Long-term ecological and evolutionary dynamics in the gut microbiomes of carbapenemase-producing Enterobacteriaceae colonized subjects
Supplementary Figure 1: Residual Shannon diversity, after subtracting the intercept term due to the random effect in a linear mixed-effects model that treats colonization status as the fixed effect, and the stated covariates as random effects (i.e. antibiotic usage since last visit, hospitalization status, individual subjects, gender and ethnicity; see Supplementary File 1). The p-values originate as part of the output from the respective linear mixed effect models, comparing CPE positive timepoints against other groups. (antibiotics, hospitalization: n = 290; individuals, gender, ethnicity: n = 363; respective p-values for CPE positive vs. 2 months post-clearance, and CPE positive vs. family members – antibiotics: 0.100, 1.21×10^{-5}; hospitalization: 0.00215, 3.00×10^{-7}; individuals: 0.0145, 1.94×10^{-4}; gender: 2.85×10^{-5}, <2×10^{-16}; ethnicity: 1.28×10^{-4}, 1.36×10^{-12}) Centre lines in the boxplots represent median values, box limits represent upper and lower quartile values, whiskers represent 1.5 times the interquartile range above the upper quartile and below the lower quartile. *** = p<0.01, ** = p<0.05, * = p<0.1.
Supplementary Figure 2: (a) Boxplots showing genus-level Shannon diversity distributions for different timepoints for index patients (“CPE positive” = during colonization, “At clearance” = within 1 month of decolonization, and “2 months post-clearance” = time points that were >2 months after decolonization), and all timepoints for family members. This plot is similar to Figure 1b but with all Enterobacteriaceae counts removed and relative abundances renormalized to 1. (b) Corresponding microbial richness values. (n = 346 timepoints; respective two-sided Wilcoxon rank-sum test p-values for CPE positive vs. 2 months post-clearance, and CPE positive vs. family members – Shannon diversity: 2.20×10^{-6}, 2.20×10^{-14}; richness: 9.97×10^{-7}, 2.64×10^{-13}). Centre lines in the boxplots represent median values, box limits represent upper and lower quartile values, whiskers represent 1.5 times the interquartile range above the upper quartile and below the lower quartile. *** = Wilcoxon rank-sum p<0.01 and all other comparisons were not statistically significant.
Sugar metabolism
- Rubisco short Heterocyclic fermentation
- Glucose and pyruvate degradation
- TCA and glycine bypass
- Pentose (phosphate pathway)
- Glycolysis and Entner-Doudoroff
- TCA cycle VIII
- L-1,2-propanediol degradation
- Gluconeogenesis
- TCA cycle II
- Peptidoglycan biosynthesis I
- TCA cycle II
- Glyceraldehyde bypass and TCA
- 2-methylcitrate cycle
- 3-methylcitrate cycle
- TCA cycle VII
- Incomplete reductive TCA cycle
- Myo-inositol degradation
- Glycerol degradation to 1,3-propanediol
- ADP-L-glycerate-3-dehydro-heptose biosynthesis
- Fucose and mannose degradation
- Acetyl-CoA biosynthesis
- Glucose and glycogen-1-phosphate degradation
- Tryptophan biosynthesis
- D-glucarate and D-galactarate degradation
- D-glucarate degradation
- D-glucarate degradation
- Trichloroacetic acid biosynthesis

Amino acid metabolism
- L-isoleucine biosynthesis (from threonine)
- L-arginine degradation II
- Brancd amino acid biosynthesis
- L-lysine-synthesis II
- L-lysine-synthesis II
- L-lysine-synthesis II
- L-lysine-synthesis II
- L-lysine-synthesis II
- L-lysine-synthesis II
- L-lysine-synthesis II
- L-lysine-synthesis II
- L-threonine biosynthesis
- Aromatic acid and amino acid biosynthesis

Fatty acid metabolism
- Fatty acid elongation — saturated
- Oleic acid biosynthesis IV
- Palmitoleic acid biosynthesis I
- Octanoate (fatty-acid protein) biosynthesis
- Stearate biosynthesis II
- (5Z)-dodec-6-enoyl-biosynthesis
- Fatty acid beta-oxidation II
- Fatty acid beta-oxidation II
- Fatty acid beta-oxidation II
- Fatty acid beta-oxidation II
- Fatty acid beta-oxidation II
- Fatty acid beta-oxidation II
- Fatty acid beta-oxidation II
- Fatty acid beta-oxidation II
- Fatty acid beta-oxidation II

Others
- N-acetylbranchedamine biosynthesis
- Polymyxin resistance
- Circumvent 3-counsate degradation to isopent-4-enoyl
- 3-phenylpropanate & 3-MPP degradation to isopent-4-enoyl

Supplementary Figure 3: Differentially abundant pathways in colonized (CPE positive) and post-decolonization (CPE negative) gut metagenomes (two-sided Wilcoxon rank-sum test, FDR-adjusted p-value<0.05, LDA score>2) after removal of reads assigned to Enterobacteriaceae in HUMAnN results and renormalization of abundances.
**Supplementary Figure 4:** (a) Boxplots showing antibiotic-resistant gene (ARG) abundance distributions for different timepoints for index patients ("CPE positive" = during colonization, "At clearance" = within 1 month of decolonization, and "2 months post-clearance" = time points that were >2 months after decolonization), and all timepoints for family members (p-values of comparison with the CPE positive group – 2 months post clearance: 4.03×10^{-6}, Family members: 1.18×10^{-11}). (b) Distribution of corresponding abundance values for carbapenem resistance genes (cARG; ARO:0000020) (p-values of comparison with the CPE positive group – At clearance: 1.49×10^{-5}, 2 months post clearance: 9.02×10^{-9}, Family members: 6.58×10^{-12}). (c) ARG abundance values after subtracting cARG abundances (p-values of comparison with the CPE positive group – At clearance: 0.0731, 2 months post clearance: 8.18×10^{-9}, Family members: 1.96×10^{-11}). *** = Wilcoxon rank-sum p<0.01, * = Wilcoxon rank-sum p<0.1, and all other comparisons were not statistically significant. n = 345, two-sided Wilcoxon rank-sum test. Centre lines in the boxplots represent median values, box limits represent upper and lower quartile values, whiskers represent 1.5 times the interquartile range above the upper quartile and below the lower quartile.
Supplementary Figure 5: (a) Representative allele frequency spectra for samples classified as “one strain”, “two strains” and “multiple strains”, obtained from 3 consecutive timepoints in subject 0506-T. (b) Boxplots depicting the precision of metagenomic SNVs (One strain, allele frequency ≥0.98) evaluated using SNVs present in corresponding CPE isolates where available (n = 33). (c) Fraction of metagenomic SNVs (all allele frequencies) called using a shared species reference that were recapitulated using sample-specific analysis with CPE isolate references (n = 117). Centre lines in the boxplots represent median values, box limits represent upper and lower quartile values, whiskers represent 1.5 times the interquartile range above the upper quartile and below the lower quartile.
**Supplementary Figure 6:** Summary of Figures 2A and 2B, with relative proportions of each strain composition type across the timepoints V00 to V11. Stacked bar charts were plotted for each species (*E. coli* and *K. pneumoniae*) and cohort group (Subjects and Family members).
**Supplementary Figure 7:** Pictorial depiction of haplotype populations consistent with the sub-strain clusters seen in Figure 3 for *E. coli*. Each strip of boxes represents one haplotype, with white boxes representing reference bases and coloured bases representing SNVs that are part of different time-series clusters from Figure 3a. Note that the population on the left depicts the putative state for timepoints V00-V03 and at timepoint V05 there is a loss of the CPE sub-strain to give the population seen on the right.
Supplementary Figure 8: Hierarchical clustering of plasmid sequences found in (a) *E. coli* and (b) *K. pneumoniae*, for subject 1674-T across various timepoints (based on Mash distance). Sequences with >95% identity were grouped together and a representative plasmid is shown in the leaves. (c) Table showing the proportion of each representative plasmid that is covered by contigs in various samples. Values in italics and grey indicate proportions that fall below 0.8 and were used to determine the presence, absence pattern shown in Figure 3h.
| Carbapenemase gene | *E. coli* | *K. pneumoniae* |
|--------------------|-----------|----------------|
| KPC                | 17        | 18             |
| OXA-48             | 42        | 31             |
| IMP                | 6         | 0              |
| NDM                | 8         | 13             |
| IMI                | 0         | 1              |

**Supplementary Table 1:** Types of carbapenemase genes found in CPE *E. coli* and *K. pneumoniae* isolates.