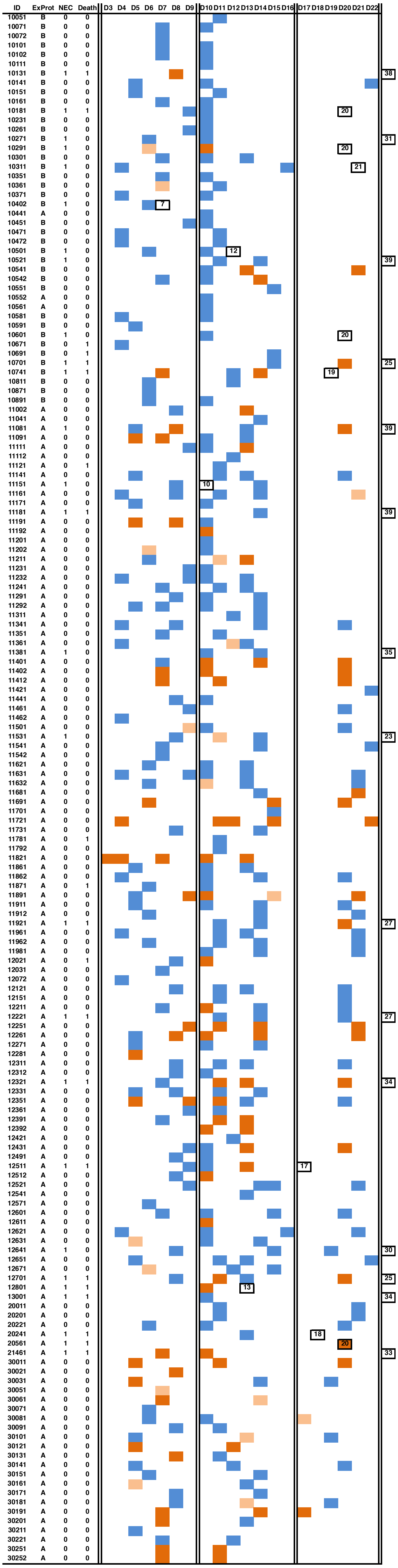
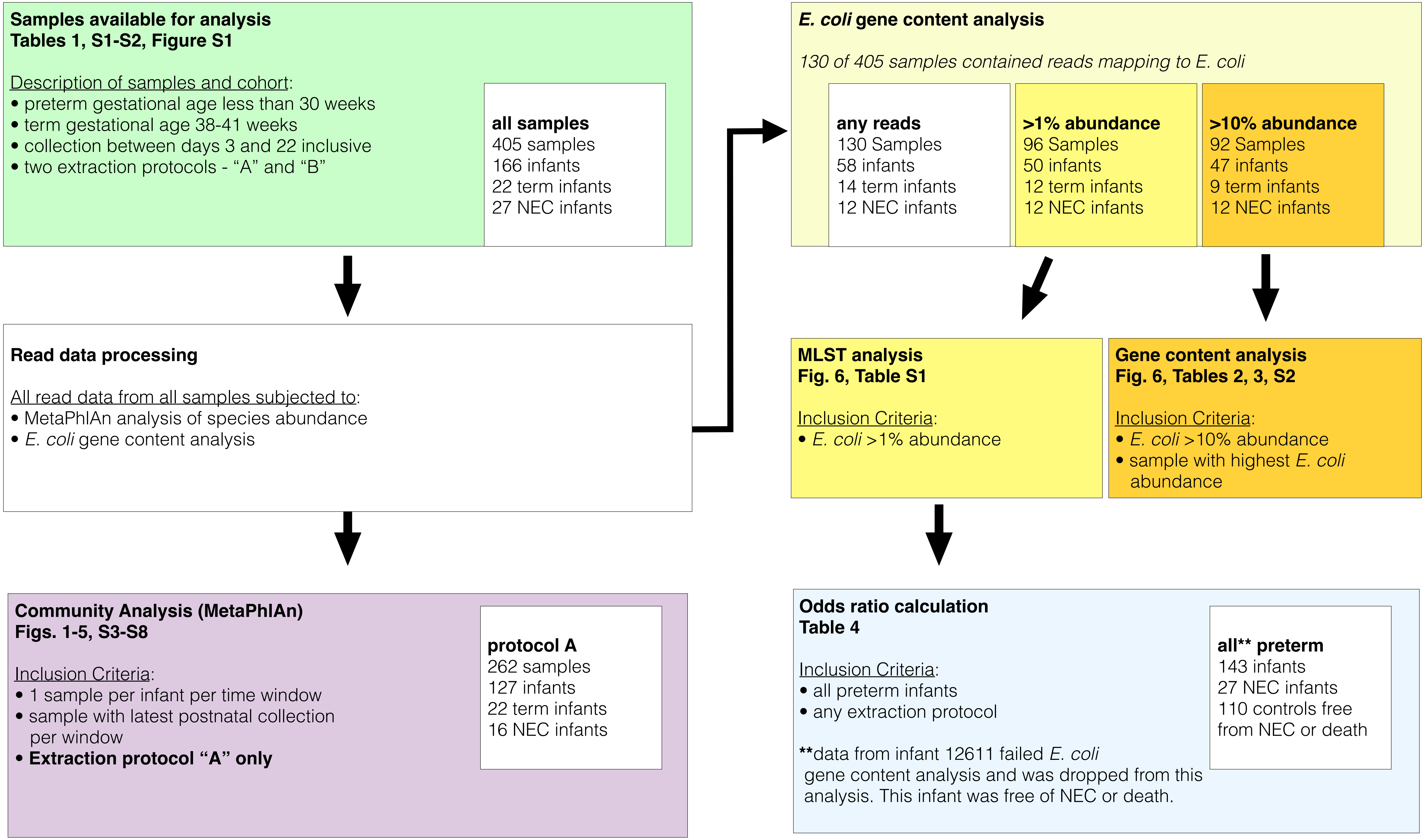

ID ExProt NEC Death
D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 D14 D15 D16 D17 D18 D19 D20 D21 D22
10051 B 0 0
10071 B 0 0
10072 B 0 0
10101 B 0 0
10102 B 0 0
10111 B 0 0
10131 B 1 1
10141 B 0 0
10151 B 0 0
10161 B 0 0
10181 B 1 1
10231 B 0 0
10261 B 0 0
10271 B 1 0
10291 B 1 0
10301 B 0 0
10311 B 1 0
10351 B 0 0
10361 B 0 0
10371 B 0 0
10402 B 1 0
10441 A 0 0
10451 B 0 0
10471 B 0 0
10472 B 0 0
10501 B 1 0
10521 B 1 0
10541 B 0 0
10542 B 0 0
10551 B 0 0
10552 A 0 0
10561 A 0 0
10581 B 0 0
10591 B 0 0
10601 B 1 0
10671 B 0 1
10691 B 0 1
10701 B 1 1
10741 B 1 1
10811 B 0 0
10871 B 0 0
10891 B 0 0
11002 A 0 0
11041 A 0 0
11081 A 1 0
11091 A 0 0
11111 A 0 0
11112 A 0 0
11121 A 0 1
11141 A 0 0
11151 A 1 0
11161 A 0 0
11171 A 0 0
11181 A 1 1
11191 A 0 0
11192 A 0 0
11201 A 0 0
11202 A 0 0
11211 A 0 0
11231 A 0 0
11232 A 0 0
11241 A 0 0
11291 A 0 0
11292 A 0 0
11311 A 0 0
11341 A 0 0
11351 A 0 0
11361 A 0 0
11381 A 1 0
11401 A 0 0
11402 A 0 0
11412 A 0 0
11421 A 0 0
11441 A 0 0
11461 A 0 0
11462 A 0 0
11501 A 0 0
11531 A 1 0
11541 A 0 0
11542 A 0 0
11621 A 0 0
11631 A 0 0
11632 A 0 0
11681 A 0 0
11691 A 0 0
11701 A 0 0
11721 A 0 0
11731 A 0 0
11781 A 0 1
11792 A 0 0
11821 A 0 0
11861 A 0 0
11862 A 0 0
11871 A 0 1
11891 A 0 0
11911 A 0 0
11912 A 0 0
11921 A 1 1
11961 A 0 0
11962 A 0 0
11981 A 0 0
12021 A 0 1
12031 A 0 0
12072 A 0 0
12121 A 0 0
12151 A 0 0
12211 A 0 0
12221 A 1 1
12251 A 0 0
12261 A 0 0
12271 A 0 0
12281 A 0 0
12311 A 0 0
12312 A 0 0
12321 A 1 1
12331 A 0 0
12351 A 0 0
12361 A 0 0
12391 A 0 0
12392 A 0 0
12421 A 0 0
12431 A 0 0
12491 A 0 0
12511 A 1 1
12512 A 0 0
12521 A 0 0
12541 A 0 0
12571 A 0 0
12601 A 0 0
12611 A 0 0
12621 A 0 0
12631 A 0 0
12641 A 1 0
12651 A 0 0
12671 A 0 0
12701 A 1 1
12801 A 1 1
13001 A 1 1
20011 A 0 0
20201 A 0 0
20221 A 0 0
20241 A 1 1
20561 A 1 1
21461 A 1 1
30011 A 0 0
30021 A 0 0
30031 A 0 0
30051 A 0 0
30061 A 0 0
30071 A 0 0
30081 A 0 0
30091 A 0 0
30101 A 0 0
30121 A 0 0
30131 A 0 0
30141 A 0 0
30151 A 0 0
30161 A 0 0
30171 A 0 0
30181 A 0 0
30191 A 0 0
30201 A 0 0
30211 A 0 0
30221 A 0 0
30251 A 0 0
30252 A 0 0
38
20
31
20
21
7
12
39
20
25
19
39
10
39
35
23
27
27
34
17
30
25
13
34
18
20
33

Figure S2



Samples available for analysis
Tables 1, S1-S2, Figure S1

<u>Description of samples and cohort</u>:
- preterm gestational age less than 30 weeks
- term gestational age 38-41 weeks
- collection between days 3 and 22 inclusive
- two extraction protocols - “A” and “B”

all samples
405 samples
166 infants
22 term infants
27 NEC infants

Read data processing

<u>All read data from all samples subjected to</u>:
- MetaPhlAn analysis of species abundance
- *E. coli* gene content analysis

Community Analysis (MetaPhlAn)
Figs. 1-5, S3-S8

<u>Inclusion Criteria</u>:
- 1 sample per infant per time window
- sample with latest postnatal collection per window
- **Extraction protocol “A” only**

protocol A
262 samples
127 infants
22 term infants
16 NEC infants

***E. coli* gene content analysis**

130 of 405 samples contained reads mapping to E. coli

any reads	**>1% abundance**	**>10% abundance**
130 Samples	96 Samples	92 Samples
58 infants	50 infants	47 infants
14 term infants	12 term infants	9 term infants
12 NEC infants	12 NEC infants	12 NEC infants

MLST analysis
Fig. 6, Table S1

<u>Inclusion Criteria</u>:
- *E. coli* >1% abundance

Gene content analysis
Fig. 6, Tables 2, 3, S2

<u>Inclusion Criteria</u>:
- *E. coli* >10% abundance
- sample with highest *E. coli* abundance

Odds ratio calculation
Table 4

<u>Inclusion Criteria</u>:
- all preterm infants
- any extraction protocol

all preterm**
143 infants
27 NEC infants
110 controls free from NEC or death

**data from infant 12611 failed *E. coli* gene content analysis and was dropped from this analysis. This infant was free of NEC or death.

Figure S3

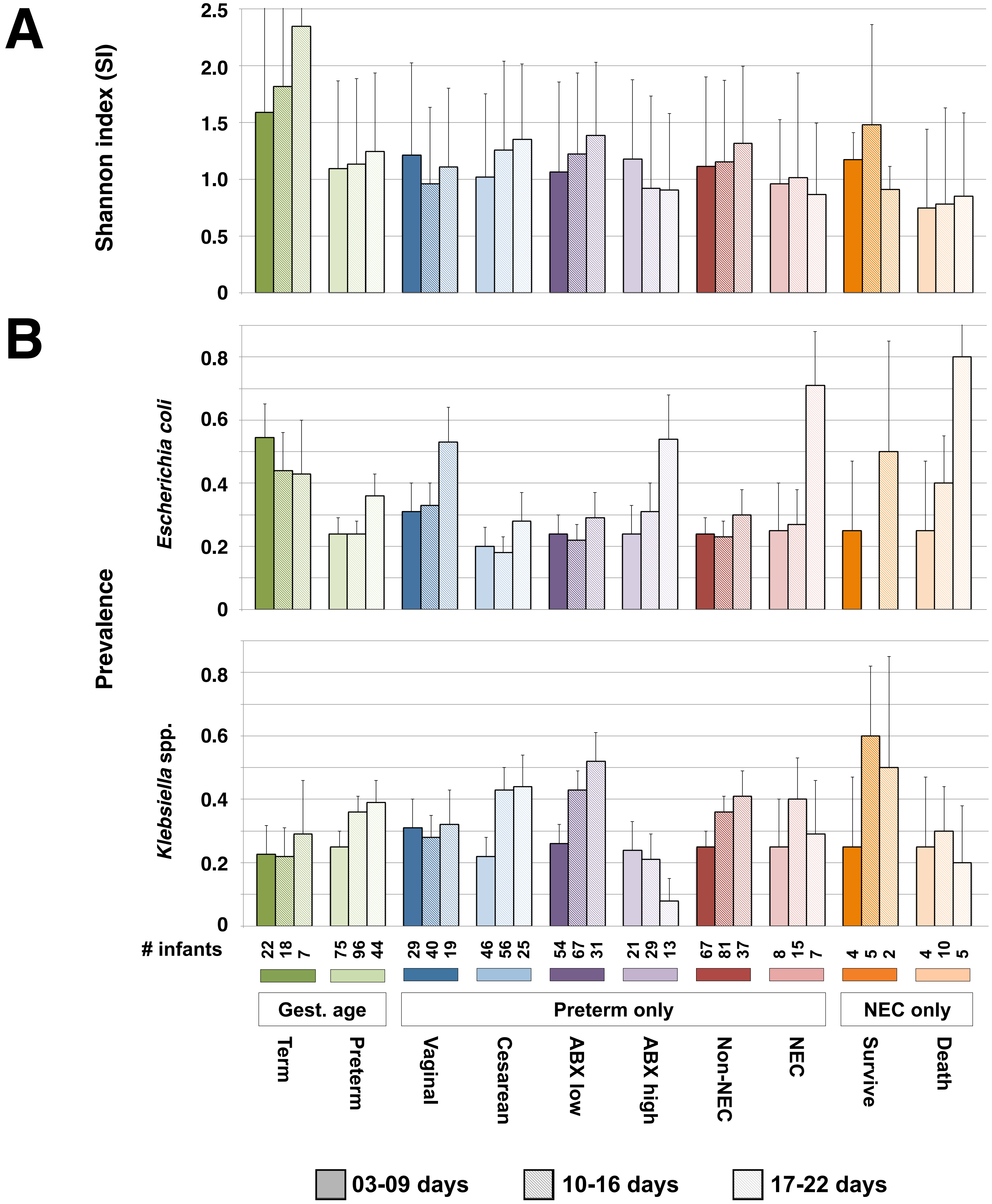

A
B
Shannon index (SI)
2.5
2.0
1.5
1.0
0.5
0
Prevalence
Escherichia coli
0.8
0.6
0.4
0.2
0
Klebsiella spp.
0.8
0.6
0.4
0.2
0
infants
22 18 7
75 96 44
29 40 19
46 56 25
54 67 31
21 29 13
67 81 37
8 15 7
4 5 2
4 10 5
Gest. age
Preterm only
NEC only
Term
Preterm
Vaginal
Cesarean
ABX low
ABX high
Non-NEC
NEC
Survive
Death
03-09 days
10-16 days
17-22 days

Figure S4

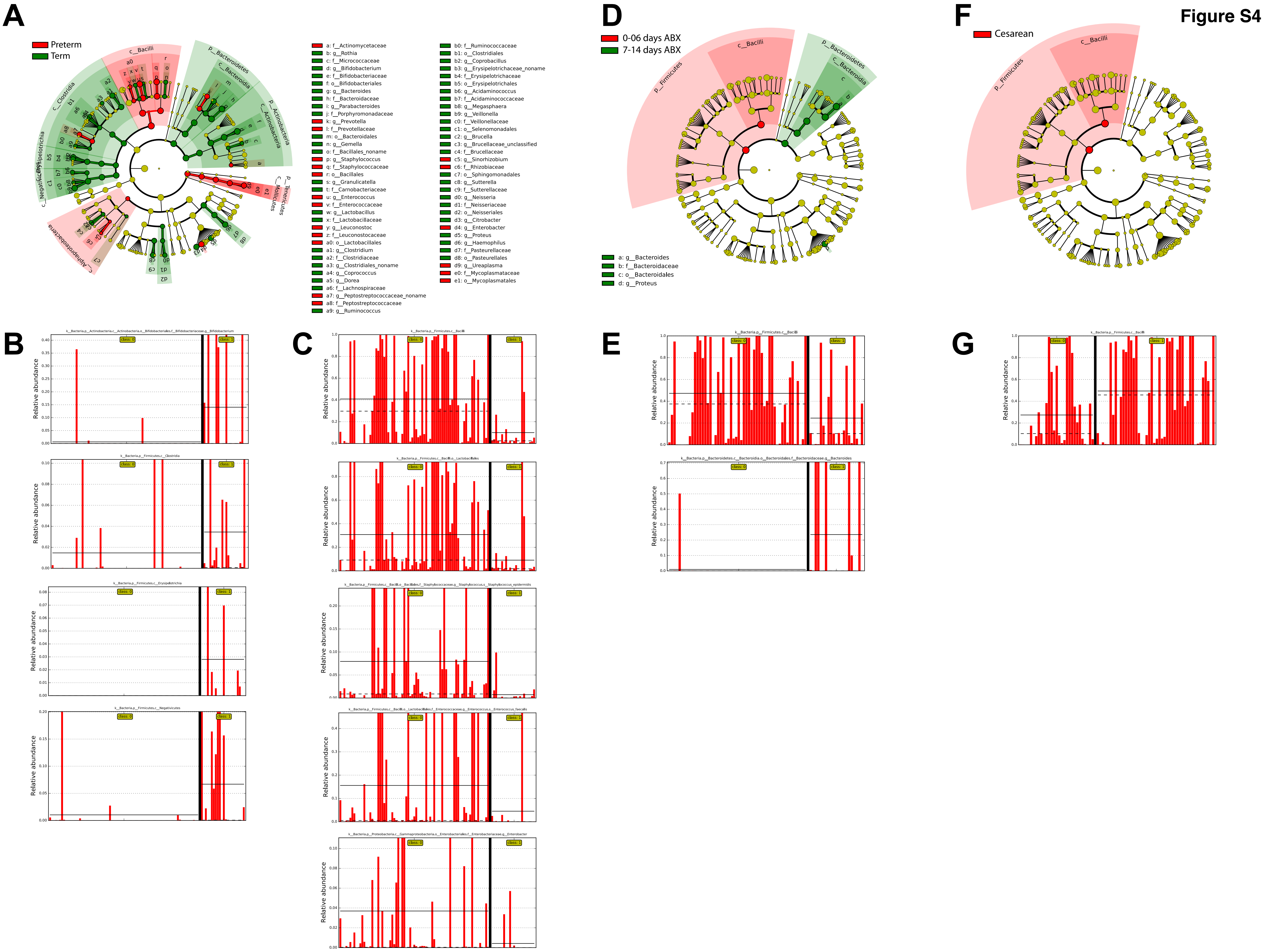

A
Preterm
Term
c__Bacilli
p__Bacteroidetes
c__Bacteroidia
p__Actinobacteria
c__Actinobacteria
p__Tenericutes
c__Mollicutes
c__Alphaproteobacteria
c__Negativicutes
c__Erysipelotrichia
c__Clostridia
a: f__Actinomycetaceae
b: g__Rothia
c: f__Micrococcaceae
d: g__Bifidobacterium
e: f__Bifidobacteriaceae
f: o__Bifidobacteriales
g: g__Bacteroides
h: f__Bacteroidaceae
i: g__Parabacteroides
j: f__Porphyromonadaceae
k: g__Prevotella
l: f__Prevotellaceae
m: o__Bacteroidales
n: g__Gemella
o: f__Bacillales_noname
p: g__Staphylococcus
q: f__Staphylococcaceae
r: o__Bacillales
s: g__Granulicatella
t: f__Carnobacteriaceae
u: g__Enterococcus
v: f__Enterococcaceae
w: g__Lactobacillus
x: f__Lactobacillaceae
y: g__Leuconostoc
z: f__Leuconostocaceae
a0: o__Lactobacillales
a1: g__Clostridium
a2: f__Clostridiaceae
a3: g__Clostridiales_noname
a4: g__Coprococcus
a5: g__Dorea
a6: f__Lachnospiraceae
a7: g__Peptostreptococcaceae_noname
a8: f__Peptostreptococcaceae
a9: g__Ruminococcus
b0: f__Ruminococcaceae
b1: o__Clostridiales
b2: g__Coprobacillus
b3: g__Erysipelotrichaceae_noname
b4: f__Erysipelotrichaceae
b5: o__Erysipelotrichales
b6: g__Acidaminococcus
b7: f__Acidaminococcaceae
b8: g__Megasphaera
b9: g__Veillonella
c0: f__Veillonellaceae
c1: o__Selenomonadales
c2: g__Brucella
c3: g__Brucellaceae_unclassified
c4: f__Brucellaceae
c5: g__Sinorhizobium
c6: f__Rhizobiaceae
c7: o__Sphingomonadales
c8: g__Sutterella
c9: f__Sutterellaceae
d0: g__Neisseria
d1: f__Neisseriaceae
d2: o__Neisseriales
d3: g__Citrobacter
d4: g__Enterobacter
d5: g__Proteus
d6: g__Haemophilus
d7: f__Pasteurellaceae
d8: o__Pasteurellales
d9: g__Ureaplasma
e0: f__Mycoplasmataceae
e1: o__Mycoplasmatales
D
0-06 days ABX
7-14 days ABX
p__Firmicutes
c__Bacilli
p__Bacteroidetes
c__Bacteroidia
a: g__Bacteroides
b: f__Bacteroidaceae
c: o__Bacteroidales
d: g__Proteus
F
Cesarean
p__Firmicutes
c__Bacilli
B
k__Bacteria.p__Actinobacteria.c__Actinobacteria.o__Bifidobacteriales.f__Bifidobacteriaceae.g__Bifidobacterium
k__Bacteria.p__Firmicutes.c__Clostridia
k__Bacteria.p__Firmicutes.c__Erysipelotrichia
k__Bacteria.p__Firmicutes.c__Negativicutes
C
k__Bacteria.p__Firmicutes.c__Bacilli
k__Bacteria.p__Firmicutes.c__Bacilli.o__Lactobacillales
k__Bacteria.p__Firmicutes.c__Bacilli.o__Bacillales.f__Staphylococcaceae.g__Staphylococcus.s__Staphylococcus_epidermidis
k__Bacteria.p__Firmicutes.c__Bacilli.o__Lactobacillales.f__Enterococcaceae.g__Enterococcus.s__Enterococcus_faecalis
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Enterobacter
E
k__Bacteria.p__Firmicutes.c__Bacilli
k__Bacteria.p__Bacteroidetes.c__Bacteroidia.o__Bacteroidales.f__Bacteroidaceae.g__Bacteroides
G
k__Bacteria.p__Firmicutes.c__Bacilli
class: 0
class: 1
Relative abundance

Figure S5

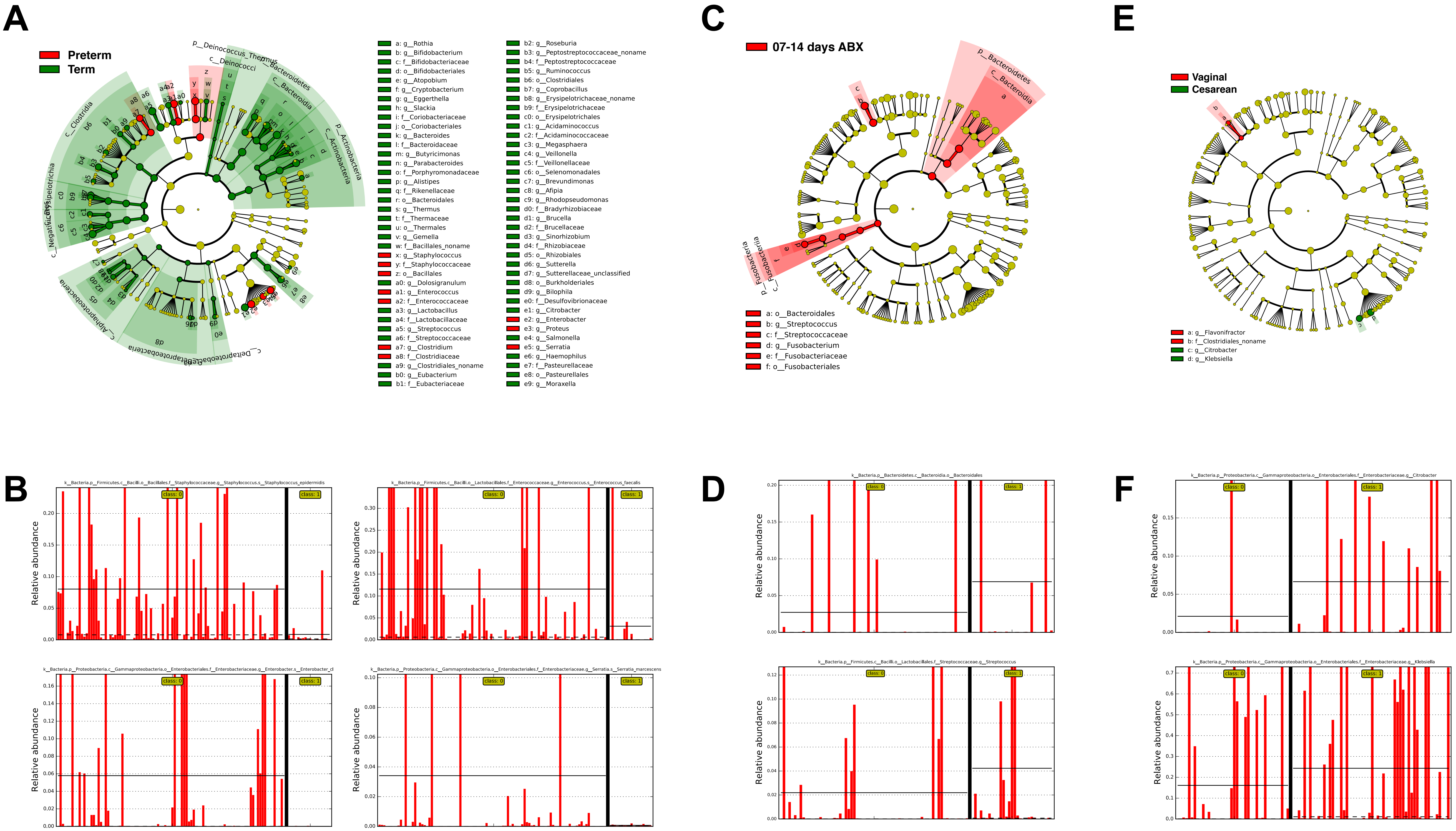

A
🟥 Preterm
🟩 Term
p__Deinococcus_Thermus
c__Deinococci
c__Bacteroidetes
c__Bacteroidia
p__Actinobacteria
c__Actinobacteria
c__Clostridia
c__Erysipelotrichia
c__Negativicutes
c__Alphaproteobacteria
c__Betaproteobacteria
c__Deltaproteobacteria
🟩 a: g__Rothia
🟩 b: g__Bifidobacterium
🟩 c: f__Bifidobacteriaceae
🟩 d: o__Bifidobacteriales
🟩 e: g__Atopobium
🟩 f: g__Cryptobacterium
🟩 g: g__Eggerthella
🟩 h: g__Slackia
🟩 i: f__Coriobacteriaceae
🟩 j: o__Coriobacteriales
🟩 k: g__Bacteroides
🟩 l: f__Bacteroidaceae
🟩 m: g__Butyricimonas
🟩 n: g__Parabacteroides
🟩 o: f__Porphyromonadaceae
🟩 p: g__Alistipes
🟩 q: f__Rikenellaceae
🟩 r: o__Bacteroidales
🟩 s: g__Thermus
🟩 t: f__Thermaceae
🟩 u: o__Thermales
🟩 v: g__Gemella
🟩 w: f__Bacillales_noname
🟥 x: g__Staphylococcus
🟥 y: f__Staphylococcaceae
🟥 z: o__Bacillales
🟩 a0: g__Dolosigranulum
🟥 a1: g__Enterococcus
🟥 a2: f__Enterococcaceae
🟩 a3: g__Lactobacillus
🟩 a4: f__Lactobacillaceae
🟩 a5: g__Streptococcus
🟩 a6: f__Streptococcaceae
🟥 a7: g__Clostridium
🟥 a8: f__Clostridiaceae
🟩 a9: g__Clostridiales_noname
🟩 b0: g__Eubacterium
🟩 b1: f__Eubacteriaceae
🟩 b2: g__Roseburia
🟩 b3: g__Peptostreptococcaceae_noname
🟩 b4: f__Peptostreptococcaceae
🟩 b5: g__Ruminococcus
🟩 b6: o__Clostridiales
🟩 b7: g__Coprobacillus
🟩 b8: g__Erysipelotrichaceae_noname
🟩 b9: f__Erysipelotrichaceae
🟩 c0: o__Erysipelotrichales
🟩 c1: g__Acidaminococcus
🟩 c2: f__Acidaminococcaceae
🟩 c3: g__Megasphaera
🟩 c4: g__Veillonella
🟩 c5: f__Veillonellaceae
🟩 c6: o__Selenomonadales
🟩 c7: g__Brevundimonas
🟩 c8: g__Afipia
🟩 c9: g__Rhodopseudomonas
🟩 d0: f__Bradyrhizobiaceae
🟩 d1: g__Brucella
🟩 d2: f__Brucellaceae
🟩 d3: g__Sinorhizobium
🟩 d4: f__Rhizobiaceae
🟩 d5: o__Rhizobiales
🟩 d6: g__Sutterella
🟩 d7: g__Sutterellaceae_unclassified
🟩 d8: o__Burkholderiales
🟩 d9: g__Bilophila
🟩 e0: f__Desulfovibrionaceae
🟩 e1: g__Citrobacter
🟥 e2: g__Enterobacter
🟥 e3: g__Proteus
🟩 e4: g__Salmonella
🟥 e5: g__Serratia
🟩 e6: g__Haemophilus
🟩 e7: f__Pasteurellaceae
🟩 e8: o__Pasteurellales
🟩 e9: g__Moraxella
B
k__Bacteria.p__Firmicutes.c__Bacilli.o__Bacillales.f__Staphylococcaceae.g__Staphylococcus.s__Staphylococcus_epidermidis
k__Bacteria.p__Firmicutes.c__Bacilli.o__Lactobacillales.f__Enterococcaceae.g__Enterococcus.s__Enterococcus_faecalis
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Enterobacter.s__Enterobacter_cl
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Serratia.s__Serratia_marcescens
Relative abundance
class: 0
class: 1
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.02
0.04
0.06
0.08
0.12
0.14
0.16
C
🟥 07-14 days ABX
p__Bacteroidetes
c__Bacteroidia
p__Fusobacteria
c__Fusobacteriia
🟥 a: o__Bacteroidales
🟥 b: g__Streptococcus
🟥 c: f__Streptococcaceae
🟥 d: g__Fusobacterium
🟥 e: f__Fusobacteriaceae
🟥 f: o__Fusobacteriales
D
k__Bacteria.p__Bacteroidetes.c__Bacteroidia.o__Bacteroidales
k__Bacteria.p__Firmicutes.c__Bacilli.o__Lactobacillales.f__Streptococcaceae.g__Streptococcus
Relative abundance
class: 0
class: 1
E
🟥 Vaginal
🟩 Cesarean
🟥 a: g__Flavonifractor
🟥 b: f__Clostridiales_noname
🟩 c: g__Citrobacter
🟩 d: g__Klebsiella
F
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Citrobacter
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Klebsiella
Relative abundance
class: 0
class: 1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

Figure S6

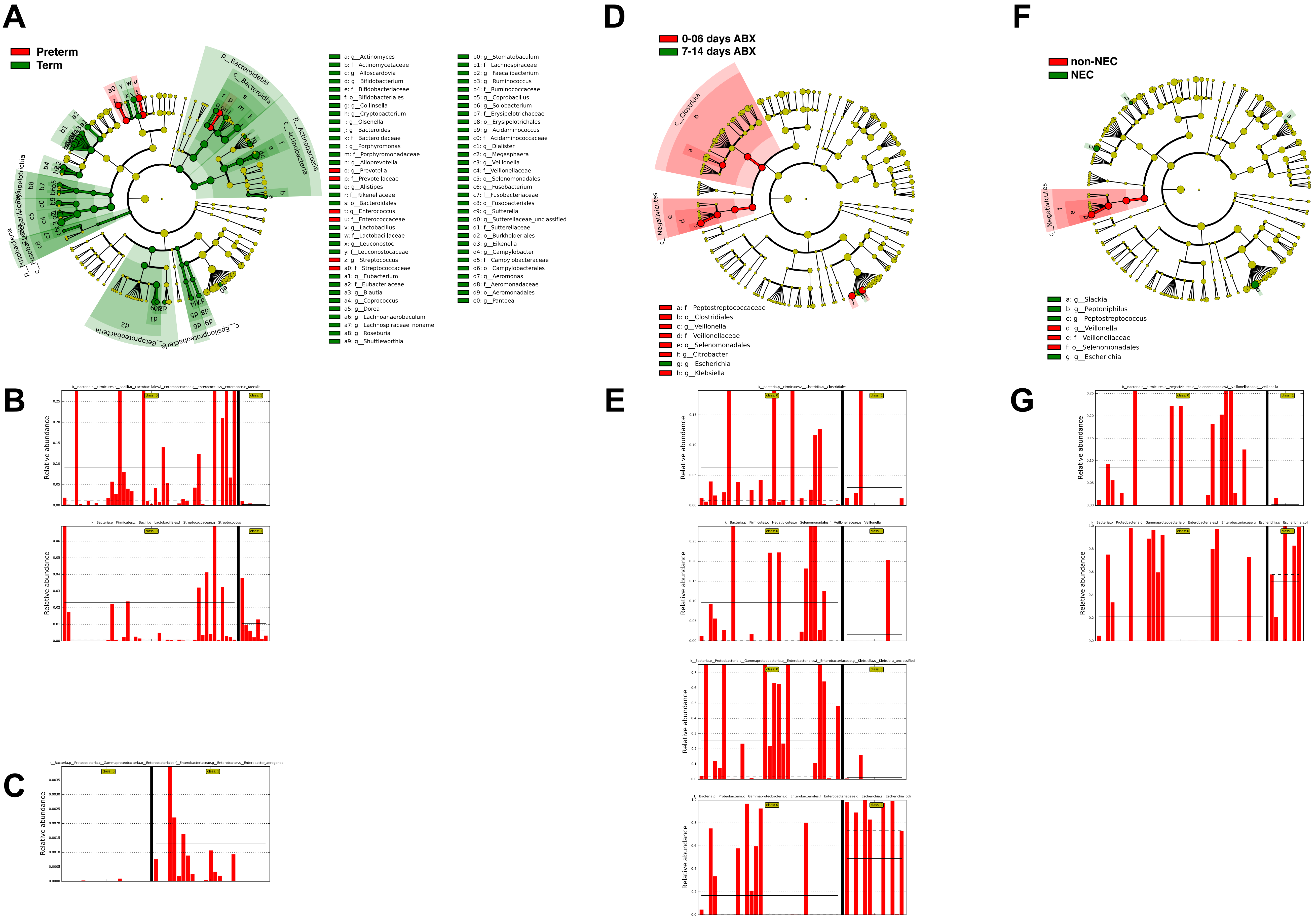
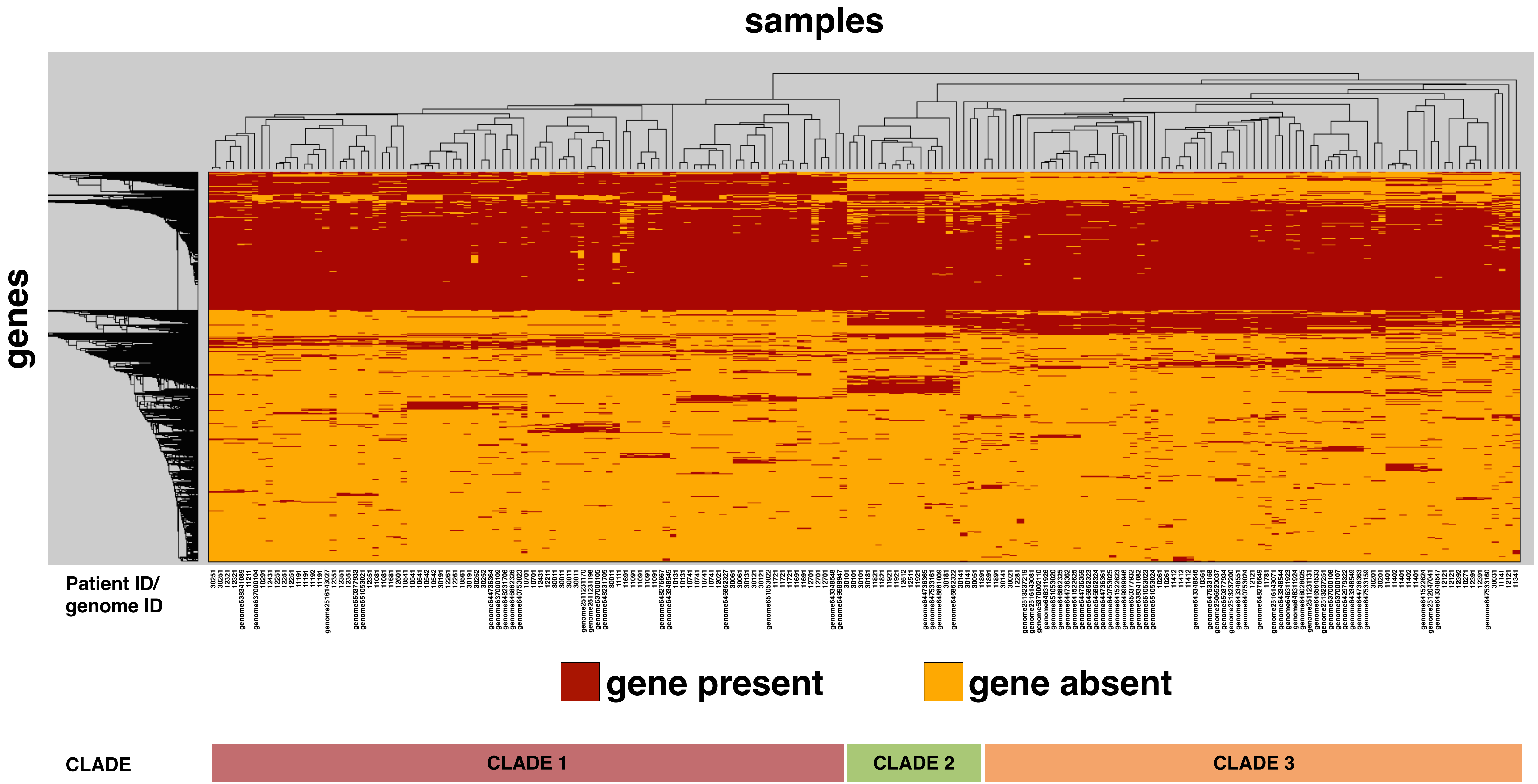

A
🟥 Preterm
🟩 Term
p__Bacteroidetes
c__Bacteroidia
p__Actinobacteria
c__Actinobacteria
c__Erysipelotrichia
c__Negativicutes
p__Fusobacteria
c__Fusobacteriia
c__Betaproteobacteria
c__Epsilonproteobacteria
🟩 a: g__Actinomyces
🟩 b: f__Actinomycetaceae
🟩 c: g__Alloscardovia
🟩 d: g__Bifidobacterium
🟩 e: f__Bifidobacteriaceae
🟩 f: o__Bifidobacteriales
🟩 g: g__Collinsella
🟩 h: g__Cryptobacterium
🟩 i: g__Olsenella
🟩 j: g__Bacteroides
🟩 k: f__Bacteroidaceae
🟩 l: g__Porphyromonas
🟩 m: f__Porphyromonadaceae
🟩 n: g__Alloprevotella
🟥 o: g__Prevotella
🟥 p: f__Prevotellaceae
🟩 q: g__Alistipes
🟩 r: f__Rikenellaceae
🟩 s: o__Bacteroidales
🟥 t: g__Enterococcus
🟥 u: f__Enterococcaceae
🟩 v: g__Lactobacillus
🟩 w: f__Lactobacillaceae
🟩 x: g__Leuconostoc
🟩 y: f__Leuconostocaceae
🟥 z: g__Streptococcus
🟥 a0: f__Streptococcaceae
🟩 a1: g__Eubacterium
🟩 a2: f__Eubacteriaceae
🟩 a3: g__Blautia
🟩 a4: g__Coprococcus
🟩 a5: g__Dorea
🟩 a6: g__Lachnoanaerobaculum
🟩 a7: g__Lachnospiraceae_noname
🟩 a8: g__Roseburia
🟩 a9: g__Shuttleworthia
🟩 b0: g__Stomatobaculum
🟩 b1: f__Lachnospiraceae
🟩 b2: g__Faecalibacterium
🟩 b3: g__Ruminococcus
🟩 b4: f__Ruminococcaceae
🟩 b5: g__Coprobacillus
🟩 b6: g__Solobacterium
🟩 b7: f__Erysipelotrichaceae
🟩 b8: o__Erysipelotrichales
🟩 b9: g__Acidaminococcus
🟩 c0: f__Acidaminococcaceae
🟩 c1: g__Dialister
🟩 c2: g__Megasphaera
🟩 c3: g__Veillonella
🟩 c4: f__Veillonellaceae
🟩 c5: o__Selenomonadales
🟩 c6: g__Fusobacterium
🟩 c7: f__Fusobacteriaceae
🟩 c8: o__Fusobacteriales
🟩 c9: g__Sutterella
🟩 d0: g__Sutterellaceae_unclassified
🟩 d1: f__Sutterellaceae
🟩 d2: o__Burkholderiales
🟩 d3: g__Eikenella
🟩 d4: g__Campylobacter
🟩 d5: f__Campylobacteraceae
🟩 d6: o__Campylobacterales
🟩 d7: g__Aeromonas
🟩 d8: f__Aeromonadaceae
🟩 d9: o__Aeromonadales
🟩 e0: g__Pantoea
B
k__Bacteria.p__Firmicutes.c__Bacilli.o__Lactobacillales.f__Enterococcaceae.g__Enterococcus.s__Enterococcus_faecalis
k__Bacteria.p__Firmicutes.c__Bacilli.o__Lactobacillales.f__Streptococcaceae.g__Streptococcus
Relative abundance
class: 0
class: 1
C
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Enterobacter.s__Enterobacter_aerogenes
Relative abundance
class: 0
class: 1
D
🟥 0-06 days ABX
🟩 7-14 days ABX
c__Clostridia
c__Negativicutes
🟥 a: f__Peptostreptococcaceae
🟥 b: o__Clostridiales
🟥 c: g__Veillonella
🟥 d: f__Veillonellaceae
🟥 e: o__Selenomonadales
🟥 f: g__Citrobacter
🟩 g: g__Escherichia
🟥 h: g__Klebsiella
E
k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales
k__Bacteria.p__Firmicutes.c__Negativicutes.o__Selenomonadales.f__Veillonellaceae.g__Veillonella
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Klebsiella.s__Klebsiella_unclassified
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Escherichia.s__Escherichia_coli
Relative abundance
class: 0
class: 1
F
🟥 non-NEC
🟩 NEC
c__Negativicutes
🟩 a: g__Slackia
🟩 b: g__Peptoniphilus
🟩 c: g__Peptostreptococcus
🟥 d: g__Veillonella
🟥 e: f__Veillonellaceae
🟥 f: o__Selenomonadales
🟩 g: g__Escherichia
G
k__Bacteria.p__Firmicutes.c__Negativicutes.o__Selenomonadales.f__Veillonellaceae.g__Veillonella
k__Bacteria.p__Proteobacteria.c__Gammaproteobacteria.o__Enterobacteriales.f__Enterobacteriaceae.g__Escherichia.s__Escherichia_coli
Relative abundance
class: 0
class: 1

Figure S7


samples
genes
Patient ID/
genome ID
gene present
gene absent
CLADE
CLADE 1
CLADE 2
CLADE 3

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Samples available for analysis related to Experimental Procedures. Samples available per infant are plotted by day postpartum and color-coded to indicate relative abundance of *E. coli* at <1% (blue), 1-<10% (light orange), and at least 10% (orange). Samples are divided by analysis group (days 3-9, days 10-16 and days 17-22). Day of NEC onset is indicated by the number of the day postpartum in an outlined box. Also indicated is which of the two extraction protocols applied, as well as whether the infant developed NEC or died. Patient IDs for term infants begin with “3”.

Figure S2. Samples flow chart for analysis related to Experimental Procedures. General characteristics of all 405 samples available for analysis are described (green). Read data from all samples was analyzed by MetaPhlAn analysis and also mapped to *E. coli* genomes for gene content analysis (white). All samples extracted under protocol A were included in community analysis (Purple). All samples with data mapping to *E. coli* genomes were considered for *E. coli* gene content analysis (light yellow); several samples did not contain enough read data to determine MLST type or call gene content. Samples with at least 1% relative abundance of *E. coli* and had a determined MLST type were considered for MLST analysis (yellow). Samples with at least 10% relative abundance of *E. coli* and represented the sample with highest *E. coli* abundance for any sample from an individual infant were included in gene content analysis (orange). Odds ratio calculations with respect to UPEC included all preterm infants (blue) and used criteria of MLST analysis (yellow) to determine UPEC presence in an infant.

Figure S3. Clinical factors associated with community diversity and prevalence of *E. coli* and *Klebsiella* spp related to Figure 1. (A) Shannon's index of alpha diversity per infant group. Standard deviation is indicated by error bar. (B) Prevalence of *E. coli* and *Klebsiella* spp. per infant group. Prevalence is defined as the frequency each organism is carried in infants at greater than 1% relative abundance. The number of infants per group is indicated at bottom of figure. Date for each of the three collection windows is presented from left to right as day 3-9 (dark), day 10-16 (medium), and day 17-22 (light). Except for the 'Term' group (darker green), all other groups include only preterm infants less than 30 weeks gestational age.

Figure S4. Taxonomic biomarkers enriched in infants sampled at days 3-9 postpartum related to Figure 1. (A) Cladogram generated using LEfSe (Segata et al., 2011) indicates taxonomies found to be significantly enriched in either preterm or term infants. (B) Taxonomies significantly enriched in Term infants (“class:1”, right side). (C) Taxonomies significantly enriched in Preterm infants (“class:0”, left side). (D) Cladogram indicates taxonomies found to be significantly enriched in preterm infants with low (0-6 days) or high (7-14 days) antibiotic administration during their first 14 days of life. (E) Taxonomies significantly enriched in preterm infants with high antibiotic administration (“class:1”, right side) or low antibiotic administration (“class:0”, left side). (F) Cladogram indicating taxonomies enriched in preterm Cesarean births. (G) Bacilla were significantly enriched in preterm Cesarean births (“class:1”, right side). Histograms plot relative abundance for each patient and were generated using LEfSe. Relative abundance scales differ for each plot. Median (dashed line) and Mean (solid line) relative abundances per group are indicated.

Figure S5. Taxonomic biomarkers enriched in infants sampled at days 10-16 postpartum related to Figure 1. (A) Cladogram generated using LEfSe (Segata et al., 2011) indicates taxonomies found to be significantly enriched in either preterm or term infants. (B) Taxonomies significantly enriched in preterm infants (“class:0”, left side). (C) indicates taxonomies found to be significantly enriched in preterm infants with high (7-14 days) antibiotic administration. (D) Taxonomies significantly enriched in preterm infants receiving 7-14 days (“class:1”, right side) and 0-6 days (“class:0”, left side) antibiotic administration. (E) Cladogram indicating taxonomies enriched in preterm vaginal or Cesarean births. (F) Taxonomies significantly enriched in preterm Cesarean births (“class:1”, right side). Histograms plot relative abundance for each patient and were generated using LEfSe. Relative abundance scales differ for each plot. Median (dashed line) and Mean (solid line) relative abundances per group are indicated.

Figure S6. Taxonomic biomarkers enriched in infants sampled at days 17-22 postpartum related to Figure 1. (A) Cladogram generated using LEfSe (Segata et al., 2011) indicates taxonomies found to be significantly enriched in either preterm or term infants. (B) Taxonomies significantly enriched in preterm infants ("class:0", left side). (C) *Enterobacter aerogenes* is enriched in preterm infants delivered by Cesarean ("class:1; right side). (D) Cladogram indicates taxonomies found to be significantly enriched in preterm infants with low (0-6 days) or high (7-14 days) antibiotic administration. (E) Taxonomies significantly enriched in preterm infants with 0-6 days ("class:0", left side) and 7-14 days ("class:1", right side) antibiotic administration. (F) Cladogram indicating taxonomies enriched in preterm infants who did or did not develop NEC. (G) *E. coli* is significantly enriched in preterm NEC infants ("class:1", right side) while *Veillonella* is enriched in preterm infants who did not develop NEC ("class:0"; left side). Histograms plot relative abundance for each patient and were generated using LEfSe. Relative abundance scales differ for each plot. Median (dashed line) and Mean (solid line) relative abundances per group are indicated.

Figure S7. Infant samples clustered by *E. coli* gene content indicate samples from same infant collected over time contain the same strain of *E. coli* related to Figure 3. Infant samples subjected to *E. coli* gene content analysis were clustered using gene present/absence data using the Bray-Curtis dissimilarity metric. Reference *E. coli* genomes were also included in the cluster analysis. Patient and genome IDs are indicated. Clade identities relating to Figure 3 are also indicated.