



OPEN

Long-term ecological and evolutionary dynamics in the gut microbiomes of carbapenemase-producing Enterobacteriaceae colonized subjects

Jonathan T. L. Kang^{1,4}, Jonathan J. Y. Teo^{1,4}, Denis Bertrand^{1,4}, Amanda Ng¹, Aarthi Ravikrishnan¹, Melvin Yong², Oon Tek Ng³, Kalisvar Marimuthu³, Swaine L. Chen^{1,2}, Kern Rei Chng¹, Yunn-Hwen Gan² and Niranjan Nagarajan^{1,2}

Long-term colonization of the gut microbiome by carbapenemase-producing Enterobacteriaceae (CPE) is a growing area of public health concern as it can lead to community transmission and rapid increase in cases of life-threatening CPE infections. Here, leveraging the observation that many subjects are decolonized without interventions within a year, we used longitudinal shotgun metagenomics (up to 12 timepoints) for detailed characterization of ecological and evolutionary dynamics in the gut microbiome of a cohort of CPE-colonized subjects and family members ($n = 46$; 361 samples). Subjects who underwent decolonization exhibited a distinct ecological shift marked by recovery of microbial diversity, key commensals and anti-inflammatory pathways. In addition, colonization was marked by elevated but unstable Enterobacteriaceae abundances, which exhibited distinct strain-level dynamics for different species (*Escherichia coli* and *Klebsiella pneumoniae*). Finally, comparative analysis with whole-genome sequencing data from CPE isolates ($n = 159$) helped identify substrain variation in key functional genes and the presence of highly similar *E. coli* and *K. pneumoniae* strains with variable resistance profiles and plasmid sharing. These results provide an enhanced view into how colonization by multi-drug-resistant bacteria associates with altered gut ecology and can enable transfer of resistance genes, even in the absence of overt infection and antibiotic usage.

The global dissemination of antibiotic resistance genes (ARGs) among pathogenic bacteria is a major public health problem that, if left unaddressed, would reduce efficacy of treatments, elevate costs and increase mortality¹. Of particular concern is the spread of carbapenemase-producing Enterobacteriaceae (CPE)², with their ability to degrade carbapenems often acquired from plasmids encoding carbapenemases³, thus rapidly endangering these antibiotics of last resort⁴. In addition to life-threatening infections, asymptomatic gut colonization by CPE is increasingly common⁵, creating ARG transmission reservoirs⁶. While previous studies have focused on epidemiology^{2,7} and molecular aspects^{4,8}, the natural history of gut colonization including ecological and evolutionary changes linked to ARG transmission or CPE decolonization remains unexplored.

Recent studies into host–microbiome–pathogen interactions have provided important insights into pathogenesis⁹, immune response¹⁰ and treatment avenues¹¹ for various pathogens. Some

have leveraged metagenomics to temporally track microbiomes and understand ecological responses to overt infection¹¹. As microbes have rapid turnover, whole-genome sequencing (WGS) of pathogens has helped characterize intra-host evolution during chronic infections, identifying key enzymes for host adaptation or colonization¹². Alternatively, deep metagenomic sequencing can reveal nucleotide-level variation for many species of interest¹³, shedding light on strain-level dynamics in microbiomes. This approach has characterized stable microbiomes in healthy individuals as well as dynamic changes during faecal microbiota transplantation¹⁴. Asymptomatic gut colonization by CPE strains presents a unique opportunity to study an intermediate phenomenon, that is, strain competition with commensals, and associated ecological and evolutionary adaptations, without overt infection or disease.

In this Letter, we conducted longitudinal gut microbial analysis for a cohort of index subjects ($n = 29$, CPE colonized at recruitment) and their family members ($n = 17$, not CPE colonized) with up to 12 timepoints within a year, to obtain multi-scale¹⁵ (microbiome composition-, strain- and gene-level) characterization of ecological and evolutionary changes during CPE colonization. On the basis of deep shotgun metagenomics of stool DNA, we observed distinct ecological shifts marked by recovery of diversity and key commensals in association with CPE decolonization. CPE colonization was marked by elevated but unstable Enterobacteriaceae abundances, which exhibited specific strain-level dynamics for different species (*Escherichia coli* and *Klebsiella pneumoniae*). Comparative analysis with WGS data from CPE isolates ($n = 159$) identified the presence of highly similar *E. coli* and *K. pneumoniae* strains with variable resistance profiles and plasmid sharing. These results provide an enhanced view into how colonization by multi-drug-resistant bacteria associates with altered gut ecology and can enable transfer of resistance genes, even without overt infection.

Results

Ecological shifts and recovery after CPE decolonization. Leveraging the observation that CPE carriage is resolved within 3 months, with 98.5% probability within a year for our cohort¹⁶ (longer for other cohorts¹⁷), we tracked gut microbiome composition in these individuals for a year to understand ecological changes associated with decolonization (12 timepoints, 361 samples in total;

¹Genome Institute of Singapore, Singapore, Singapore. ²Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ³Institute of Infectious Diseases and Epidemiology, Tan Tock Seng Hospital, Singapore, Singapore. ⁴These authors contributed equally: Jonathan T. L. Kang, Jonathan J. Y. Teo, Denis Bertrand. ✉e-mail: nagarajann@gis.a-star.edu.sg

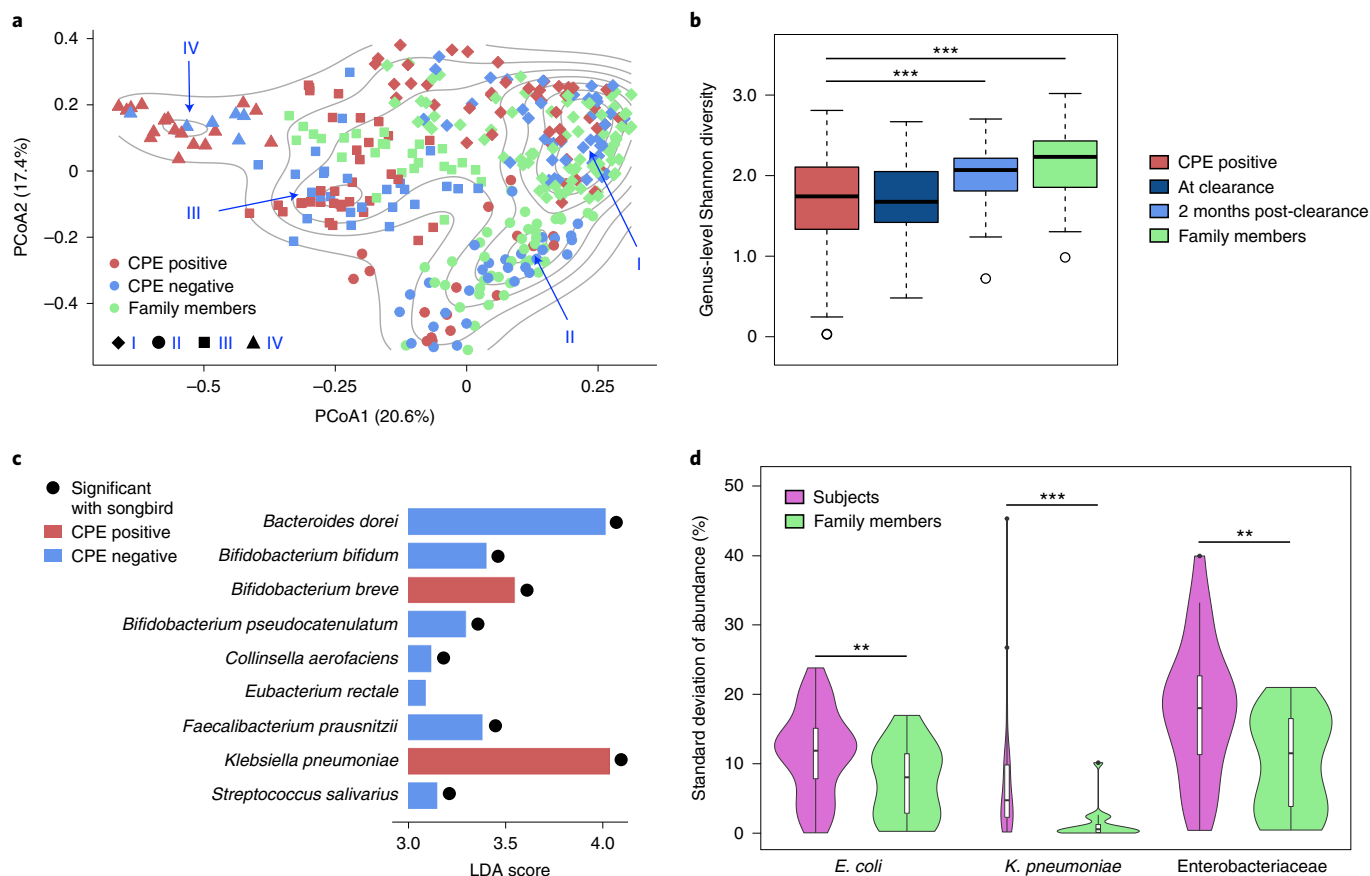


Fig. 1 | Shifts in gut microbial ecology associated with CPE colonization. **a**, PCoA plot showing how gut microbial community composition varies in relation to CPE colonization status (genus-level Bray–Curtis dissimilarity; Unifrac plot in Extended Data Fig. 1). Contour lines indicate similar density with regions of locally higher density associated with distinct community configurations (defined by average linkage clustering in Extended Data Fig. 1; marked with labels I, II, III and IV). **b**, Box plots showing genus-level Shannon diversity distributions for different timepoints for index patients ('CPE positive', during colonization; 'At clearance', within 1 month of decolonization; '2 months post-clearance', timepoints that were >2 months after decolonization) and all timepoints for family members ($n = 347$ timepoints; two-sided Wilcoxon rank-sum test P values: CPE positive versus 2 months post-clearance, 1.15×10^{-4} ; CPE positive versus family members, 2.46×10^{-12}). **c**, Species that were found to be enriched in CPE-positive (red) and CPE-negative (blue) samples along with their linear discriminant analysis (LDA) scores based on LEfSe analysis. Results that were significant on the basis of Songbird analysis as well are indicated with a solid circle. **d**, Violin plots showing the standard deviation of relative abundances (ignoring relative abundances <0.1% to avoid the effect of detection limit for metagenomics) over time of various taxa in different individuals (subjects and family members) ($n = 43$ individuals, two-sided Wilcoxon rank-sum test P values: *E. coli*, 0.021; *K. pneumoniae*, 4.96×10^{-5} ; Enterobacteriaceae, 0.011). For all subfigures: * $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$ based on the Wilcoxon rank-sum test, and all other comparisons were not statistically significant. Centre lines in the boxplots represent median values, box limits represent upper and lower quartile values, and whiskers represent 1.5 times the interquartile range above the upper quartile and below the lower quartile.

Extended Data Table 1 and Supplementary File 1). Specifically, stool samples were obtained from hospital patients who screened positive for CPE carriage ($n = 29$, index subjects) and non-CPE-colonized family members ($n = 17$, home-environment-matched controls) and characterized via deep shotgun metagenomics (>50 million 2×100 bp reads; Methods). Principal coordinates analysis (PCoA) with average-linkage clustering based on taxonomic profiles identified multiple distinct community configurations (I, II, III and IV), where CPE-positive samples (from stool culture and quantitative PCR¹⁶) were less common in configurations I and II, and more common in configurations III and IV (Fig. 1a, Extended Data Fig. 1 and Supplementary File 2). This statistically significant shift of CPE-positive samples along PCoA1 (Wilcoxon P value < 1.4×10^{-8} ; Extended Data Fig. 2a) is defined by a gradient of abundances strongly correlated with *Escherichia* (negative, that is, more abundant in configuration IV samples) and *Bacteroides* (positive; Extended Data Fig. 2b). A similar shift was observed when comparing taxonomic profiles for configuration IV versus configuration I (Supplementary File 3). Interestingly, while configuration IV has no

microbiomes from family members, a few CPE-negative samples from index subjects also cluster here.

Grouping samples on the basis of temporal proximity to CPE clearance highlighted that, while CPE-positive samples have lowest average diversity¹⁸, there is a gradual increase around the time of decolonization and post-decolonization, with diversity reaching the higher levels seen in family members after 2 months (Fig. 1b). This pattern was seen even after accounting for potential confounders, including antibiotic usage, hospitalization status, multiple timepoints for an individual, gender and ethnicity, in a linear mixed-effects model (Supplementary Fig. 1 and Methods). We investigated whether CPE colonization alone could explain these changes by computationally subtracting Enterobacteriaceae from taxonomic profiles and recomputing diversity metrics. We noted that richness and Shannon diversity consistently preserved the trend of increasing during decolonization (Supplementary Fig. 2), pointing to a wider microbiome shift besides Enterobacteriaceae.

Shifts in diversity during decolonization were also reflected in microbiome similarity, with Bray–Curtis distances to family

members being highest in CPE-positive samples, reducing during and post-decolonization towards baseline values seen among family members (Extended Data Fig. 3). These results highlight the ecological shift associated with CPE colonization that resolves post-decolonization, but with residual effects in some individuals.

To identify species associated with decolonization, we conducted differential-abundance analysis based on CPE status (Methods and Supplementary File 2). While most Enterobacteriaceae species were not differentially abundant, *K. pneumoniae*^{2,4} had strong association with CPE-positive status (Fig. 1c). Only one other species (*Bifidobacterium breve*) was enriched in CPE-positive samples, while seven species were depleted in CPE-negative samples. These included important commensals known to reduce gut inflammation through diverse pathways, including *Bacteroides dorei* (decreasing gut microbial lipopolysaccharide production¹⁹), *Faecalibacterium prausnitzii* (butyrate production²⁰) and *Bifidobacterium (bifidum, pseudocatenulatum)*; inhibition of NF- κ B activation²¹), and could thus aid CPE decolonization²².

Differential pathway analysis based on CPE status provided further supporting evidence that key inflammatory pathways (for example, sulphate reduction) are enriched during colonization, highlighting their role in the process (Extended Data Fig. 4 and Supplementary File 4). Furthermore, aerobic respiration and oxidative phosphorylation pathways (for example, pentose phosphate) were enriched during CPE colonization consistent with a model of gut oxygenation proposed previously²³. These results were recapitulated after excluding Enterobacteriaceae from functional profiles, emphasizing that they are not solely from CPE colonization and have substantial contributions from other species (Supplementary Fig. 3). Micro-aerophilic niches for Enterobacteriaceae due to antibiotic treatment provide another potential explanation²⁴, as antibiotic usage was common (preceding ~25% timepoints; Supplementary File 1). Notably, while ARGs were enriched in gut microbiomes for CPE-positive timepoints, no significant differences were observed between index subjects and family members at other timepoints (Supplementary Fig. 4).

While Enterobacteriaceae species were enriched in CPE-positive samples relative to CPE-negative samples (Wilcoxon P value $< 3.5 \times 10^{-5}$), index subjects at CPE-negative timepoints also showed enrichment compared with family members (Wilcoxon P value = 0.01; Extended Data Fig. 5). Overall, Enterobacteriaceae composition varied across individuals, with *E. coli* and *K. pneumoniae* being most common, but other *Escherichia*, *Klebsiella*, *Enterobacter* and *Proteus* species also being moderately abundant across individuals and timepoints (Extended Data Fig. 5). The abundant Enterobacteriaceae species in an individual were not necessarily the CPE species (for example, subject 0505-T, timepoints 1–3). Enterobacteriaceae profiles also shifted rapidly (for example, in 0457-T and 0512-T, timepoint 6) with higher temporal variation in index subjects (Wilcoxon P value < 0.05 ; Fig. 1d and Extended Data Fig. 5). Together, these results suggest that CPE colonization is maintained by an altered, dynamic pro-inflammatory microenvironment supporting Enterobacteriaceae, which resolves concurrently with microbiome recovery²⁵.

Distinct strain-level dynamics of Enterobacteriaceae species.

Analysing the deep metagenomic data at higher resolution, we investigated strain-level dynamics for the two most prevalent Enterobacteriaceae (*E. coli* and *K. pneumoniae*). Read mapping to references identified single-nucleotide variants (SNVs), and allele frequency modes were used with a classical population genetics approach to determine strain count¹³ (Methods and Supplementary Fig. 5). For 53% of samples (*E. coli*, 63%; *K. pneumoniae*, 38%) where a species was detected (abundance $> 0.1\%$), read coverage was sufficient to identify strain count (one, two or multiple strains, otherwise low coverage; Fig. 2a,b and Supplementary Fig. 6). As

expected for a gut commensal²⁶, *E. coli* was at comparable frequencies in index subjects (86%) and family members (90%), and more frequently detected relative to *K. pneumoniae* (Fisher's exact P value $< 5 \times 10^{-20}$; Fig. 2a). *K. pneumoniae* was, however, more frequently found in index subjects (70%) relative to family members (39%), consistent with the hypothesis that a pro-inflammatory environment might be facilitating colonization (Fisher's exact P value $< 2 \times 10^{-9}$; Fig. 2b).

Among classified samples, *E. coli* was frequently single strain in index patients (44%) while family members often had multiple strains (44%; Supplementary Fig. 6), pointing to clonal dominance in pro-inflammatory environments. A few individuals maintained a single strain over several months (for example, 0505-T, 1667-T and 0506-T), indicating that this can be a stable state (Fig. 2a). For *K. pneumoniae* despite sporadic detection, the multi-strain state was more common (49%), consistent with the hypothesis that even in pro-inflammatory environments, no clone outcompetes others²⁷ (Fig. 2b and Supplementary Fig. 6). Overall, in agreement with previous observations (Fig. 1d and Extended Data Fig. 5), Enterobacteriaceae strain compositions were highly variable temporally.

Capturing transition frequencies between strain compositions as first-order Markov models, we noted distinct patterns for *E. coli* and *K. pneumoniae*, and index subjects versus family members (Fig. 2c and Methods). For example, *E. coli* colonization is more likely to stay single strain for index patients (69%) versus family members (57%) and single-strain *K. pneumoniae* colonization (52%; Fig. 2c). When *E. coli* is not detected, this state is more likely to be maintained in index subjects (47%) versus family members (22%; Fig. 2c). Overall, the Markov model predicts that *E. coli* is maintained as single strains (index subjects, 41%; family members, 39%). Contrastingly, *K. pneumoniae* frequently converges to not detected state in subjects (42%) and family members (75%). Grouping classes of strain detection, we tested if transition probabilities differed across cohorts and species (Fig. 2d). For *E. coli*, transition probabilities were not significantly different between cohorts (Fisher's exact P value > 0.05 ; Fig. 2d). In contrast, driven by stark detected/not detected patterns for *K. pneumoniae*, index subjects had significantly different transition probabilities compared with family members for various groupings involving the not detected state ('All classifications', 'Detected versus not detected' and 'Fixed versus variable strains'; Fisher's exact P value $< 10^{-2}$; Fig. 2d). These results further highlight differences in strain-level dynamics for Enterobacteriaceae in the potentially pro-inflammatory gut microbiome milieu of index subjects.

Substrain variation and plasmid sharing in Enterobacteriaceae.

Samples determined to have a single strain can still exhibit sub-strain variation relative to this genomic background, similar to viral quasi-species. Characterizing distribution of such intra-host variations across genes could identify adaptive changes important for CPE colonization, similar to recent studies with mouse models and strain isolates^{28,29}. To analyse standing variation in Enterobacteriaceae, we identified low-frequency ($< 50\%$) SNVs in single-strain timepoints (*E. coli*, 30,155; *K. pneumoniae*, 13,061), annotating them for function-altering changes to reveal adaptive signatures in Enterobacteriaceae during gut colonization (Supplementary File 5 and Methods). We found 5,919 and 1,787 function-altering changes in *E. coli* and *K. pneumoniae*, including several in key polysaccharide utilization and virulence genes (for example, *lacZ*, *lacY* and ECIA139_4258 (putative invasins/intimin protein)) similar to key genes previously identified from isolate WGS as undergoing selection for human gut colonization³⁰ (Table 1). Visualizing function-altering SNVs in polysaccharide-utilization genes, where adaptive mutations can reflect pressures to use diet-derived polysaccharides, identified structural motifs that might be key functionally (Extended Data Fig. 6). Consistent with

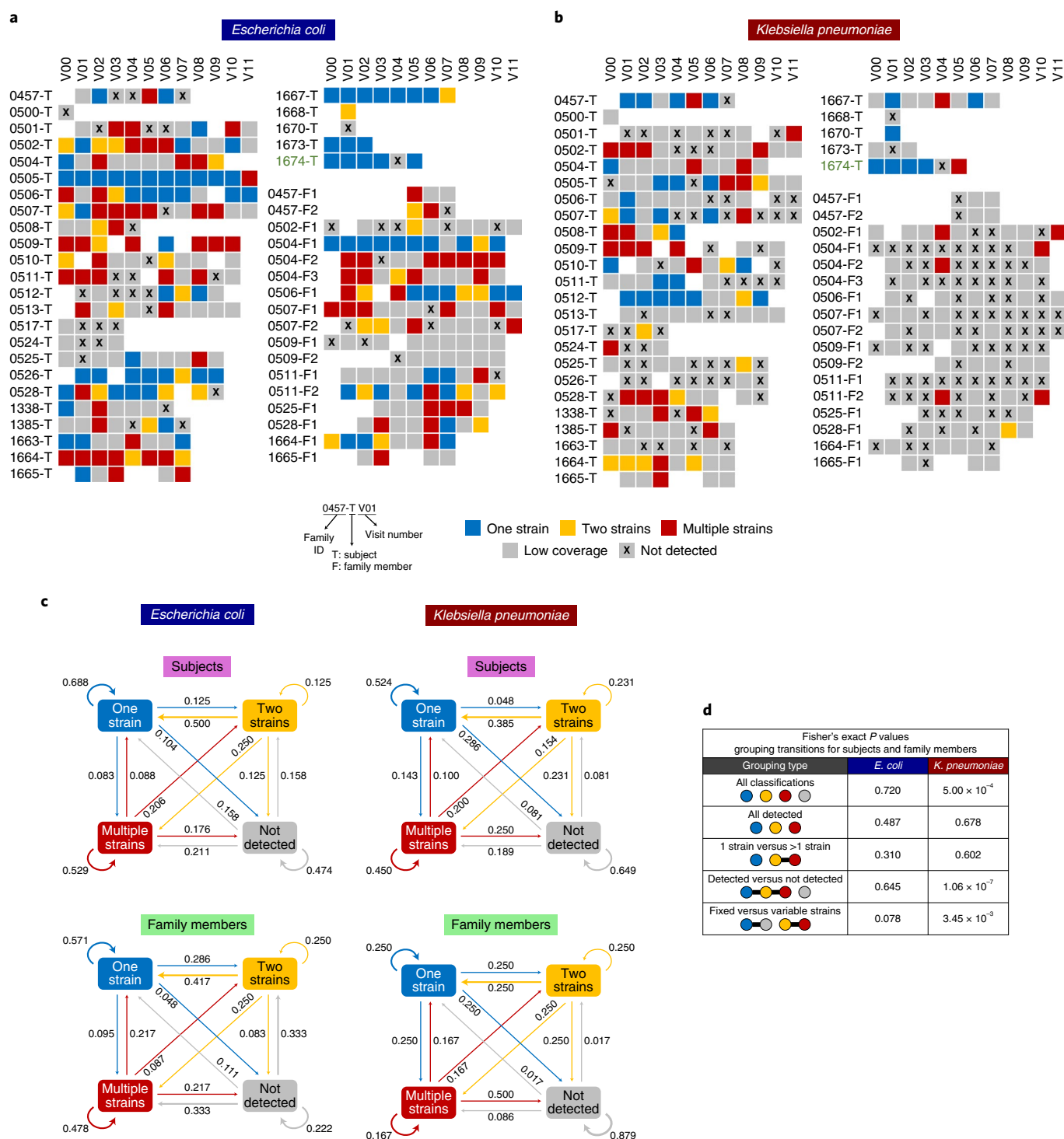


Fig. 2 | Enterobacteriaceae strain variations and dynamics in index subjects and family members. **a,b**, Strain composition for *E. coli* (**a**) and *K. pneumoniae* (**b**) determined on the basis of allele frequency distributions per sample. Samples where a species was not detected (relative abundance <0.1%) or in which the genome had low coverage (<10x) were distinguished from those where one, two or more strains were confidently detected. Each row depicts the multiple timepoints for a subject (where available), with index subjects indicated with a T and family members with an F in subject IDs. **c**, First-order Markov models showing the probability of transition between different states (excluding the low coverage state where information is missing). **d**, Table showing statistical significance of differences in transition probabilities between index subjects and family members, for various groupings of states and Enterobacteriaceae species. States placed in the same grouping are connected by a horizontal line.

our study of low-frequency SNVs, most regions bear signatures of purifying selection ($dN/dS < 0.5$; Extended Data Fig. 6), though identified genes were significantly enriched relative to genome-wide average for non-synonymous SNVs (Table 1).

To study SNVs further relative to CPE decolonization, time-series information was used to cluster co-varying SNVs (Methods). Interestingly, in some subjects multiple clusters were identified, revealing distinct substrain lineages differing by few hundred SNVs

Table 1 | Top genes with substrain variation

| Gene | Protein | SNV count | P value |
|----------------------|---|-----------|--------------------|
| <i>E. coli</i> | | | |
| ECIAI39_4258 | Putative invasin/intimin protein | 48 | <10 ⁻¹⁰ |
| ECIAI39_0530 | Host specificity protein J of prophage | 34 | <10 ⁻¹³ |
| <i>ydcM</i> | Putative transposase | 24 | <10 ⁻¹⁵ |
| ECIAI39_1027 | Putative GTP-binding domain | 19 | <10 ⁻¹² |
| ECIAI39_4240 | Putative anti-restriction protein | 18 | <10 ⁻¹⁴ |
| <i>yeeS</i> | Putative DNA repair protein; CP4-44 prophage | 18 | <10 ⁻¹⁶ |
| <i>lacZ</i> | β-D-galactosidase | 17 | <0.05 |
| ECIAI39_4893 | Putative tail fibre component K of prophage | 14 | <10 ⁻⁷ |
| <i>lacY</i> | Lactose/galactose transporter | 14 | <10 ⁻⁴ |
| <i>Rz</i> | Endopeptidase from phage origin (lysis protein) | 13 | <10 ⁻⁹ |
| ECIAI39_1018 | Hypothetical protein | 11 | <10 ⁻⁴ |
| ECIAI39_4867 | Hypothetical protein | 10 | <10 ⁻² |
| <i>yncK</i> | Transposase ORF A | 10 | <10 ⁻⁸ |
| <i>K. pneumoniae</i> | | | |
| KPHS_51490 | Putative transposase | 30 | <10 ⁻³⁹ |
| KPHS_35720 | Putative transposase | 18 | <10 ⁻¹³ |
| KPHS_22580 | Hypothetical protein | 12 | <10 ⁻⁴ |

Top ten genes with most function-altering SNVs (≥10) with low frequencies (allele frequency < 0.5) for *E. coli* and *K. pneumoniae*. SNVs in distinct individuals were counted separately, but multiple timepoints were counted as one. P values were computed by performing a binomial test on the observed number of SNVs given the probability of a gene acquiring SNVs after accounting for (1) the probabilities of acquiring mutations across all genes and (2) the codon composition of the genes where the respective SNVs are found. Bonferroni correction ($\alpha=0.05$) was applied.

genome-wide (for example, 1674-T; Fig. 3a,b, Extended Data Figs. 7 and 8 and Supplementary Fig. 7). For subject 1674-T, both *E. coli* and *K. pneumoniae* have a dominant cluster during CPE-positive timepoints (V00–V03) that match SNV signatures seen in corresponding CPE isolates (*Shared* and *Cluster 1* SNVs; Fig. 3c,d and Supplementary Fig. 7). The subdominant cluster (*Cluster 2*; Fig. 3c,d and Supplementary Fig. 7; possibly sublineage of *Cluster 1*) has SNV signature not seen in CPE isolates but detected post-decolonization (V05), indicating that these lineages are discordant for CPE status despite high genomic similarity. For both species, decolonization coincides with appearance of distinct strains with >1,000 SNVs relative to CPE strains (V05 *unique*; Fig. 3c,d and Supplementary Fig. 7). Despite shared strain patterns for *E. coli* and *K. pneumoniae*, they exhibited dissimilar relative abundance trends, with *E. coli* being reduced at decolonization (V05) while *K. pneumoniae* peaks at this point (Fig. 3e,f). While a few dense trajectories of co-varying SNVs were detected in other individuals, many SNVs varied independently of these clusters (Extended Data Figs. 7 and 8). These results suggest that Enterobacteriaceae may share substrain dynamics during CPE decolonization while having species-specific ecological properties.

Leveraging availability of multiple timepoints across subjects, we identified SNVs whose frequencies varied notably over time (>30%). These were overlapped across subjects to identify SNVs that recurrently vary (Extended Data Table 2 and Supplementary File 6), identifying several polysaccharide utilization (*lacZ*, *lacI* and *treA*), pyruvate metabolism (*pf1B* and *pykF*) and protein synthesis

(*dnaK*, 30S and 50S ribosomal subunits) genes implicated in adaptation to nutrient limitation³¹, antibiotics³² and environmental stress³³ conditions. Several genes were common to *E. coli* and *K. pneumoniae* (*srnB*, *pnp*, *nlpI* and *pheT*), indicating that similar selection constraints may act on strains for both species. The *lacZ* gene was highlighted as having the most recurrent, frequency-varying SNVs ($n=12$), with all SNVs occurring in surface-exposed regions (Fig. 3g). Comparing accessible surface area of variant residues versus others highlighted that variants are significantly more exposed to solvent (mean 73.1Å² versus 34.1Å², Welch's *t*-test *P* value < 0.01). Many SNVs occurred in the activating interface, a region near amino-terminus required for tetramerization³⁴, indicating that they could impact *lacZ* function via complex formation dynamics.

Among genetic features prominent in CPE strains from subject 1674-T, variants in polysaccharide utilization genes were common as discussed previously (Fig. 3g). Additionally, plasmid sequence analysis identified two shared plasmids between *E. coli* and *K. pneumoniae* CPE isolates (Fig. 3h, Supplementary Fig. 8 and Methods). This included pKPC2, recently identified in hypervirulent carbapenem-resistant *K. pneumoniae* from Singapore, and harbouring *bla*_{KPC-2}, the carbapenemase that provided CPE designation for these isolates¹⁶. Additionally, pMS6192B was shared between all *E. coli* and *K. pneumoniae* from the first visit (V00; Fig. 3h). The shared plasmids have total sequence length >140 kbp and no SNVs across species, indicating a recent common source. Plasmid transfer experiments with pKPC2 between *E. coli* and *K. pneumoniae* suggest moderate conjugation frequency under in vitro conditions (~0.1%; Methods). Furthermore, 50% of plasmid-bearing clones (3/6) were observed to have SNV in pKPC2 after 300 generations, defining an upper-bound on divergence of plasmid-bearing isolates having no SNVs being 5 months (binomial *P* value < 0.05).

Discussion

The availability of metagenomic data for up to 12 timepoints over a year in this study allowed examination of long-term dynamics, enabling microbiome comparisons before and after CPE decolonization in a subject-matched fashion to study microbiome shifts associated with decolonization. This revealed ecological shifts that cannot be explained solely by CPE strain loss (for example, increase in species richness), and the specific taxonomic/functional changes observed point to the role of inflammation in maintaining Enterobacteriaceae-favourable gut environment in index subjects (for example, *Pantoea* species; Extended Data Fig. 2 and Supplementary File 3). Additionally, our data indicate that this configuration may be unstable in many individuals, opening up the possibility that interventions that reduce gut inflammation directly or via action of probiotics could reduce Enterobacteriaceae and promote CPE decolonization.

Gut inflammation has been known to create a niche for enterics such as *Salmonella*³⁵, where some species use sulphate, nitrate and tetrathionate as terminal electron acceptors for anaerobic respiration (for example, *E. coli*). The enriched pathways in CPE-colonized subjects are marked by menaquinol biosynthesis, glycolysis and respiration (tricarboxylic acid cycle), even after computationally subtracting Enterobacteriaceae contributions, indicating that the gut environment in this group is qualitatively different in oxygenation. Additionally, fucose/rhamnose and 1,2-propanediol degradation are enriched during CPE colonization, potentially serving as carbon sources for Enterobacteriaceae such as *K. pneumoniae*, which demonstrates competitive fitness with oxygen as terminal electron acceptor under such conditions²⁷. The enrichment of pentose phosphate pathway could indicate need for reducing equivalents of NADPH⁺ to maintain redox conditions or serve as nucleic acid precursors to fuel growth. Overall, the shift in microbial pathways during CPE colonization is independent of Enterobacteriaceae but favours their growth. Further work is needed to understand if this

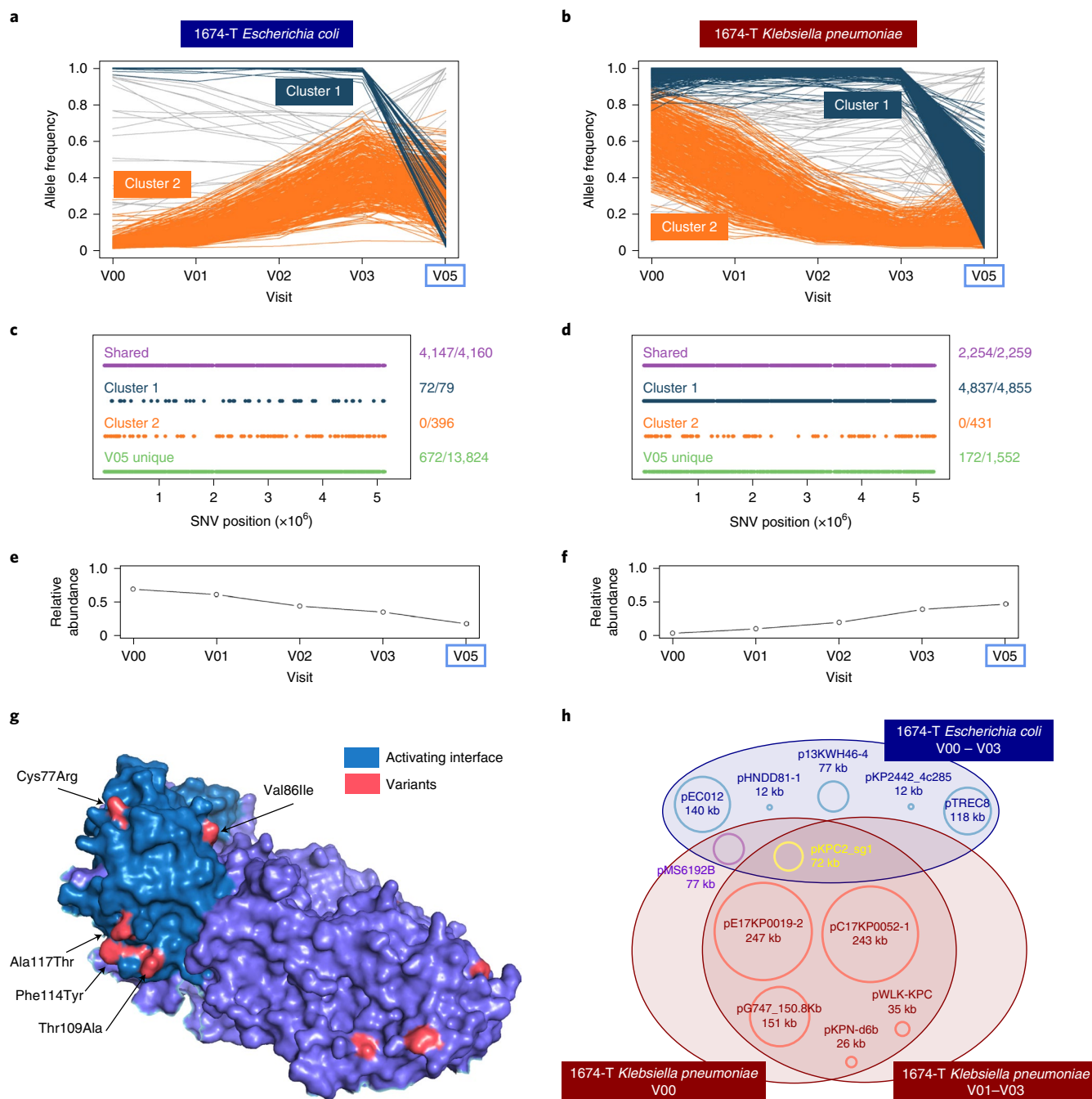


Fig. 3 | Substrain variation and plasmid sharing in Enterobacteriaceae species. a,b, Line charts showing the allele frequencies of metagenomics-derived SNVs across timepoints for *E. coli* (**a**) and *K. pneumoniae* (**b**), for subject 1674-T. SNVs that belong to different substrain clusters are coloured accordingly, and a potential model for the corresponding haplotypes is shown in Supplementary Fig. 7. All genome-wide SNVs detected at all the depicted timepoints are shown here. The blue box indicates the CPE-negative timepoint. **c,d**, Corresponding genomic distribution for SNVs belonging to cluster 1 or cluster 2 and those that are fixed at all timepoints ('Shared') or just at the CPE-negative timepoint ('V05 unique'). The associated fraction describes the number of SNVs in a grouping that are common with CPE isolates, in relation to the total number of SNVs in this grouping. **e,f**, Relative abundance of *E. coli* and *K. pneumoniae* in the gut metagenome for subject 1674-T across timepoints. **g**, Visualization of amino acid changes (red) seen in a key polysaccharide utilization protein (lacZ, pdb structure 4DUX chain A, identified from Table 1 and Extended Data Table 2) based on metagenomics-derived SNVs that changed in frequency over time and across multiple individuals. SNVs are found primarily on the protein surface, with six SNVs located on the activating interface responsible for tetramerization (teal). **h**, Diagram depicting plasmid sharing between *E. coli* and *K. pneumoniae* strains at various timepoints for subject 1674-T. Plasmid sequences were clustered at 95% similarity (Supplementary Fig. 8), and a representative plasmid for each cluster is shown in the Venn diagram as an approximately sized circle, with plasmid name and size.

shift is established by gut inflammation (as in colitis³⁶, via direct biomarker measurement, for example, calprotectin) or if diverse factors play a strong role in an individual-specific manner (for example,

antibiotics for some subjects³⁴). Notably, while reduction in microbial diversity during CPE colonization could not be attributed solely to antibiotic usage or hospitalization at a timepoint, these could be

delayed effects and would need a more controlled study design to explore further.

An alternative strategy for CPE decolonization involves introduction of species depleted in the colonized state (for example, *Faecalibacterium prausnitzii* or *Bifidobacterium bifidum*), either as probiotic formulations or through faecal microbiota transplants³⁷. Matching donors to recipients, supplementing missing species or promoting further instability in Enterobacteriaceae abundances based on ecological models could be a promising avenue to explore, similar to studies for *Clostridoides difficile*³⁸. The observed differences in colonization dynamics for Enterobacteriaceae suggest that decolonization strategies might also have to be species specific. For example, a possible explanation for presence of multiple *K. pneumoniae* strains is that they are not well adapted for gut colonization but instead opportunistically exploit a niche. Decolonization of *K. pneumoniae* strains may therefore require elimination of conditions favouring this niche such as inflammation or availability of specific sugars. The presence of single *E. coli* strains in many subjects supports a model where gut-adapted strains acquire ARGs, and plasmid-targeting strategies might be better suited in this case. Interestingly, our data suggest that human gut microbiomes can harbour multiple strains of commensals such as *E. coli* (in contrast to observations in mouse studies³⁹), even among non-CPE-colonized family members⁴⁰ (Fig. 2a). Further studies using high-throughput culturing and single-cell sequencing could help reconstruct strain genomes and unravel factors that determine niche competition⁴¹.

Identifying genes that enable CPE gut colonization provides another avenue to target interventions. As shown here, analysis of high-coverage metagenomic data to identify substrain variations can provide promising hypotheses based on in vivo evolution, similar to viral quasi-species analysis⁴², or mutagenesis-based experiments⁴³. Furthermore, isolation of lineages with distinct advantages in colonizing or avoiding decolonization (for example, cluster 2 in 1674-T), could narrow down genetic features to be investigated in vitro. Finally, the role of gut microbiome as an ARG reservoir, and plasmid sharing across Enterobacteriaceae, is of particular concern. While we cannot definitively conclude that data for 1674-T represent plasmid transfer, these observations and isolated strains serve as important resources for further investigations into plasmid transmission and CPE decolonization.

Methods

Sample collection and CPE classification. A prospective cohort study involving CPE carriers was conducted from October 2016 to February 2018. Study participants were recruited with informed consent from two tertiary healthcare centres in Singapore. CPE carriers were identified by routine collection of rectal swab samples for clinical care and/or infection prevention and control measures, in accordance with local infection control policies. The study received ethics approval from the Singapore National Healthcare Group Domain Specific Review Board 74 (NHG DSRB Reference 2016/00364) before commencement. Stool samples were first collected weekly for 4 weeks, then monthly for 5 months and finally once every 2 months for 6 months. In addition to the CPE-colonized subjects, stool samples from a number of family members were also obtained to provide a control dataset. Samples obtained from index subjects were classified as either CPE positive or CPE negative, on the basis of whether CPE genes (including *bla*_{NDM-1}, *bla*_{KPC}, *bla*_{OXA-48}, *bla*_{IMI-1} and *bla*_{IMP}) were positively identified from Enterobacteriaceae isolates found to be resistant to either meropenem or ertapenem¹⁶ (Supplementary Table 1). The presence of CPE-negative samples was used to detect CPE clearance, and samples were further classified on the basis of the amount of time elapsed since clearance, that is, before clearance, within 2 months post-clearance and more than 2 months post-clearance. Owing to the focus on household transmission and CPE carriage, dietary information was not collected in the study.

Isolate sequencing and assembly. DNA for all CPE isolates obtained from stool samples in this study (all subjects, all timepoints) was collected from Tan Tock Seng Hospital and transferred to the Genome Institute of Singapore (GIS) for WGS. Library preparation was performed using the NEBNext Ultra DNA Library Prep Kit for Illumina, and 2×151 base-pair sequencing was performed using the Illumina HiSeq 4000. Raw FASTQ reads were processed using in-house pipelines

at GIS for de novo assembly with the Velvet assembler⁴⁴ (v1.2.10), parameters optimized by Velvet Optimiser (*k*-mer length ranging from 81 to 127), contig scaffolding with Opera⁴⁵ (v1.4.1) and finishing with FinIS⁴⁶ (v0.3).

Shotgun metagenomic sequencing. DNA from 361 stool samples was extracted using the PowerSoil DNA Isolation Kit (12888, MoBio Laboratories) with modifications to the manufacturer's protocol. Specifically, to avoid spin-filter clogging, we extended the centrifugation to twice the original duration, and solutions C2, C3 and C4 were doubled in volume. DNA was eluted in 80 µl of Solution C6. Concentration of DNA was determined by Qubit dsDNA BR assay (Q32853, Thermo Fisher Scientific). For library construction, 50 ng of DNA was resuspended in a total volume of 50 µl and was sheared using Adaptive Focused Acoustics (Covaris) with the following parameters: duty factor of 30%, peak incident power of 450, 200 cycles per burst and treatment time of 240 s. Sheared DNA was cleaned up with 1.5× Agencourt AMPure XP beads (A63882, Beckman Coulter). Gene Read DNA Library I Core Kit (180434, Qiagen) was used for end-repair, A-addition and adapter ligation. Custom barcode adapters were used for cost considerations (HPLC purified, double stranded, first strand: 5' P-GATCGGAAGAGCACACGTCT; second strand: 5' ACACTCTTCCCTACACGACGCTCTTCCGATCT) in replacement of Gene Read Adapter I Set for library preparation. Library was cleaned up twice using 1.5× Agencourt AMPure XP beads (A63882, Beckman Coulter). Enrichment was carried out with indexed primers according to an adapted protocol from Multiplexing Sample Preparation Oligonucleotide kit (Illumina). We pooled the enriched libraries in equimolarity and sequenced them on an Illumina HiSeq 2500 sequencing instrument at GIS to generate 2×101 base-pair reads, which yielded around 17.7 billion paired-end reads in total and 49 million paired-end reads on average per library.

Taxonomic and functional profiling. Read quality trimming was performed using famas (<https://github.com/andreas-wilm/famas>, v0.10, -no-order-check), and microbial reads were identified by mapping and filtering out reads aligned to the human reference genome (hg19) using bwa-mem⁴⁷ (v0.7.9a, default parameters; >90% microbial reads on average). Taxonomic profiling was done using MetaPhlan⁴⁸ (v2.0, default parameters, filtering taxa with relative abundance <0.1%), and functional profiles were obtained with HUMAnN⁴⁹ (v2.0, default parameters). As a sanity check, we confirmed that species- and genus-level taxonomic profiles were not dominated by taxa that are commonly attributed to reagent or laboratory contamination⁵⁰ (Supplementary File 2). Average-linkage hierarchical clustering of taxonomic profiles was used to group samples with the number of clusters determined using Akaike information criterion. Sample α-diversity was computed using the Shannon diversity index with the vegan library in R. Differential abundance analysis was performed using LEfSe⁵¹ (v1.0.8), as a non-parametric and conservative approach to identify significantly varying taxa and functions across groups⁵². These results were further validated using Songbird⁵³ (v1.0.3; -epochs 10000 -differential-prior 0.5) analysis with Bonferroni-corrected *P* value < 0.05. Abundances of ARGs in the metagenomes was computed using a direct read mapping approach implemented in SRST2 (ref. ⁵⁴) with default parameters and the CARD_v3.0.8_SRST2 database⁵⁵.

Linear mixed-effects modelling. Linear mixed-effects modelling was conducted using the *lmer* function from the *lme4* package in R. For each model, genus-level Shannon diversity was set as the response variable, with colonization status as the fixed effect and potential confounders (for example, antibiotic usage since last visit, hospitalization status, individual subjects, gender and ethnicity; Supplementary File 1) as random-effect covariates. Residual Shannon diversity values were derived for visualization by subtracting the intercept terms corresponding to random effects.

SNV analysis. Genome assemblies were aligned to their respective reference genomes using nucmer (v3.23, -maxmatch -nosimplify), and consensus SNVs were called using the show-SNVs function in MUMmer⁵⁶. References for *E. coli* (NC_011750) and *K. pneumoniae* (NC_016845.1) were selected to minimize median distance from isolate genomes. Metagenomic SNVs (consensus and low frequency) were identified on the basis of read mapping using bwa-mem⁴⁷ to the *E. coli* and *K. pneumoniae* references (v0.7.10a; soft-clipped reads and reads with more than three or four mismatches for *K. pneumoniae* and *E. coli*, respectively, were filtered out to avoid mismapped reads) and variant calling with LoFreq⁵⁷ (v1.2.1; default parameters). Note that our stringent mapping approach restricts to only reads with >96% identity with the reference and thus will typically exclude mismapping of reads from other genomes. Additionally, genomic regions with frequent ambiguous mappings were identified on the basis of isolate sequencing data and metagenome data from samples without target species as determined from taxonomic classification (>5× coverage with *E. coli* reads on *K. pneumoniae* genome or vice versa). Calls in these regions that match positions where variants were called between isolate reads and reference sequence (allele frequency >95%) were excluded from downstream analysis. The validity of this pipeline was confirmed by noting that very few *K. pneumoniae* SNVs (median 2, mean 5.5) were called genome-wide when analysing metagenomes where taxonomic

profiling detected few *K. pneumoniae* reads (ten samples with 107–288 reads). Note that SNVs from such ‘low coverage’ samples are also excluded from further analysis in this study as defined below. To assess the impact of a shared, but potentially divergent, reference on SNV calling, reads were also mapped onto CPE isolate genomes (where available) to call SNVs and compute concordance. Isolate genome-based SNVs were translated to the common reference coordinate system using the UCSC liftover tool⁵⁸ with chain file generated using flow⁵⁹ (-fastMap -tileSize=12 -minIdentity=90).

Strain analysis. Metagenomic coverage of samples for *E. coli* and *K. pneumoniae* was determined from bwa-mem read mappings using genomeCoverageBed⁶⁰ (v2.25.0). Samples with too low relative abundance for confident identification (<0.1%) were designated as ‘not detected’, while samples with low median read coverage (<8) were designated as ‘low coverage’. Of the remaining samples, those with >90% of the SNVs at or above an allele frequency of 0.9 were designated as ‘one strain’, exhibiting a unimodal distribution as is classically expected in the single haplotype setting¹³ (Supplementary Fig. 5a). A *k*-means clustering approach (based on allele frequency values <0.98, *k*=2) was used on other samples to identify ‘two strains’ (silhouette score >0.8, indicating good concordance with two clusters for a bimodal distribution) and ‘multiple strains’ cases where there may be more than two clusters (Supplementary Fig. 5a). Note that this analysis was only used to determine strain ‘states’ (Fig. 2), and the corresponding clusters were not used for downstream haplotype analysis. To confirm metagenomic SNV calling quality and strain designations, ‘one strain’ cases were compared with SNVs from corresponding isolates (where available) and noted to have high precision for both *E. coli* and *K. pneumoniae* (>98%; Supplementary Fig. 5b). A first-order Markov model of the transition frequencies between the strain compositions was estimated using the markovchain package in R (ref. ⁶¹) (maximum likelihood estimator with Laplace smoothing parameter = 1).

Substrain analysis. SNVs with mean allele frequency >0.9 across timepoints were identified as probably fixed across all strains in a sample. Non-fixed SNVs from ‘one strain’ cases were further annotated for their impact on protein function using SnpEff⁶² (v4.3). The ratio of the rate of non-synonymous (dN) to synonymous (dS) mutations was calculated using the package Biopython.codonalign.codonseq with the ‘NG86’ method.

Leveraging the availability of multiple ‘one strain’ timepoints in some individuals, non-fixed SNV trajectories were clustered to identify co-varying SNVs that may belong to a common substrain background. Specifically, the DBSCAN algorithm in R (ref. ⁶³) was used to cluster SNV trajectories in selected individuals with multiple ‘one strain’ timepoints ($\epsilon = 0.2$, $\text{minPts} = 2n$ as recommended), and identified clusters were visualized as a sanity check.

Plasmid analysis. A Mash screen search approach was used with PLSDb⁶⁴ to obtain a list of plasmids that are potentially present in the CPE isolate genomes. The union of all such plasmid sequences was then aligned with isolate genome assemblies to identify plasmid hits with >85% coverage at >95% identity (only alignments >500 bp). Plasmid hits were clustered into groups using hierarchical clustering at 95% identity (hclust function in R, average linkage based on Mash distance⁶⁵), with the longest plasmid serving as a representative. Only plasmids longer than 10 kbp are included in the figure to avoid spurious/redundant matches to shorter plasmids.

Plasmid conjugation assay. Donor *E. coli* harbouring the pKPC2 plasmid with a kanamycin selection cassette (MG1655) and recipient *K. pneumoniae* strains (ATCC13883) were streaked on selective LBA and incubated overnight at 37°C. Bacterial colonies were resuspended in LB (1 ml), diluted to OD₆₀₀ 0.5 and mixed in a 1:1 ratio and spotted onto 0.22 µm nitrocellulose membrane (Sartorius) placed on top of LBA (20 µl). After 4 hours of incubation at 37°C, the bacterial mixture was resuspended in 2 ml of PBS, serially diluted and plated on LBA with appropriate antibiotic selection. Kanamycin (50 µg ml⁻¹) and fosfomycin (40 µg ml⁻¹) were used for selection of transconjugants. Plates were incubated at 37°C overnight, and colonies were enumerated. Conjugation frequency was calculated as the total number of transconjugants per total number of recipients.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Isolate and shotgun metagenomic sequencing data is available from the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena/browser/home>) under project accession number PRJEB49334. Source data are provided with this paper.

Code availability

Source code for scripts used to analyse the data are available in a GitHub project at <https://github.com/CSB5/CPE-microbiome>.

Received: 9 September 2021; Accepted: 29 July 2022;

Published online: 15 September 2022

References

1. von Wintersdorff, C. J. et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front. Microbiol.* **7**, 173 (2016).
2. Suay-García, B. & Pérez-Gracia, M. T. Present and future of carbapenem-resistant Enterobacteriaceae (CRE) infections. *Antibiotics* **8**, 122 (2019).
3. Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* **13**, 42–51 (2015).
4. Codjoe, F. S. & Donkor, E. S. Carbapenem resistance: a review. *Med Sci.* **6**, 1 (2017).
5. Schechner, V. et al. Asymptomatic rectal carriage of blaKPC producing carbapenem-resistant Enterobacteriaceae: who is prone to become clinically infected? *Clin. Microbiol. Infect.* **19**, 451–456 (2013).
6. Penders, J., Stobberingh, E. E., Savelkoul, P. H. & Wolfs, P. F. The human microbiome as a reservoir of antimicrobial resistance. *Front. Microbiol.* **4**, 87 (2013).
7. Nordmann, P., Naas, T. & Poirel, L. Global spread of Carbapenemase-producing Enterobacteriaceae. *Emerg. Infect. Dis.* **17**, 1791–1798 (2011).
8. Tooke, C. L. et al. β-Lactamases and β-lactamase inhibitors in the 21st century. *J. Mol. Biol.* **431**, 3472–3500 (2019).
9. Sun, X. et al. Microbiota-derived metabolic factors reduce campylobacteriosis in mice. *Gastroenterology* **154**, 1751–1763.e2 (2018).
10. Ichinohe, T. et al. Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proc. Natl Acad. Sci. USA* **108**, 5354–5359 (2011).
11. Buffie, C. G. et al. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517**, 205–208 (2015).
12. Lieberman, T. D. et al. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat. Genet.* **43**, 1275–1280 (2011).
13. Garud, N. R., Good, B. H., Hallatschek, O. & Pollard, K. S. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. *PLoS Biol.* **17**, e3000102 (2019).
14. Chu, N. D., Smith, M. B., Perrotta, A. R., Kassam, Z. & Alm, E. J. Profiling living bacteria informs preparation of fecal microbiota transplantations. *PLoS ONE* **12**, e0170922 (2017).
15. Ferreira, A., Crook, N., Gasparrini, A. J. & Dantas, G. Multiscale evolutionary dynamics of host-associated microbiomes. *Cell* **172**, 1216–1227 (2018).
16. Mo, Y. et al. Duration of carbapenemase-producing Enterobacteriaceae carriage in hospital patients. *Emerg. Infect. Dis.* **26**, 2182–2185 (2020).
17. Haverkate, M. R. et al. Duration of colonization with *Klebsiella pneumoniae* carbapenemase-producing bacteria at long-term acute care hospitals in Chicago, Illinois. *Open Forum Infect. Dis.* **3**, ofw178 (2016).
18. Korach-Rechtman, H. et al. Intestinal dysbiosis in carriers of carbapenem-resistant Enterobacteriaceae. *mSphere* **5**, e00173–20 (2020).
19. Yoshida, N. et al. *Bacteroides vulgatus* and *Bacteroides dorei* reduce gut microbial lipopolysaccharide production and inhibit atherosclerosis. *Circulation* **138**, 2486–2498 (2018).
20. Lenoir, M. et al. Butyrate mediates anti-inflammatory effects of. *Gut Microbes* **12**, 1–16 (2020).
21. Riedel, C. U. et al. Anti-inflammatory effects of bifidobacteria by inhibition of LPS-induced NF-κB activation. *World J. Gastroenterol.* **12**, 3729–3735 (2006).
22. Zeng, M. Y., Inohara, N. & Núñez, G. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol.* **10**, 18–26 (2017).
23. Winter, S. E. & Bäuml, A. J. A breathtaking feat: to compete with the gut microbiota, *Salmonella* drives its host to provide a respiratory electron acceptor. *Gut Microbes* **2**, 58–60 (2011).
24. Rivera-Chávez, F., Lopez, C. A. & Bäuml, A. J. Oxygen as a driver of gut dysbiosis. *Free Radic. Biol. Med.* **105**, 93–101 (2017).
25. Chng, K. R. et al. Metagenome-wide association analysis identifies microbial determinants of post-antibiotic ecological recovery in the gut. *Nat. Ecol. Evol.* **4**, 1256–1267 (2020).
26. Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).
27. Stacy, A. et al. Infection trains the host for microbiota-enhanced resistance to pathogens. *Cell* **184**, 615–627.e17 (2021).
28. Barreto, H. C., Sousa, A. & Gordo, I. The landscape of adaptive evolution of a gut commensal bacteria in aging mice. *Curr. Biol.* **30**, 1102–1109.e5 (2020).
29. Ernst, C. M. et al. Adaptive evolution of virulence and persistence in carbapenem-resistant *Klebsiella pneumoniae*. *Nat. Med.* **26**, 705–711 (2020).
30. Zhao, S. et al. Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* **25**, 656–667.e8 (2019).
31. Warsi, O. M., Andersson, D. I. & Dykhuizen, D. E. Different adaptive strategies in *E. coli* populations evolving under macronutrient limitation and metal ion limitation. *BMC Evol. Biol.* **18**, 72 (2018).
32. Hickman, R. A., Munck, C. & Sommer, M. O. A. Time-resolved tracking of mutations reveals diverse allele dynamics during *Escherichia coli* antimicrobial adaptive evolution to single drugs and drug pairs. *Front. Microbiol.* **8**, 893 (2017).

33. Auriol, C., Bestel-Corre, G., Claude, J. B., Soucaille, P. & Meynial-Salles, I. Stress-induced evolution of *Escherichia coli* points to original concepts in respiratory cofactor selectivity. *Proc. Natl Acad. Sci. USA* **108**, 1278–1283 (2011).
34. Juers, D. H., Matthews, B. W. & Huber, R. E. LacZ β -galactosidase: structure and function of an enzyme of historical and molecular biological importance. *Protein Sci.* **21**, 1792–1807 (2012).
35. Rogers, A. W. L., Tsolis, R. M. & Bäumlner, A. J. *Salmonella* versus the microbiome. *Microbiol. Mol. Biol. Rev.* **85**, e00027–19 (2021).
36. Hughes, E. R. et al. Microbial respiration and formate oxidation as metabolic signatures of inflammation-associated dysbiosis. *Cell Host Microbe* **21**, 208–219 (2017).
37. Gupta, S., Allen-Vercoe, E. & Petrof, E. O. Fecal microbiota transplantation: in perspective. *Ther. Adv. Gastroenterol.* **9**, 229–239 (2016).
38. Wortelboer, K., Nieuwoudorp, M. & Herrema, H. Fecal microbiota transplantation beyond *Clostridioides difficile* infections. *EBioMedicine* **44**, 716–729 (2019).
39. Lee, S. M. et al. Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501**, 426–429 (2013).
40. Martinson, J. N. V. et al. Rethinking gut microbiome residency and the Enterobacteriaceae in healthy human adults. *ISME J.* **13**, 2306–2318 (2019).
41. Woyke, T., Doud, D. F. R. & Schulz, F. The trajectory of microbial single-cell sequencing. *Nat. Methods* **14**, 1045–1054 (2017).
42. Domingo, E. & Perales, C. Viral quaspecies. *PLoS Genet.* **15**, e1008271 (2019).
43. Yamada, C. et al. Molecular insight into evolution of symbiosis between breast-fed infants and a member of the human gut microbiome *Bifidobacterium longum*. *Cell Chem. Biol.* **24**, 515–524.e5 (2017).
44. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
45. Gao, S., Bertrand, D., Chia, B. K. & Nagarajan, N. OPERA-LG: efficient and exact scaffolding of large, repeat-rich eukaryotic genomes with performance guarantees. *Genome Biol.* **17**, 102 (2016).
46. Gao, S., Bertrand, D. & Nagarajan, N. FinIS: improved in silico finishing using an exact quadratic programming formulation. *Lect. Notes Comput. Sci.* **7534**, 314–325 (2012).
47. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997v2 (2013).
48. Segata, N. et al. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814 (2012).
49. Franzosa, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).
50. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
51. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
52. Hawinkel, S., Mattiello, F., Bijmans, L. & Thas, O. A broken promise: microbiome differential abundance methods do not control the false discovery rate. *Brief. Bioinformatics* **20**, 210–221 (2019).
53. Morton, J. T. et al. Establishing microbial composition measurement standards with reference frames. *Nat. Commun.* **10**, 2719 (2019).
54. Inouye, M. et al. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6**, 90 (2014).
55. Alcock, B. P. et al. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, D517–D525 (2020).
56. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
57. Wilm, A. et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **40**, 11189–11201 (2012).
58. Hinrichs, A. S. et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
59. Pracana, R., Priyam, A., Levantis, I., Nichols, R. A. & Wurm, Y. The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. *Mol. Ecol.* **26**, 2864–2879 (2017).
60. Quinlan, A. R. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–34 (2014).
61. Spedicato, G. Discrete time Markov chains with R. *R J.* **9.2**, 84 (2017).
62. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
63. Hahsler, M., Piekenbrock, M. & Doran, D. dbscan: fast density-based clustering with R. *J. Stat. Softw.* **91**, 1–30 (2019).
64. Galata, V., Fehlmann, T., Backes, C. & Keller, A. PLSDb: a resource of complete bacterial plasmids. *Nucleic Acids Res.* **47**, D195–D202 (2019).
65. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
66. Quan, S. et al. Adaptive evolution of the lactose utilization network in experimentally evolved populations of *Escherichia coli*. *PLoS Genet.* **8**, e1002444 (2012).
67. Tsuchido, T., VanBogelen, R. A. & Neidhardt, F. C. Heat shock response in *Escherichia coli* influences cell division. *Proc. Natl Acad. Sci. USA* **83**, 6959–6963 (1986).
68. Trubetskoy, D., Proux, F., Allemand, F., Dreyfus, M. & Iost, I. SrmB, a DEAD-box helicase involved in *Escherichia coli* ribosome assembly, is specifically targeted to 23S rRNA in vivo. *Nucleic Acids Res.* **37**, 6540–6549 (2009).
69. Garoff, L., Huseby, D. L., Praski Alzrigat, L. & Hughes, D. Effect of aminoacyl-tRNA synthetase mutations on susceptibility to ciprofloxacin in *Escherichia coli*. *J. Antimicrob. Chemother.* **73**, 3285–3292 (2018).
70. Aponte, R. A., Zimmermann, S. & Reinstein, J. Directed evolution of the DnaK chaperone: mutations in the lid domain result in enhanced chaperone activity. *J. Mol. Biol.* **399**, 154–167 (2010).
71. Mundhada, H. et al. Increased production of L-serine in *Escherichia coli* through adaptive laboratory evolution. *Metab. Eng.* **39**, 141–150 (2017).
72. Conrad, T. M. et al. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proc. Natl Acad. Sci. USA* **107**, 20500–20505 (2010).
73. Li, Y. et al. LPS remodeling is an evolved survival strategy for bacteria. *Proc. Natl Acad. Sci. USA* **109**, 8716–8721 (2012).

Acknowledgements

This work was supported by grants from Biomedical Research Council Industry Alignment Fund (IAF311018) to N.N., S.L.C., K.M. and N.O.T., National Medical Research Clinician Scientist Individual Research Grant (CIRG18nov-0034) to K.M., National Medical Research Council Collaborative Grant (NMRC CGAug16C005) to K.M. and N.O.T. and National Medical Research Council Clinician Scientist Award (NMRC/CSA-INV/0002/2016 and MOH-CSAINV18nov-0004) to K.M.

Author contributions

N.N., S.L.C., K.M. and N.O.T. conceived and designed the study. K.M. and N.O.T. oversaw the clinical work. A.N. conducted metagenomic library preparation and sequencing under C.K.R.'s supervision. S.L.C. conducted the isolate sequence analysis. J.T.L.K., J.T., D.B. and A.R. coordinated bioinformatic analysis of the metagenomic data under N.N.'s supervision. M.Y. conducted plasmid transfer experiments under Y-H.G.'s supervision. J.T.L.K., J.T., D.B., A.N., A.R., C.K.R. and N.N. wrote the manuscript with inputs from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41564-022-01221-w>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41564-022-01221-w>.

Correspondence and requests for materials should be addressed to Niranjan Nagarajan.

Peer review information *Nature Microbiology* thanks Sean Gibbons and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

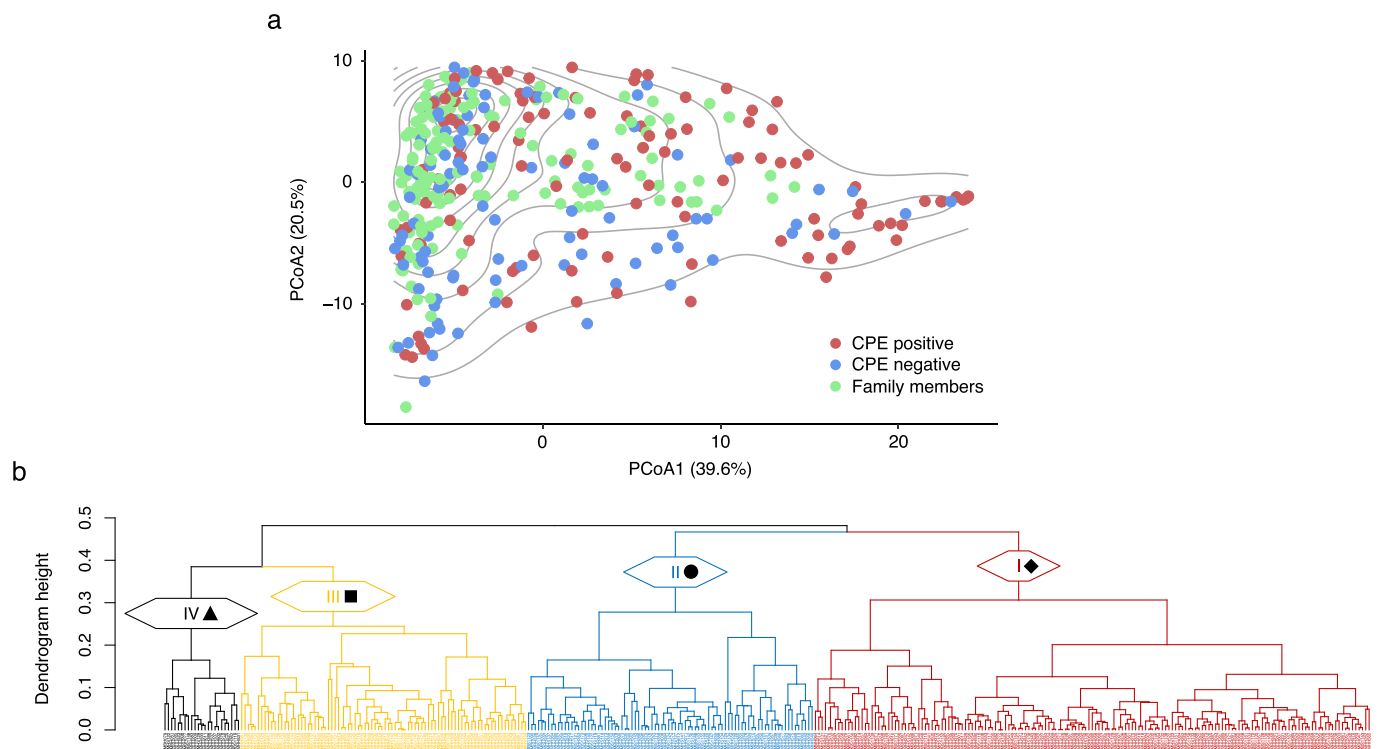
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

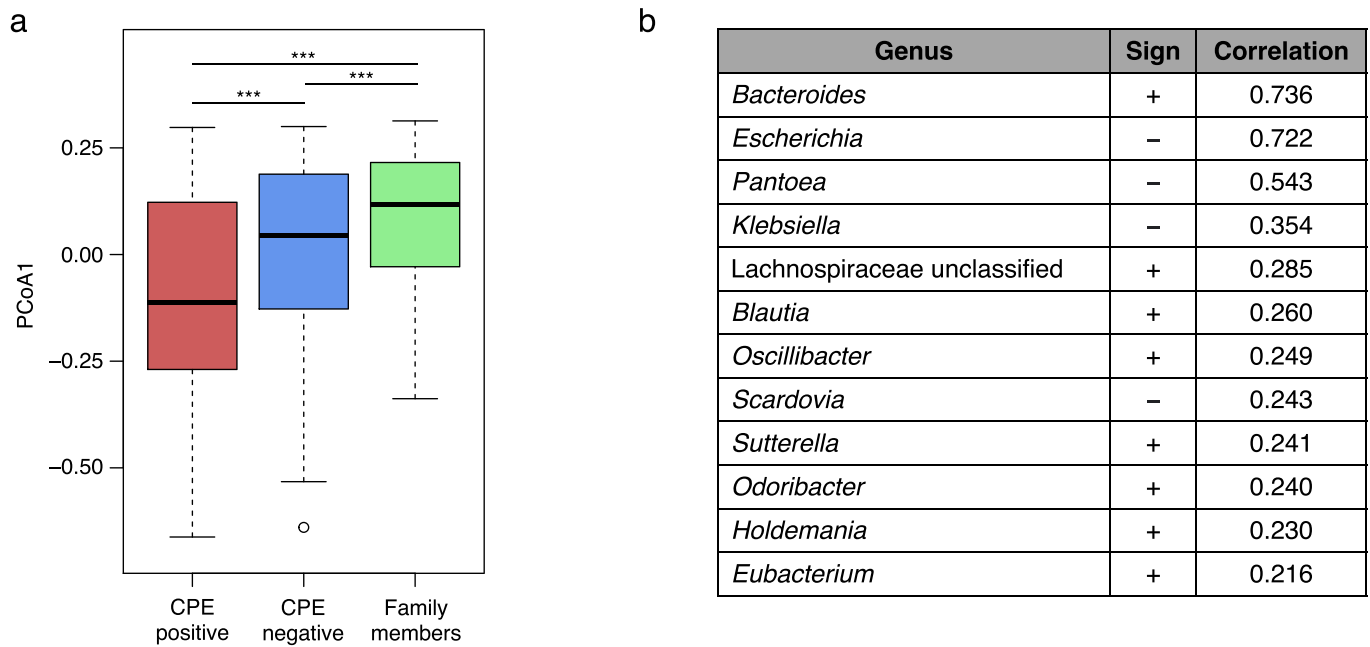


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

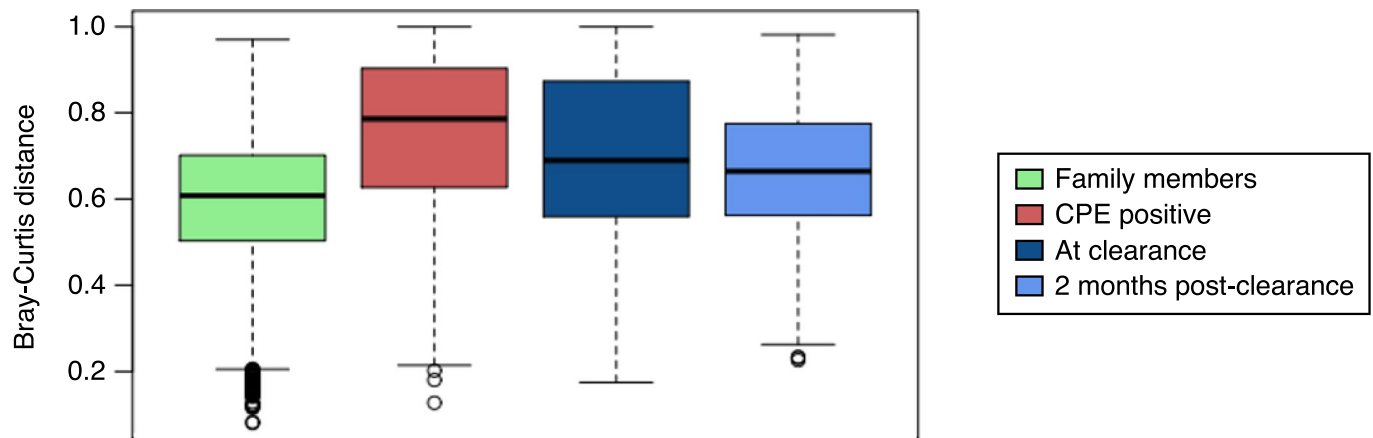
© The Author(s) 2022



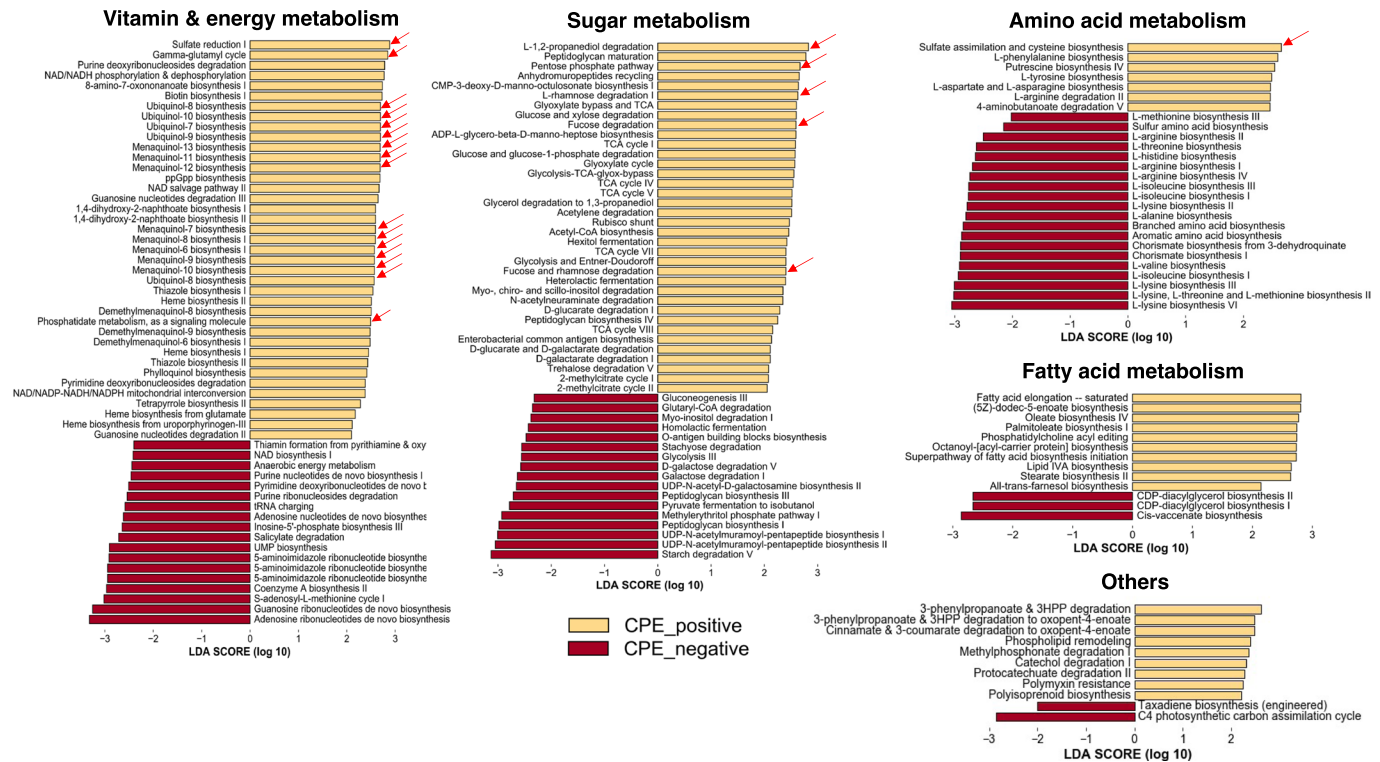
Extended Data Fig. 1 | Additional principal coordinates analysis. (a) Principal coordinates analysis (PCoA) plot similar to Fig. 1a but calculated based on genus-level weighted UniFrac distance. Note that while the location of points is flipped with respect to the x-axis for Fig. 1a, the general distribution of points is similar with regions of high density on the left having more family members while a less dense cluster on the right has more CPE positive points. **(b)** Dendrogram for hierarchical clustering based on distance in PCoA1-PCoA2 space in Fig. 1a, and number of clusters determined by AIC. The labels I-IV correspond to those in Fig. 1a. Dendrogram tips are labeled with the library ID of each sample (see Supplementary File 1).



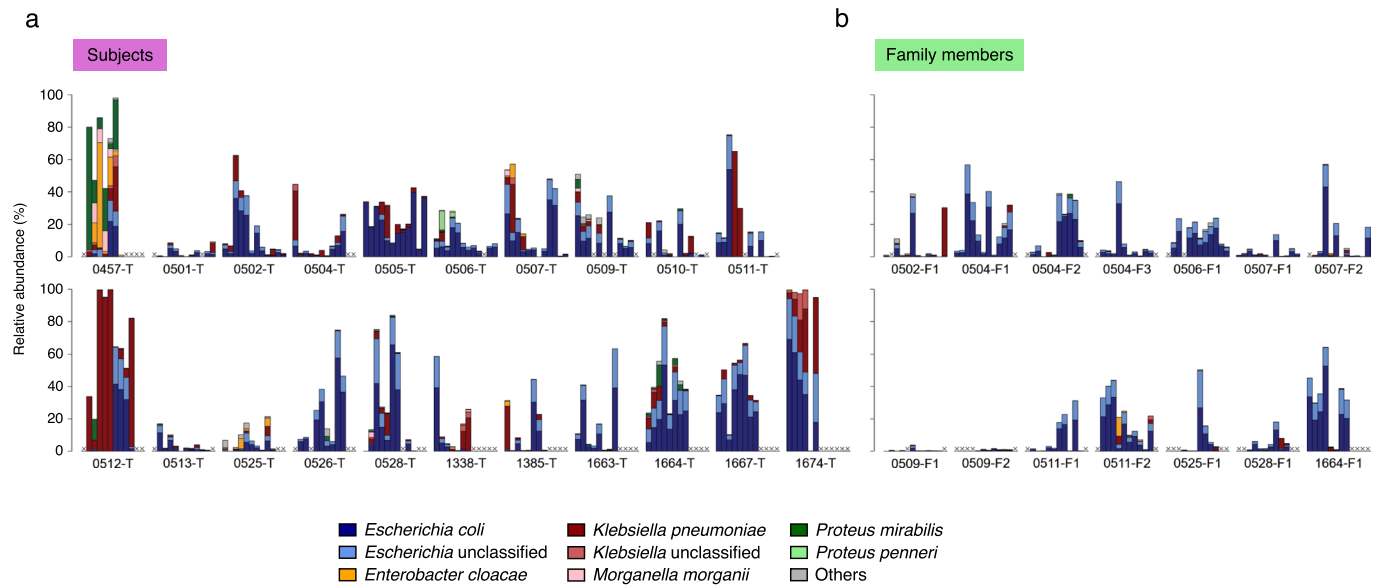
Extended Data Fig. 2 | Associations with the first principal coordinate. (a) Comparison of principal coordinates analysis (PCoA) values (Fig. 1a) across different groups of samples based on the first principal coordinate (PCoA1). Wilcoxon rank-sum $p < 0.01$ for all pairwise comparisons ($n = 362$ samples; two-sided Wilcoxon rank-sum test p -values - CPE positive vs. CPE negative: 5.14×10^{-4} , CPE negative vs. family members: 0.00694, CPE positive vs. family members: 7.48×10^{-11}). Centre lines in the boxplots represent median values, box limits represent upper and lower quartile values, and whiskers represent 1.5 times the interquartile range above the upper quartile and below the lower quartile. (b) Genera most correlated with PCoA1 (Fig. 1a). Sign (-): Negative correlation with PCoA1 (most abundant in configuration IV samples). Sign (+): Positive correlation with PCoA1 (least abundant in configuration IV samples).



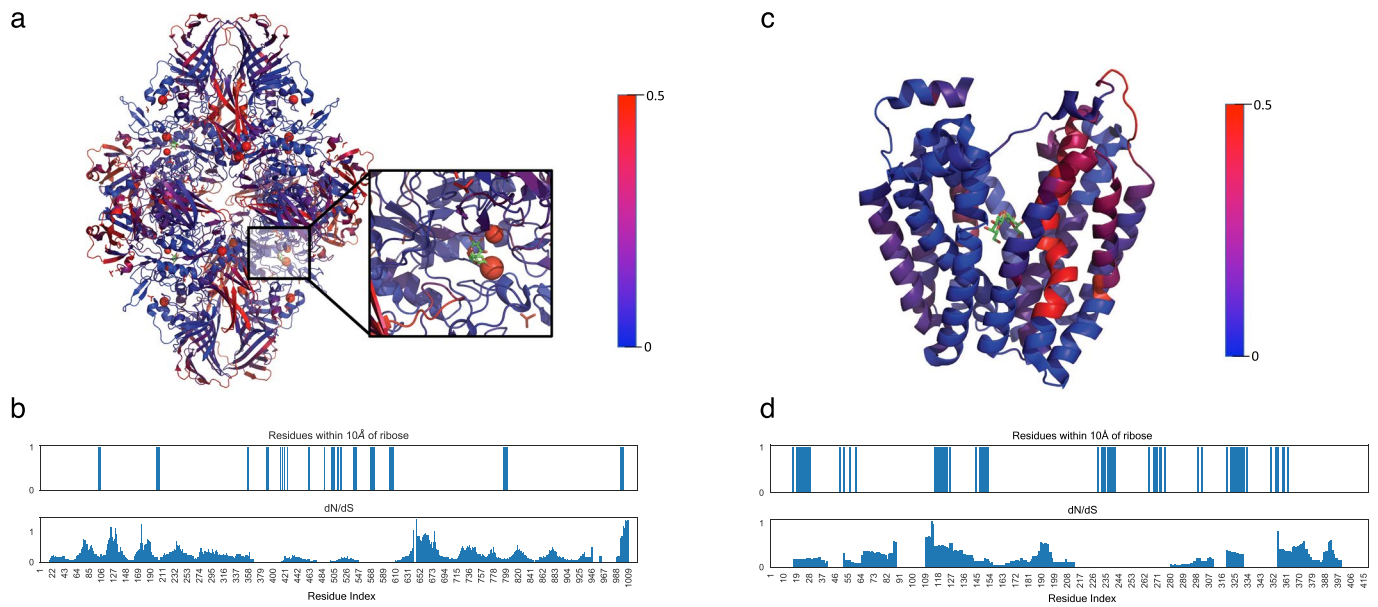
Extended Data Fig. 3 | Pairwise genus-level Bray-Curtis distances with family members. Boxplots quantifying the distributions of pairwise genus-level Bray-Curtis distances. Each pair consists of one sample from a family member, and a second sample from one of the following groups: family members, CPE positive, CPE negative at point of clearance, as well as more than two months post-clearance. Wilcoxon rank-sum $p < 0.01$ for all pairwise comparisons with the family member group. ($n = 39,420$ sample pairs; two-sided Wilcoxon rank-sum test p -values of comparison with the family member group - CPE positive: $< 2.2 \times 10^{-16}$; At clearance: $< 2.2 \times 10^{-16}$; 2 months post-clearance: $< 2.2 \times 10^{-16}$). Centre lines in the boxplots represent median values, box limits represent upper and lower quartile values, and whiskers represent 1.5 times the interquartile range above the upper quartile and below the lower quartile.



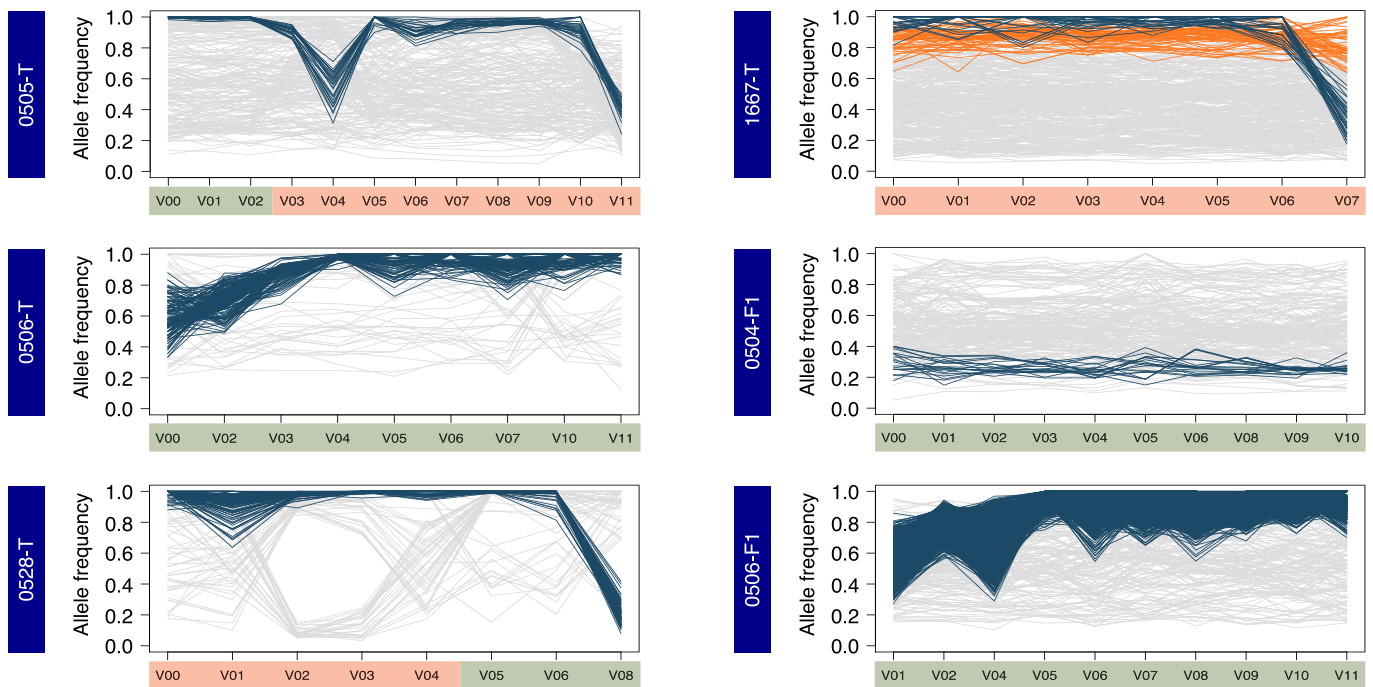
Extended Data Fig. 4 | Pathway comparisons in colonized and decolonized gut metagenomes. Differentially abundant pathways in colonized (CPE positive) and post-decolonization (CPE negative) gut metagenomes (two-sided Wilcoxon rank-sum test FDR-adjusted p -value < 0.05 , LDA score > 2). Pathways highlighted with an arrow are those that have potential associations with gut inflammation and aerobic respiration.



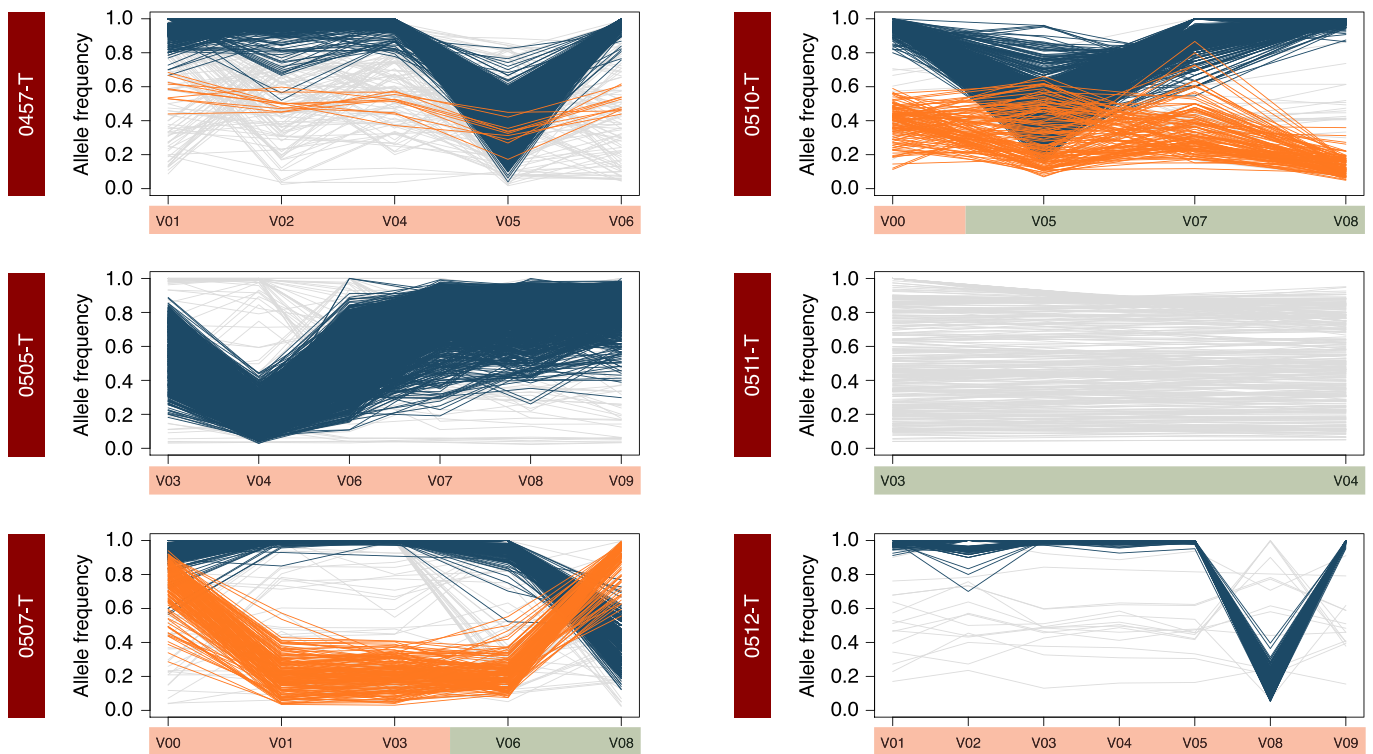
Extended Data Fig. 5 | Enterobacteriaceae relative abundances. Bar plots showing the relative abundances of various Enterobacteriaceae species in (a) subjects and (b) family members, over the course of sample collection. Each sub-panel consists of 12 bars, representing consecutively visit 0 (V0) to visit 11 (V11). An 'X' at a particular position indicates that a sample corresponding to that visit number was not collected. Only individuals with samples collected at six or more visits are shown.



Extended Data Fig. 6 | dN/dS values in *lacZ* and *lacY*. (a) Visualization of dN/dS rates for the *lacZ* tetramer (4dux) using a 30aa window based on metagenomics-derived SNVs (Table 1). The inset shows a close-up view of *lacZ*'s sugar binding active site. In general, dN/dS rates were lower for *lacZ* in regions close to the active site (two-sided Wilcoxon rank-sum test p -value $< 2.6 \times 10^{-9}$) except for around position 1000. (b) Plot of dN/dS values and regions of *lacZ* that are close to ribose. (c) Visualization of dN/dS rates for *lacY* (1pv7) using a 30aa window. Interestingly, *lacY* exhibited a distinct trend with a few regions of higher dN/dS rates close to the active site (two-sided Wilcoxon rank-sum test < 0.029). (d) Plot of dN/dS values and regions close to the *lacY* lactose symporter channel.



Extended Data Fig. 7 | Sub-strain variations in *E. coli*. Sub-strain variations over time for other individuals (*E. coli*). Selected figures were plotted for subjects with >3 consecutive single-strain timepoints. Coloured lines depict clusters of SNVs that were identified. Timepoints are shaded based on whether they were determined to be CPE colonized (red) or not (green). Note that while a cluster of SNVs typically tends to be dominant in a subject across timepoints (coloured in dark blue, except for 0504-F1), they can drop in frequency at some timepoints, and this does not necessarily coincide with the CPE decolonization event as was observed for 1674-T (for example, V08 for 0528-T).



Extended Data Fig. 8 | Sub-strain variations in *K. pneumoniae*. Sub-strain variations over time for other individuals (*K. pneumoniae*). Figures were plotted for subjects with ≥ 2 single-strain timepoints. Coloured lines depict clusters of SNVs that were identified. Timepoints are shaded based on whether they were determined to be CPE colonized (red) or not (green). Note that similar to the observation in Extended Data Fig. 7, the dominant sub-strain in a subject can switch at timepoints that do not coincide with change in CPE colonization status (for example, V01 and V08 for 0507-T).

Extended Data Table 1 | Cohort and sample statistics

| | Number of individuals | Total number of samples | Mean number of samples per individual |
|-----------------------|------------------------------|--------------------------------|--|
| Index subjects | 29 | 216 | 7.5 |
| Family members | 17 | 145 | 8.5 |
| Total | 46 | 361 | 7.9 |

Stool samples were collected at recruitment (V00), weekly for 4 weeks (V01-V04), monthly for 5 months (V05-V09) and bimonthly for 6 months (V10-V12) when provided by study participants.

Extended Data Table 2 | SNVs whose population frequencies varied notably over time recurrently

| Gene | Protein | SNV position (no. of indiv.) | Ref |
|-----------------------------|--|--|-------------------|
| <i>E. coli</i> | | | |
| <i>lacZ</i> | Beta-D-galactosidase | Cys77Arg (2), Val86Ile (2), Thr109Ala (2), Phe114Tyr (2), Ala117Thr (3), Ser133Cys (2), Leu350Val (2), Ser656Thr (2), Ser656Leu (2), Thr746Met (2), Asp753Gly (3), Ala821Val (2) | Ref ⁶⁶ |
| <i>pf1B</i> | Pyruvate formate lyase I | Thr16Ala (3), Asn383Ser (3) | |
| <i>lpp</i> | Murein lipoprotein | Ala3Pro (2), Lys2Asn (3) | Ref ⁶⁷ |
| <i>pnp</i> | Polynucleotide phosphorylase/polyadenylase | Ile122Val (3) | Ref ³¹ |
| <i>rnpA</i> | Protein C5 component of RNase P | Lys91Arg (3) | |
| <i>rpsE</i> | 30S ribosomal subunit protein S5 | His121Arg (3) | |
| <i>pheT</i> | Phenylalanine tRNA synthetase, beta subunit | Glu699Asp (3) | |
| <i>rsxC</i> | Putative 4Fe-4S ferredoxin-type protein fused with unknown protein | Val712Ile (3) | |
| <i>srnB</i> | ATP-dependent RNA helicase | Ser14Asn (3) | Ref ⁶⁸ |
| <i>thrS</i> | Threonyl-tRNA synthetase | Lys638Gln (3) | Ref ⁶⁹ |
| <i>yfgA</i> | Conserved hypothetical protein; putative HTH-type transcriptional regulator | Gln332Pro (3) | |
| <i>nuoE</i> | NADH:ubiquinone oxidoreductase, chain E | Met26Lys (3) | Ref ³³ |
| <i>nlpl</i> | Lipoprotein precursor | Thr57Ser (3) | |
| <i>K. pneumoniae</i> | | | |
| <i>atpB</i> | F0 sector of membrane-bound ATP synthase, subunit a | Ile236Val (3) | |
| <i>dnaK</i> | Chaperone Hsp70, co-chaperone with DnaJ | Ala449Ser (3) | Ref ⁷⁰ |
| <i>eno</i> | Enolase | Glu132Ala (3) | Ref ⁷¹ |
| <i>gapA</i> | Glyceraldehyde-3-phosphate dehydrogenase A | Ile66Val (3) | |
| <i>ho1C</i> | DNA polymerase III, chi subunit | Val136Met (3) | |
| <i>fadB</i> | 3-hydroxyacyl-CoA dehydrogenase/3-hydroxybutyryl-CoA epimerase/delta(3)-cis-delta(2)-trans-enoyl-CoA isomerase/enoyl-CoA hydratase | Ala725Asp (2) | |
| <i>srnB</i> | ATP-dependent RNA helicase | Asn14Ser (2) | Ref ⁶⁸ |
| KPHS_52800 | Ribonuclease P protein component | Arg9Lys (2) | |
| KPHS_48690 | Hypothetical protein | Leu28Met (2) | |
| <i>rpoA</i> | DNA-directed RNA polymerase subunit alpha | Ala157Thr (2) | Ref ⁷² |
| <i>rp1M</i> | 50S ribosomal protein L13 | Glu86Gln (2) | |
| KPHS_47160 | Hypothetical protein | Val10Ile (2) | |
| <i>pnp</i> | Polynucleotide phosphorylase/polyadenylase | Val99Ile (2) | Ref ³¹ |
| <i>nlpl</i> | Lipoprotein | Ser57Thr (2) | |
| <i>deaD</i> | ATP-dependent RNA helicase | Thr22Asn (2) | |
| <i>rodZ</i> | Cytoskeleton protein | Ser327Pro (2) | |
| <i>lpxD</i> | UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase | Leu9Phe (2) | Ref ⁷³ |
| KPHS_36220 | Hypothetical protein | Val20Met (2) | |
| <i>pheT</i> | Phenylalanyl-tRNA synthetase subunit beta | Asp699Glu (2) | |
| KPHS_27510 | Hypothetical protein | Ser29Tyr (2) | |
| <i>gapDH</i> | Glyceraldehyde-3-phosphate dehydrogenase | Val66Ile (2) | |
| <i>loiD</i> | Lipoprotein-releasing system ATP-binding protein | Asn19Ser (2) | |
| <i>mukE</i> | Condesin subunit E | Leu3Ser (2) | |
| KPHS_15690 | Alpha-ketoglutarate decarboxylase | Asp935Glu (2) | |
| KPHS_15680 | Succinate dehydrogenase iron-sulfur subunit | Lys2Arg (2) | |

Genes containing non-synonymous SNVs that exhibit a large (>0.3) change in allele frequency across pairs of one strain time points in an individual, observed across at least two individuals⁶⁶⁻⁷³.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted <i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Isolate and shotgun metagenomic sequencing data is available from the European Nucleotide Archive (ENA – <https://www.ebi.ac.uk/ena/browser/home>) under project accession number PRJEB49334.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Use the terms *sex* (biological attribute) and *gender* (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Population characteristics

Carriers of carbapenemase-producing Enterobacteriaceae (CPE), and their non-colonized family members as controls. Covariates include gender, age, and ethnicity, but were not the primary focus of our research and findings.

Recruitment

Study participants were recruited from two tertiary care centres in Singapore, based on their status as carriers of carbapenemase-producing Enterobacteriaceae (CPE). We do not identify any self-selection bias in our participant pool.

Ethics oversight

Ethics approval received from the Singapore National Healthcare Group Domain Specific Review Board 74 (NHG DSRB Reference: 2016/00364) prior to commencement.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size was determined based on the number of recruited study participants. As stool samples were collected from participants at multiple time points, the eventual number of samples (361) was sufficient for our analyses.

Data exclusions

No data was excluded from the analyses.

Replication

Our data set consists of multiple CPE-colonized participants and their family members, and our findings are generalizable across our participant group. As this is a cohort study, there is no experimentally-based replication.

Randomization

Randomization is not relevant as participants were not assigned to experimental groups.

Blinding

Blinding is not relevant as this study does not involve different treatment groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

| n/a | Involved in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |