
















# Evolutionary and functional history of the *Escherichia coli* K1 capsule

Received: 28 January 2023

Accepted: 26 May 2023

Published online: 15 June 2023

 Check for updates

Sergio Arredondo-Alonso<sup>1,2</sup>, George Blundell-Hunter <sup>3,20</sup>, Zuyi Fu<sup>4,20</sup>, Rebecca A. Gladstone<sup>1,2,20</sup>, Alfred Fillol-Salom<sup>4</sup>, Jessica Loraine<sup>3</sup>, Elaine Cloutman-Green<sup>5</sup>, Pål J. Johnsen <sup>6</sup>, Ørjan Samuelsen <sup>6,7</sup>, Anna K. Pöntinen <sup>1,7</sup>, François Cléon<sup>6</sup>, Susana Chavez-Bueno <sup>8,9</sup>, Miguel A. De la Cruz <sup>10,11</sup>, Miguel A. Ares<sup>10</sup>, Manivanh Vongsouvath<sup>12</sup>, Agnieszka Chmielarczyk<sup>13</sup>, Carolyne Horner<sup>14</sup>, Nigel Klein <sup>5</sup>, Alan McNally <sup>15</sup>, Joice N. Reis<sup>16,17</sup>, José R. Penadés <sup>4</sup>, Nicholas R. Thomson <sup>2,18</sup>, Jukka Corander <sup>1,2,19</sup> , Peter W. Taylor<sup>3</sup>  & Alex J. McCarthy <sup>4</sup> 

*Escherichia coli* is a leading cause of invasive bacterial infections in humans. Capsule polysaccharide has an important role in bacterial pathogenesis, and the K1 capsule has been firmly established as one of the most potent capsule types in *E. coli* through its association with severe infections. However, little is known about its distribution, evolution and functions across the *E. coli* phylogeny, which is fundamental to elucidating its role in the expansion of successful lineages. Using systematic surveys of invasive *E. coli* isolates, we show that the K1-*cps* locus is present in a quarter of bloodstream infection isolates and has emerged in at least four different extraintestinal pathogenic *E. coli* (ExPEC) phylogroups independently in the last 500 years. Phenotypic assessment demonstrates that K1 capsule synthesis enhances *E. coli* survival in human serum independent of genetic background, and that therapeutic targeting of the K1 capsule re-sensitizes *E. coli* from distinct genetic backgrounds to human serum. Our study highlights that assessing the evolutionary and functional properties of bacterial virulence factors at population levels is important to better monitor and predict the emergence of virulent clones, and to also inform therapies and preventive medicine to effectively control bacterial infections whilst significantly lowering antibiotic usage.

Bacteria across a wide range of phyla produce capsular polysaccharides that are associated with diverse experimentally validated functions and have been shown to improve bacterial persistence and adaptation to new environments<sup>1,2</sup>. Human bacterial pathogens use these capsules as major virulence determinants to promote colonization and persistence in the gastrointestinal, respiratory, and urogenital tracts, as well as other tissues<sup>3–5</sup>. Capsular polysaccharides are often identical or similar in structure and chemical composition to polysaccharides found in human tissues, thereby providing

non-immunogenic coatings for bacteria to survive in human tissues and to cause disease<sup>5</sup>. In addition, capsules can reduce the efficiency of antimicrobial peptides and complement<sup>6–9</sup>, suppress phagocytosis by innate immune cells and promote intracellular survival<sup>10–17</sup>, and contribute to defence against antimicrobial agents<sup>8,18</sup>. Although *E. coli* can produce around 80 distinct capsular chemotypes that are organised into four major groups<sup>19</sup>, only a subset of these chemically distinct capsular types are associated with the capacity to cause invasive extraintestinal diseases; such infections include bloodstream

infections (BSI), pyelonephritis and meningitis<sup>20</sup>. In particular, the polysialic acid-containing K1 capsule<sup>21,22</sup>, chondroitin-containing K4 capsule<sup>23</sup> and heparosan-containing K5 capsule<sup>24,25</sup> are associated with the extraintestinal pathogenic *E. coli* (ExPEC) clones linked to such invasive diseases<sup>26,27</sup>, likely through the propensity of these capsule type to mimic polysaccharides present on cells in the human tissues<sup>5</sup>. However, the epidemiology of the capsular types remains largely unexplored due to the absence of serological typing data or specific methods that can computationally predict *E. coli* capsular types based on whole-genome sequencing data. Our current understanding of the evolution and functional properties of the distinct capsular polysaccharides in the global *E. coli* population is therefore limited and mainly based on pre-genomic studies.

The K1 capsule polysaccharide has repeatedly been linked to BSI, neonatal meningitis and pyelonephritis<sup>5,28–31</sup>. The K1 capsule is a homopolymer of  $\alpha$ -2,8-linked N-acetylneuraminic acid (sialic acid; NeuNAc) termed polysialic acid (polySia) that mimics the polySia modification found on human neuronal and immune cells<sup>32–35</sup> and likely promotes the capacity of K1-expressing *E. coli* to hide and reside within the blood and neuronal compartments. Indeed, polySia prevents full activation of innate host defences and confers resistance to complement- and phagocyte-mediated killing<sup>36–38</sup>. In agreement with epidemiological links of K1 capsule to invasive human infections, experimental animal models using isogenic strains have revealed that K1 expression promotes stable gastrointestinal (GI) tract colonisation and promotes the development of invasive systemic infections by *E. coli*<sup>39–42</sup>. In contrast to the association of K1 with BSI and meningitis, other well-studied K antigens including types K2, K4 and K5 are mostly associated with UTIs<sup>5,6,43,44</sup>. Despite the association of K1 encapsulated *E. coli* with BSI and meningitis, and the fact that polySia can be utilized as a powerful tool for diagnosis and therapeutic targeting<sup>45,46</sup>, we lack basic knowledge on the prevalence, evolution and functional properties of the K1 capsule at the population level. This lack of knowledge limits our capacity to develop efficient strategies to combat *E. coli* infections.

In this study, we elucidated the prevalence, distribution, and evolution of the K1 capsule in *E. coli* populations by considering a global dataset of 5065 clinical isolates. The position and synteny of the K1 capsule (K1-*cps*) locus in the genomes were resolved by considering K1 complete genomes. Using a Bayesian inference approach, we estimated the introduction date of K1-*cps* in the most predominant extraintestinal lineages. To show that the K1 capsule was functional among the distinct lineages, we performed several phenotypic assays to show that the K1 capsule was expressed and conferred *E. coli* with immune resistance independent of its genetic background. Our results show that a quarter of BSIs are caused by *E. coli* carrying this the K1-*cps* locus, and this is driven by multiple introductions of K1-*cps* into the ExPEC pathotype. For the first time, we estimated the introduction times of the capsule among the main ExPEC lineages and dated that the K1-*cps* locus was acquired at least 500 years ago. In support of the role of K1 capsule in virulence, we show that the enzymatic removal of the K1 capsule renders *E. coli* susceptible to complement-replete human serum, suggesting that the therapeutic use of capsule depolymerases is likely to be a promising approach for the prevention and treatment of these infections.

## Results

### A quarter of the ExPEC population has the K1-*cps* locus

Despite the association of K1-*cps* to the ExPEC pathotype since the 1980s<sup>21,47–49</sup>, and the availability of whole-genome sequencing data for ExPEC isolates, the epidemiology of the K1 capsular type has remained largely unexplored in the post-genomic era. The K1 capsule biosynthesis locus belongs to group 2 capsules<sup>26</sup>, and is composed of eight genes in two conserved regions (regions 1 and 3) shared between all

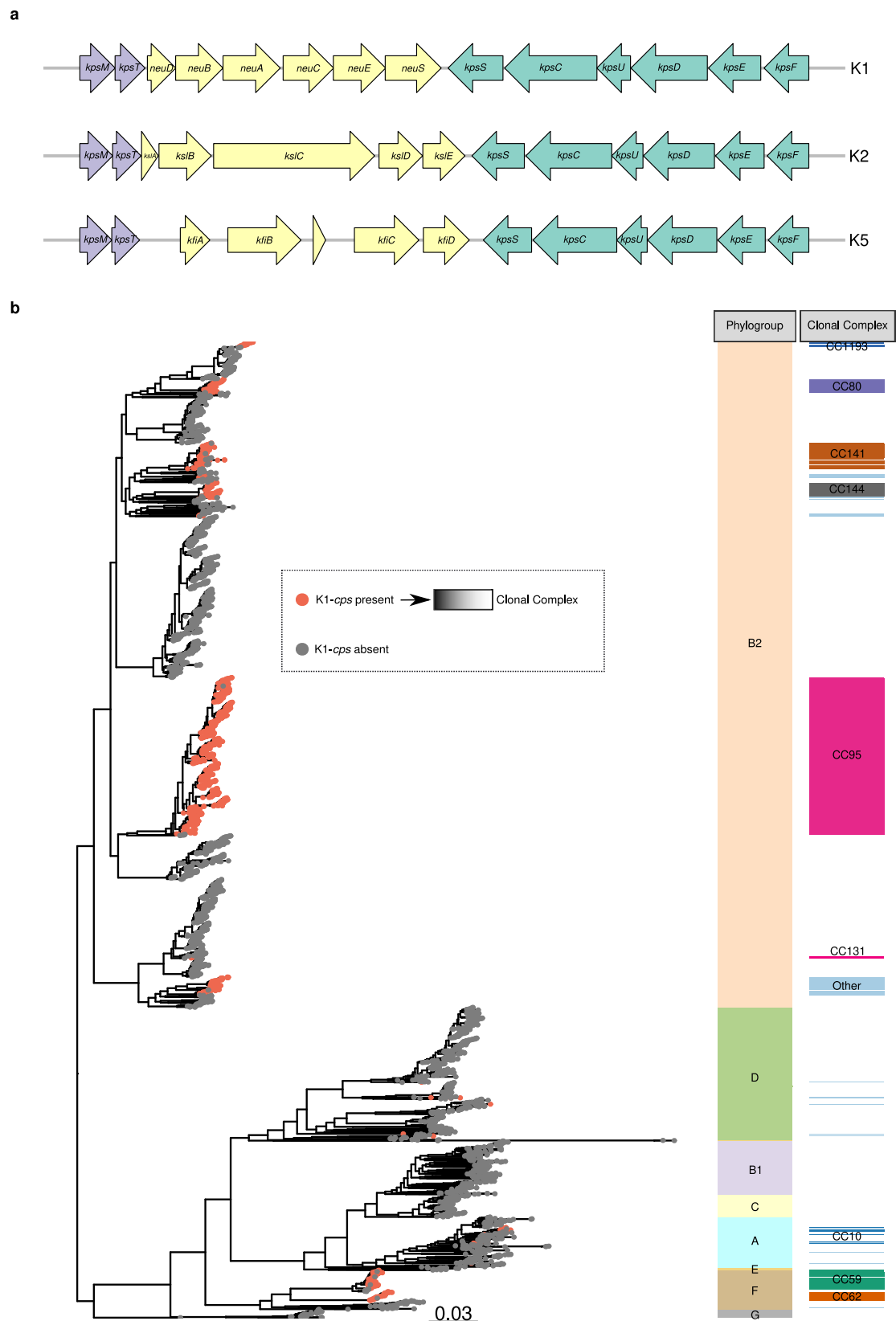
group 2 capsule types and an additional 6 genes (region 2) unique to the K1-*cps* locus (Fig. 1a).

To estimate the prevalence of the K1-*cps* locus among ExPEC isolates, we assessed two unbiased longitudinal studies, NORM<sup>50</sup> and BSAC<sup>51</sup>, that characterized BSIs in Norway ( $n = 3254$ ) and United Kingdom ( $n = 1509$ ), respectively. As the BSI isolates in both studies were collected regardless of their clonal background, antimicrobial resistance profile or other bacterial phenotypic or genotypic characteristics, the NORM and BSAC collections offer a representative survey of BSI clones that have circulated in comparable host populations during the timespan of the studies. These datasets therefore provide a valuable platform to estimate the K1-*cps* prevalence in *E. coli* BSI populations. The population prevalence of the K1-*cps* locus among BSI isolates was estimated to be 24.0% and 22.9% for the NORM and BSAC collections, respectively (Table 1). The finding that a similar proportion of BSI isolates are positive for K1-*cps* in the two independent unbiased longitudinal studies provides strong evidence that the K1 is linked with *E. coli* propensity to cause BSIs.

Group 2 capsules, including K1, are classically assumed to be expressed in *E. coli* isolates causing extraintestinal infections, but not in *E. coli* causing diarrhoeal diseases<sup>26</sup>. To clarify whether the K1-*cps* is associated with any other *E. coli* pathotypes, we analysed the Horesh et al. collection that consists of a comprehensive, high-quality and pathotype-defined collection of *E. coli* genomes<sup>52</sup>. We specifically screened for the presence of the K1-*cps* locus in the 5,236 diarrhoeagenic isolates from the Horesh et al. collection that includes the pathotypes (i) enteropathogenic *E. coli* (EPEC), (ii) enterotoxigenic *E. coli* (ETEC), (iii) enterohaemorrhagic *E. coli* (EHEC), (iv) enteroaggregative *E. coli* (EAEC), (v) enteroinvasive *E. coli* (EIEC), (vi) diffusely adherent *E. coli* (DAEC) and (vii) adherent invasive *E. coli* (AIEC). We observed that only 0.1% (5/5236) of the diarrhoeagenic isolates carried the K1-*cps* locus, therefore discarding a role of the K1 capsule in diarrhoeagenic diseases.

Given that ~25% of ExPEC isolates possess the K1-*cps* locus, it is important to analyze the clonality of these isolates to understand the population structure and to define how ExPEC clones with enhanced virulence properties emerge and evolve. We, therefore, examined the distribution of the K1-*cps* locus across the distinct ExPEC phylogroups (phylogroups A, B1, B2, C, D, E and F). For both the NORM and BSAC datasets, the K1-*cps* was frequent in phylogroup B2 (31.7% and 29.4%, respectively) and phylogroup F (50.4% and 23.2%, respectively), but also observed in phylogroups A (9.3% and 6.1%, respectively) and D (0.6% and 1.4%, respectively). We did not detect K1-*cps* in phylogroup B1, C or E. The distribution of the K1-*cps* locus among clonal complexes (CC, defined here as belonging to a lineage named after the dominant sequence type) was almost identical for both NORM and BSAC collections (Table 1). Within phylogroup B2, the K1-*cps* was mostly present in CC95 followed by CC141, CC144, and CC80 (Table 1). In addition, we observed the capsule in a few isolates of the pan-susceptible clade B of CC131, but not in the multidrug-resistant (MDR) C1 and C2 clades of CC131. The K1-*cps* locus was also present in the CC59 and CC62 within phylogroup F and in CC10 within phylogroup A (Table 1). This data indicates that 22–24% of all BSI isolates are not caused by a single successful K1-*cps*+ ExPEC clone, but by multiple distinct K1-*cps*+ clones.

To analyze the distribution and evolution of the K1-*cps* across a global collection of ExPEC isolates, we expanded the NORM and BSAC studies by incorporating isolates from other countries, time periods and sources that were genotypically or phenotypically detected as K1-*cps*+ or K1+, respectively. This batch of samples included: (i) newly generated adult and neonatal samples phenotypically detected as K1+,  $n = 201$ , from six countries (Brazil, Laos, Mexico, Poland, United Kingdom, United States), (ii) samples carrying the K1-*cps* locus from the pre-antibiotic era (from 1932 onwards) that form part of the Murray collection,  $n = 15$ <sup>53</sup> and (iii) samples carrying the K1-*cps* locus from the



Horesh collection,  $n = 86^{52}$ . We collated all genome isolates (5065) and generated a maximum-likelihood phylogeny. A total of 1427 isolates possessed the K1-*cps* locus (Fig. 1b). The phylogeny overlaid with the metadata (also available at Dataset S1) can be interactively queried in the following Microreact project <https://microreact.org/project/cm8f8adPyWoQqemAAGvrmi-k1-context>. For isolates in which the

complete K1-*cps* locus, from *kpsM* to *kpsF* (Fig. 1a), was assembled in one contig (1409/1427, 98.7%), the product of *kpsT* showed the most conserved protein sequence with 98.7% (1391/1409) of isolates carrying an identical amino acid sequence (Supplementary Fig. 1). In contrast, *neuE*, the most diverse gene in the locus, displayed a total of 76 distinct variants (Supplementary Fig. 1).

**Fig. 1 | K1-*cps* genomic organisation and distribution in the *E. coli* global dataset.** **a** Genetic organisation of the group 2 capsule locus in *E. coli* encoding K1, K2 and K5 antigens. Regions 1 and 3 (in green and purple, respectively) are responsible for capsule export and are shared among K antigens belonging to group 2 capsules. Region 2 (in yellow) is serotype-specific and involved in capsule biosynthesis. The coding sequences (CDs) are represented as arrows in the diagram and are labelled according to their gene names. **b** Maximum-likelihood phylogenetic tree of 5065 *E. coli* clinical isolates. The phylogeny was reconstructed by mapping all isolates against the *Escherichia coli* strain EC958 (accession HG941718; <https://www.ncbi.nlm.nih.gov/nuccore/HG941718.1>), using the fast and GTR model

provided in iqtree. The scale bar corresponds to the expected number of nucleotide substitutions per site. In the phylogeny, nodes were coloured, based on the absence (grey tip) or presence (red tip) of the K1-*cps* locus. Next to the phylogeny, the metadata block, Phylogroup, indicates the phylogroup of all isolates regardless of possessing or not the K1 capsule of the isolates bearing the K1-*cps* gene cluster. The metadata block, Clonal Complex, indicates the clonal complex of the isolates possessing the K1-*cps* locus in their genome. The phylogeny and metadata can be interactively queried using the following Microreact link <https://microreact.org/project/cm8f8adPyWoQqemAAgvrmi-k1-context>. Source data are provided as a Source Data file.

### The dominant ExPEC lineage CC95 acquired the K1-*cps* locus over 250 years ago

CC95 is one of the most dominant ExPEC clones associated with community-onset and nosocomial infections worldwide<sup>20</sup>, and the K1-*cps* locus is ubiquitous in invasive isolates of this lineage (99.8–100% of CC95 isolates are K1-*cps*+; Table 1). We hypothesized that the successful expansion and establishment of the K1-*cps*+ CC95 clone in the *E. coli* population has been driven by a single acquisition of the K1-*cps* locus and clonal expansion, rather than multiple independent K1-*cps* acquisitions by CC95. We further investigated the genetic context and evolution of the K1-*cps* in the lineage by obtaining a dated phylogeny of CC95 NORM genomes to determine the most common recent ancestor (tMRCA) for K1-*cps* within CC95, and by unambiguously characterizing the position and gene synteny of the K1-*cps* locus in the *E. coli* genome by retrieving all CC95 genomes with an associated RefSeq complete genome ( $n = 44$ ). Nearly all NORM genomes (441/442) from CC95 (Table 1) possessed the K1-*cps* suggesting that this locus was introduced by its most common recent ancestor (MRCA).

To determine the origin of CC95 (tMRCA), we used BactDating that makes use of a Markov Chain Monte Carlo (MCMC) model to perform Bayesian inference and produce a dating phylogeny. We estimated the origin of CC95 (tMRCA) and thus estimated the introduction of the K1-*cps* locus to be approximately around the year 1768 [95% CI, 1721–1806] (Fig. 2a). In Dataset S2, we confirmed the MCMC convergence of the model as shown by (i) the effective sample sizes of the parameters and (ii) the Gelman and Rubin's convergence diagnostic. In addition, we show a significant temporal signal by comparing the resulting model against a model with equal sampled dates using the deviance information criteria (DIC) (Dataset S2). Thus, the high-frequency of K1-*cps* in CC95 is readily explained by a single acquisition event that occurred approximately over 250 years ago that subsequently spread worldwide as a single clone.

The K1-*cps* locus was present next to the tRNA-*pheV* site, which is localized around the position -800,000 of the chromosome (matching

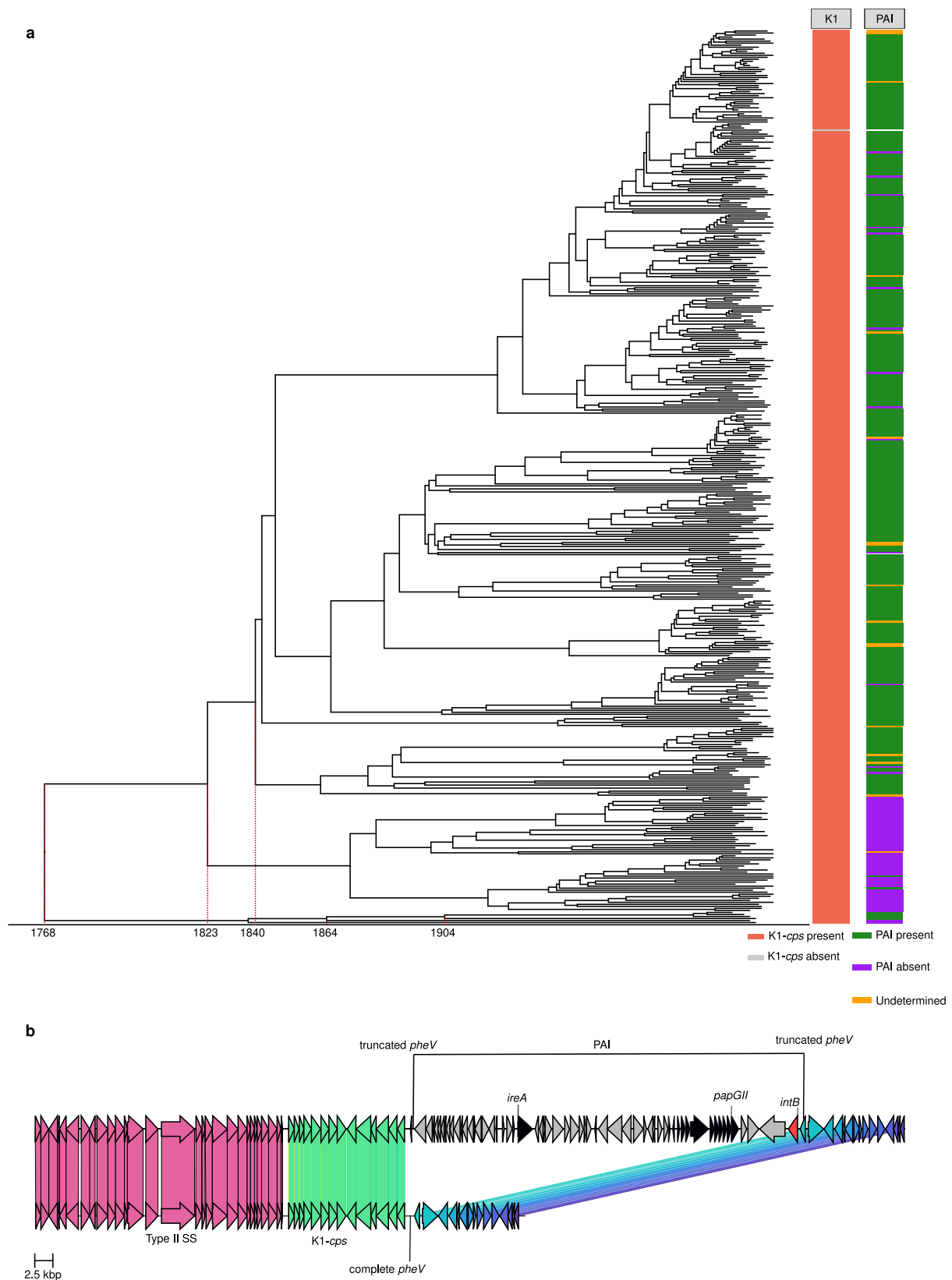
the position described in pre-genome literature as 67 min in the *E. coli* genome<sup>54,55</sup>) as revealed by analysis of the RefSeq CC95 complete genomes (Dataset S3). Upstream of the K1-*cps* locus, gene synteny is highly conserved and is characterized by the presence of a type II secretion system<sup>56</sup>. Downstream of the locus, two predominant genome configurations were defined across CC95 (i) the insertion of a pathogenicity island (PAI) that has resulted in the truncation of the tRNA *pheV* gene (26/44, 59%), and (ii) the absence of the PAI and an intact *pheV* gene (16/44, 36%), while 2 genomes (2/44, 5%) showed each one a distinct gene synteny as determined by Panaroo. In the NORM CC95 isolates, we could detect the PAI in 80.27% (354/441) of the CC95 isolates based on the truncation of the *pheV* and presence of the *intB* gene, while 15.87% (70/441) of isolates did not carry the PAI, and in 3.85% of the cases (17/441) we could not determine the presence/absence of the PAI due to fragmented short-read assemblies (Fig. 2a).

This PAI had an average size of 52.4 kbp and was previously termed PAI-V (Fig. 2b)<sup>57,58</sup>. Because PAIs carry one or more accessory genes that encode virulence factors that often function as adhesins, iron-acquisition systems, host defence mechanisms or toxins<sup>59</sup>, the acquisition and loss of this PAIs is likely to impact phenotypes and infectivity. We did not observe any trend in the association between the presence/absence of the PAI and the isolation source (BSI vs UPEC) of these 44 complete genomes (Dataset S3), indicating that acquisition or loss of this PAI is not exclusively linked to either type of invasive infection. The mosaic structure and flexible pool of virulence genes carried by PAIs could mean that a certain virulence gene is associated with the PAI in the CC95 lineage. However, a comparison of the virulence genes of the isolates that lost the PAI with the closest PAI+ isolate in the phylogeny indicated the main difference resided in the presence and absence of *papGII* locus encoding the P fimbriae and an iron-acquisition (*ireA*) locus. Interestingly, these genes have been recently identified as key features distinguishing invasive from non-invasive UPEC<sup>60</sup>. Additional genes that have roles in promoting bacterial survival or virulence were not identified on the PAI in CC95 isolates.

Based on the CC95 dated phylogeny (Fig. 2a), we determined that the MRCA did not possess the PAI and that it is likely that there were two introductions of a PAI downstream of the *kpsF* gene in this lineage. The PAI was estimated to have been introduced between 1823 [95% CI, 1789–1851] and 1840 [95% CI, 1809–1864], and then subsequently has become fixed in the genome and predominated through clonal expansion. Interestingly, we estimated a minimum of 14 excision events of the PAI linked to recombination between the two direct short repeats at *pheV* (22 bp)<sup>61</sup> that resemble *att sites* in a process likely mediated by the *intB* P4-like integrase gene<sup>58,62</sup>. This suggests that the PAI can be excised and lost from the *E. coli* genome. In support, this PAI can form a circular excision product that may be exported by larger mobile genetic elements (MGEs)<sup>61</sup>. We also observed a minor and independent introduction of an island downstream of the *kpsF* gene estimated to happen between 1864 [95% CI, 1832–1892] and 1904 [95% CI, 1878–1927] (Fig. 2b). Collectively, this data indicates that the introduction of the K1-*cps* locus in CC95 precedes the acquisition of the PAI, but that accessory genes encoded on the PAI could additionally contribute to the success of CC95 as an ExPEC clone.

**Table 1 | Prevalence of the K1-*cps* locus in the NORM and BSAC collections**

Clonal Complex (CC)	Phylogroup	K1- <i>cps</i> prevalence per CC in the NORM collection	K1- <i>cps</i> prevalence per CC in the BSAC collection
All	-	780/3254 (24.0%)	345/1509 (22.9%)
CC95	B2	441/442 (99.8%)	189/189 (100%)
CC141	B2	78/89 (87.6%)	21/24 (87.5%)
CC59	F	48/58 (82.8%)	16/22 (72.7%)
CC144	B2	38/38 (100%)	23/23 (100%)
CC80	B2	41/41 (100%)	22/22 (100%)
CC1193	B2	21/25 (84.0%)	0/1 (0%)
CC62	F	19/19 (100%)	19/20 (95.0%)
CC10	A	12/67 (17.9%)	7/40 (17.5%)
CC131	B2	4/203 (2.0%)	1/202 (0.5%)
Other	-	78/2272 (3.4%)	47/966 (4.9%)



### Multiple *K1-cps* locus acquisitions across the major ExPEC lineages

Given that the *K1-cps* + CC95 clone emerged over 250 years ago and is globally successful, we hypothesized that the CC95 population is likely to have acted as the donor of the *K1-cps* locus during the emergence of other *K1-cps* + ExPEC clones in other phylogroups. For each *K1-cps* +

lineage, we performed a hybrid assembly to obtain a complete chromosomal sequence (Dataset S4) that allowed us to infer the gene synteny present downstream of the *K1-cps* locus and shown in Fig. 4.

We first analysed the acquisition of the *K1-cps* locus in other phylogroups (CC59 and CC62 in phylogroup F; CC10 in phylogroup A). Surprisingly, the *K1-cps* locus was introduced into CC59 by its MRCA

**Fig. 2 | Downstream gene synteny of the K1-*cps* locus in CC95 complete genomes.** **a** Dated phylogeny of the CC95 lineage considering as reference genome the RefSeq assembly accession GCF\_008632595.1. The K1-*cps* presence (red) and absence (grey) is mapped as a metadata block. Carriage of the pathogenicity island (PAI) is mapped as a metadata block; present (green), absent (purple) or unknown (orange). The ticks on the x-axis of the phylogeny indicate: (i) the time of the most recent common ancestor (tMRCA) for the lineage (1768) [95% CI, 1721–1806], (ii) the time interval of the primary PAI acquisition, between 1823 [95% CI, 1789–1851] and 1840 [95% CI, 1809–1864], and (iii) the time interval of a second PAI acquisition, dated between 1864 [95% CI, 1832–1892] and 1904 [95% CI, 1878–1927]. **b** Genetic

organisation of the two most predominant synteny blocks observed downstream of the *kpsF* gene for the complete CC95 chromosomal sequences illustrated with the RefSeq assemblies GCA\_002285755.1 (top block) and GCA\_001280405.1 (bottom block). The top block illustrates the presence of the PAI downstream of the *kpsF* gene which resulted in the truncation of the tRNA *pheV* gene. The P4-like integrase *intB* present in the PAI is coloured red. The virulence genes present in the PAI, *ireA* and the *papGII* operon, are coloured black. The bottom block illustrates the second most predominant synteny block represented by an intact tRNA *pheV* gene and the absence of a PAI downstream the K1-*cps* locus. Source data are provided as a Source Data file.

around the year 1525 [95% CI, 1044–1730] (Fig. 3, Dataset S2). The BactDating model inferring the tMRCA of CC59 showed MCMC convergence of the parameters and a significant temporal signal (Dataset S2). The median number of SNP differences and the SNP profile of the *neu* region (Supplementary Fig. 2) together with the conserved gene synteny observed downstream of the *kpsF* gene (Fig. 4) indicated that CC59 does not share the same K1-*cps* locus with CC95. Instead, CC59 shared the same K1-*cps* locus with CC62, and shared a common ancestor with CC62 within the *E. coli* wide-species tree (Fig. 1). This suggests that the MRCA of these lineages in phylogroup F already carried the K1-*cps* locus. Correspondingly, the data conclusively refute that CC95 was the original donor of the K1-*cps* locus for CC59/CC62, and vice versa, that CC59/CC62 was the original donor(s) of the K1-*cps* locus for CC95.

In CC10 of phylogroup A, K1-*cps* was acquired in two independent events, estimated to have happened between the years: (i) from 1789 [95% CI, 1624–1865] to 1821 [95% CI, 1677–1890] and (ii) from 1853 [95% CI, 1738–1907] to 1873 [95% CI, 1772–1922], respectively (Fig. 3). The BactDating model predicting the date of these events showed MCMC convergence of the parameters and a significant temporal signal (Dataset S2). For one of these events, CC59/CC62 acted as a K1-*cps* donor based on the number and profile of SNP differences (Supplementary Fig. 2), and conserved gene synteny (Fig. 4). Thus, this analysis revealed that CC95 did not act as the donor of K1-*cps* for phylogroup F (CC59, CC62) or phylogroup A (CC10), but instead confirmed the transfer of the K1-*cps* locus occurred from phylogroup F to phylogroup A.

Next, we analysed the acquisition of K1-*cps* by the non-CC95 clones within phylogroup B2. We observed that *E. coli* belonging to CC59/CC62 from phylogroup F acted as K1-*cps* donors for CC1193 from phylogroup B2. We also observed the exchange of the K1-*cps* among lineages from phylogroup B2: (i) between CC95 and CC144, and (ii) between CC141 and CC80. However, we could not estimate the tMRCA of these additional lineages and this prevented us from concluding the directionality of exchange. The analysis further demonstrated that a sub-lineage of the recently globally disseminated lineage CC131 acquired the K1-*cps* locus in a single acquisition estimated to have happened between the years from 1957 [95% CI, 1946–1967] to 1999 [95% CI, 1995–2002] (Fig. 3). In Dataset S2, we show that the CC131 dating phylogeny obtained from BactDating to infer the time of the acquisition showed MCMC convergence of the parameters and a significant temporal signal (Dataset S2). The number of SNP differences with respect to other lineages suggested this lineage has acquired K1-*cps* from a distinct source outside of our collection of sequences (Supplementary Fig. 2). In summary, the emergence of K1-encapsulated ExPEC clones has not been driven by acquisition of the K1-*cps* locus from the globally successful CC95 lineage, despite the high prevalence and close association of this clone with the K1 capsule.

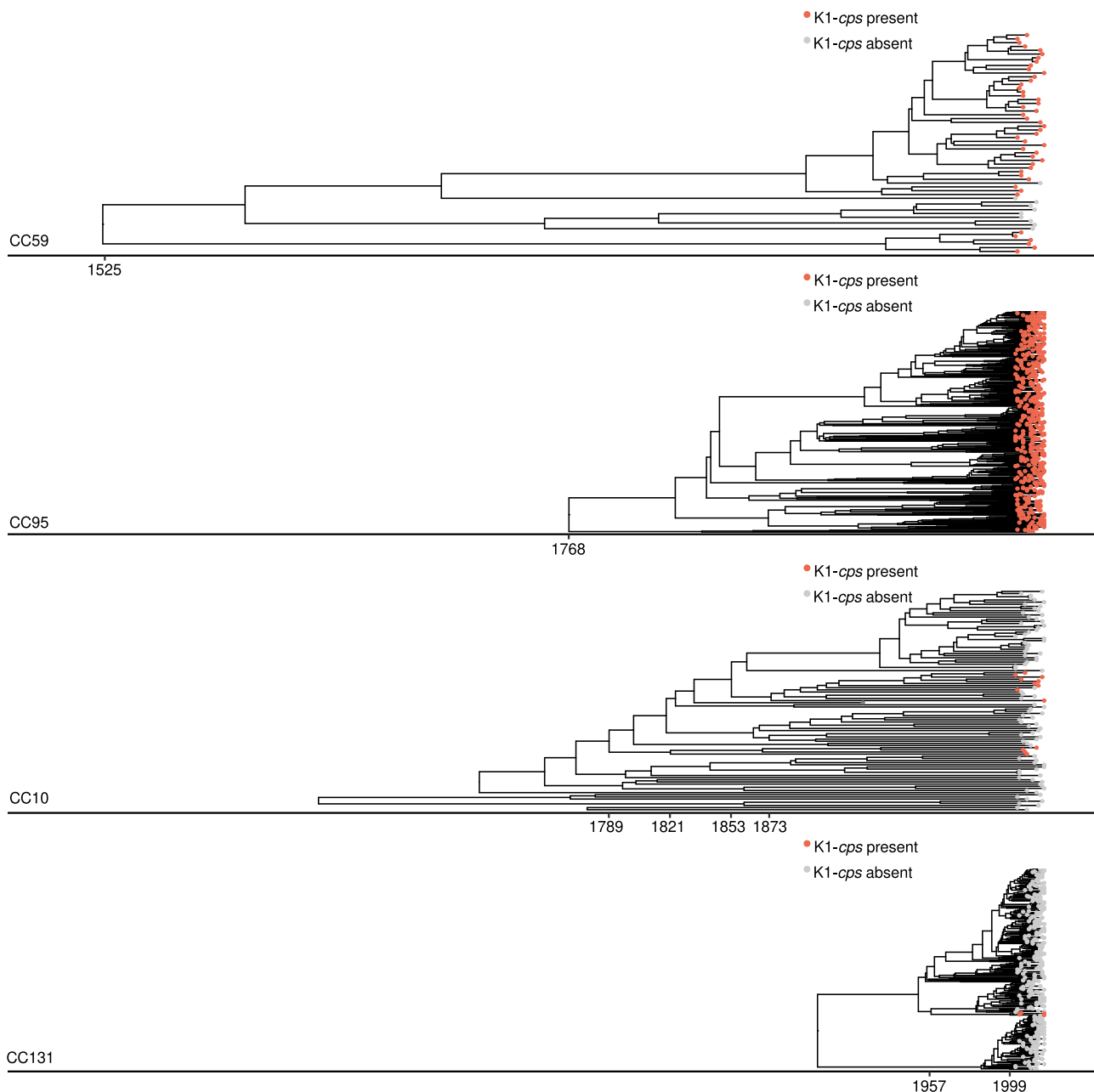
Of note, isolates from lineages in phylogroup A (CC10), phylogroup B2 (CC80, CC131, CC141, CC144, CC1193) and phylogroup F (CC59 and CC62) also frequently had a PAI inserted downstream of the K1-*cps* locus (Fig. 4). Despite high frequency in the PAIs from CC95, the gene encoding *IreA* was only additionally present in PAI

from CC131, and the gene encoding PapGII was only additionally present in the PAI from lineage CC141. Analysis of gene content of the PAI from isolates of other lineages demonstrates a heterogeneous pool of virulence genes. This includes genes encoding characterized virulence factors  $\alpha$ -hemolysin toxin (HlyA)<sup>63</sup>, cytotoxic necrotizing factor (CNF1)<sup>64</sup>, secreted autotransporter toxin (Sat)<sup>65</sup>, S-fimbrial adhesin (Sfa)<sup>66</sup>, the IrgA homolog adhesin (Ila) homolog adhesin<sup>67</sup> and aerobactin<sup>68</sup>. More comprehensive analysis is required to elucidate the function of PAI-encoded virulence genes as molecular determinants of BSIs.

### K1 capsule expression across diverse ExPEC lineages confers resistance to serum-mediated killing

The K1 capsule is strongly expressed by *E. coli* in blood and is thought to be essential for infection by enhancing resistance to complement-mediated killing<sup>36,69,70</sup>, but functional studies have traditionally been performed exclusively in the CC95 genetic background. To test the hypothesis that K1 capsule expression alters the virulence phenotype of *E. coli* independently of the genetic background, we compared the susceptibility of isogenic K1 capsule positive and negative *E. coli* strains to complement C3b deposition and serum killing. We found that inactivation of the *neuC* gene in the K1-*cps* locus enhanced the susceptibility of *E. coli* from three different genetically distinct backgrounds (CC95, CC349 and CC62 from phylogroups B2, D and F, respectively) to complement C3b deposition after incubation in human active serum (Fig. 5a; Supplementary Fig. 3a). Further, inactivation of *neuC* promoted the killing of *E. coli* during incubation in human serum (Fig. 5b; Supplementary Fig. 3a). This data demonstrates that K1 capsule synthesis, independent of genetic background, is an important determinant of *E. coli* survival in human serum, and consequently contributes to virulence.

To functionally validate that the K1 capsule is properly expressed across distinct genetic lineages, and that this expression results in an altered virulence phenotype of *E. coli*, we evaluated polySia K1 capsule production and the immune evasion properties of 149 *E. coli* isolates that possessed the K1-*cps* locus. We selected isolates that were representative of four phylogroups ( $n=20$  phylogroup A isolates,  $n=120$  phylogroup B2 isolates,  $n=3$  phylogroup D isolates and  $n=6$  phylogroup F isolates). To measure K1 expression we tested the binding of recombinant (r)GFP-labelled EndoNA2 (rEndoNA2-GFP) (Supplementary Fig. 3b), a catalytically inactive form of endosialidase from bacteriophage PK1A2 that specifically targets the polySia K1 capsule containing  $\alpha$ 2,8-linked N-acetylneuraminic acid units<sup>40,71</sup>, to the *E. coli* surface by flow cytometric analysis. rEndoNA2-GFP bound specifically and in a concentration-dependent manner to an K1 capsulated strain (Fig. 5c, Supplementary Fig. 3c). 147/149 (98.6%) isolates from all phylogroups displayed detectable levels of K1 capsule expression (Fig. 5d). This indicates that clinical *E. coli* isolates possessing the K1-*cps* locus properly synthesize polySia at their surface. The heterogeneity in the K1 expression levels could have arisen from variation in the expression of genes for endogenous NeuNAc synthesis in the K1-*cps* locus, a NeuNAc uptake transporter called NanT (encoded by *nanT*) and/or a NeuNAc degradation enzyme called NanA (encoded by *nanA*)<sup>26</sup>.



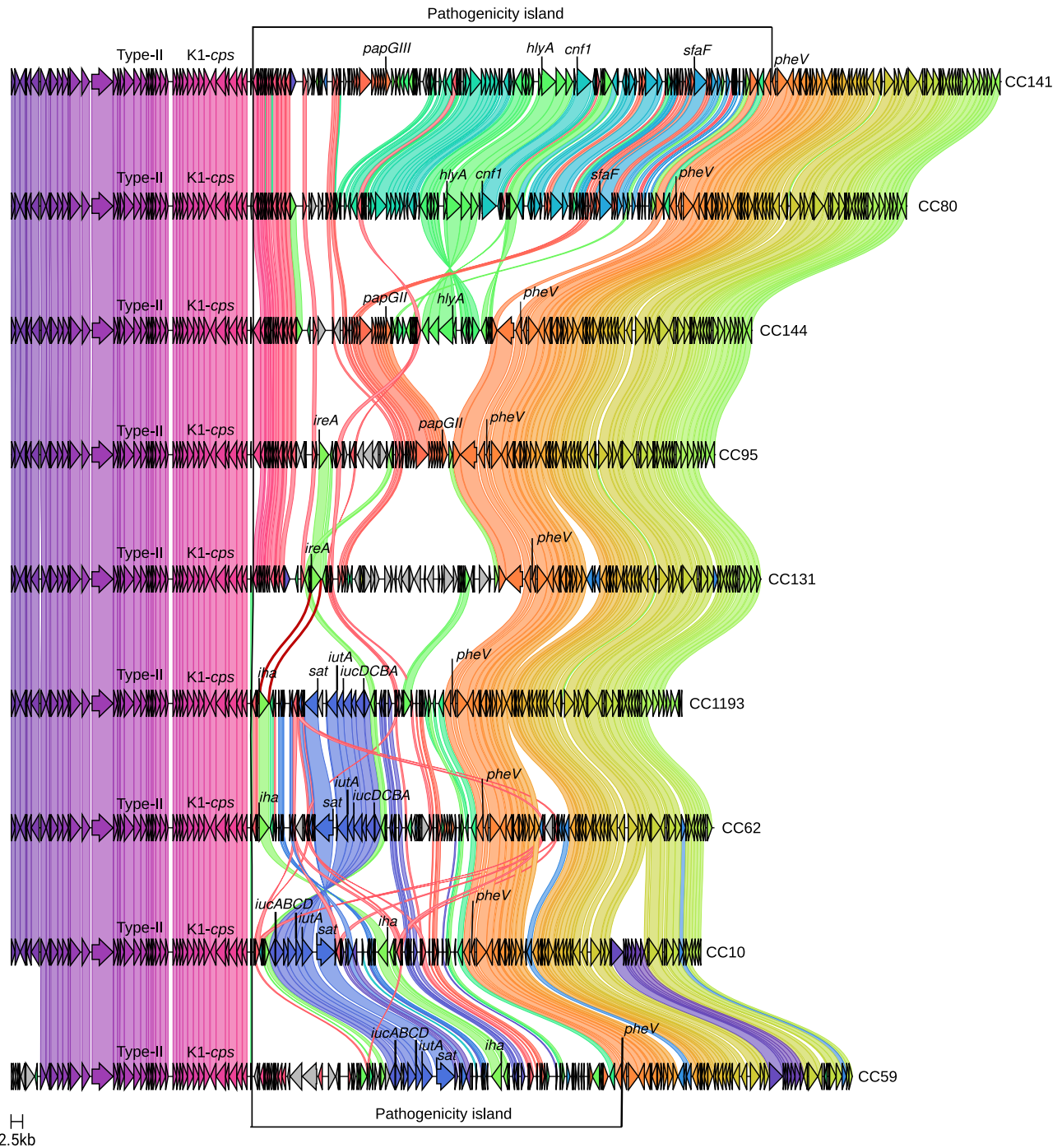
**Fig. 3 | Dated phylogenies and time acquisition of the K1-*cps* locus in the main distinct *E. coli* lineages.** In this analysis, we only included lineages for which we could confidently estimate their tMRCA using BactDating. For each lineage, a tick on the x-axis indicates either the (i) tMRCA (for lineages CC59 and CC95), or (ii) time interval of the K1-*cps* acquisition. For CC59 and CC95, the capsule was introduced by its MRCA and thus their K1-*cps* introductions overlap with the tMRCA (1525 [95% CI, 1044–1730] and 1768 [95% CI, 1721–1806], respectively). For CC10, the

most plausible scenario corresponded to two independent K1-*cps* acquisitions estimated between (i) 1789 [95% CI, 1624–1865] to 1821 [95% CI, 1677–1890] and (ii) 1853 [95% CI, 1738–1907] to 1873 [95% CI, 1772–1922]. For CC131, there was a single introduction of K1-*cps* estimated to have occurred between 1957 [95% CI, 1946–1967] to 1999 [95% CI, 1995–2002]. Source data are provided as a Source Data file.

We next assessed the capacity of the clinical *E. coli* isolates to evade deposition of C3b onto the bacterial surface during incubation in active human serum, using flow cytometric analysis. We found that 144/149 (96.6%) of *E. coli* isolates were resistant to C3b deposition (Fig. 5e). Notably, the level of K1 capsule expression was not correlated with the level of C3b deposition (Supplementary Fig. 3d). Next, we measured the capacity of the *E. coli* isolates to survive in human serum, by quantifying the number of viable bacteria before and after a 3-hour incubation period. Our analysis identified that 138/149 (92.6%) *E. coli* isolates were resistant to serum-killing (Fig. 5f). Like C3b deposition, there was no correlation between K1 capsule expression and percentage of surviving bacteria (Supplementary Fig. 3e).

### Therapeutic targeting of the K1 capsule re-sensitizes diverse ExPEC lineages to the human immune response

The development of effective strategies to prevent and treat *E. coli* diseases is urgent given the rise in hypervirulent and MDR *E. coli* isolates. However, broadly effective vaccines protective against pathogenic *E. coli* and therapeutics that target pathogenic *E. coli* have not been developed<sup>72</sup>. Given that the K1 capsule protects genetically diverse *E. coli* from serum-mediated killing, the removal or reduction of surface K1 capsule is likely to have a major impact on the capacity of *E. coli* to resist the human immune response and to cause infection. Bacteriophage enzymes have been exploited to develop novel strategies to kill bacterial pathogens<sup>73</sup>. This has included utilizing the



**Fig. 4 | Gene synteny of the neighbouring genomic region of the K1-cps locus among the most predominant CCs.** For each lineage, we obtained a chromosomal sequence of the isolates from the NORM collection 30348\_1#322 (CC141), 30224\_1#187 (CC80), 30224\_1#150 (CC144), 30134\_6#117 (CC95), 30345\_1#279 (CC131), 30348\_2#141 (CC1193), 30224\_1#376 (CC62), 30348\_2#375 (CC10) and 30348\_2#365 (CC59). For each chromosomal sequence, we extracted 50 kbp

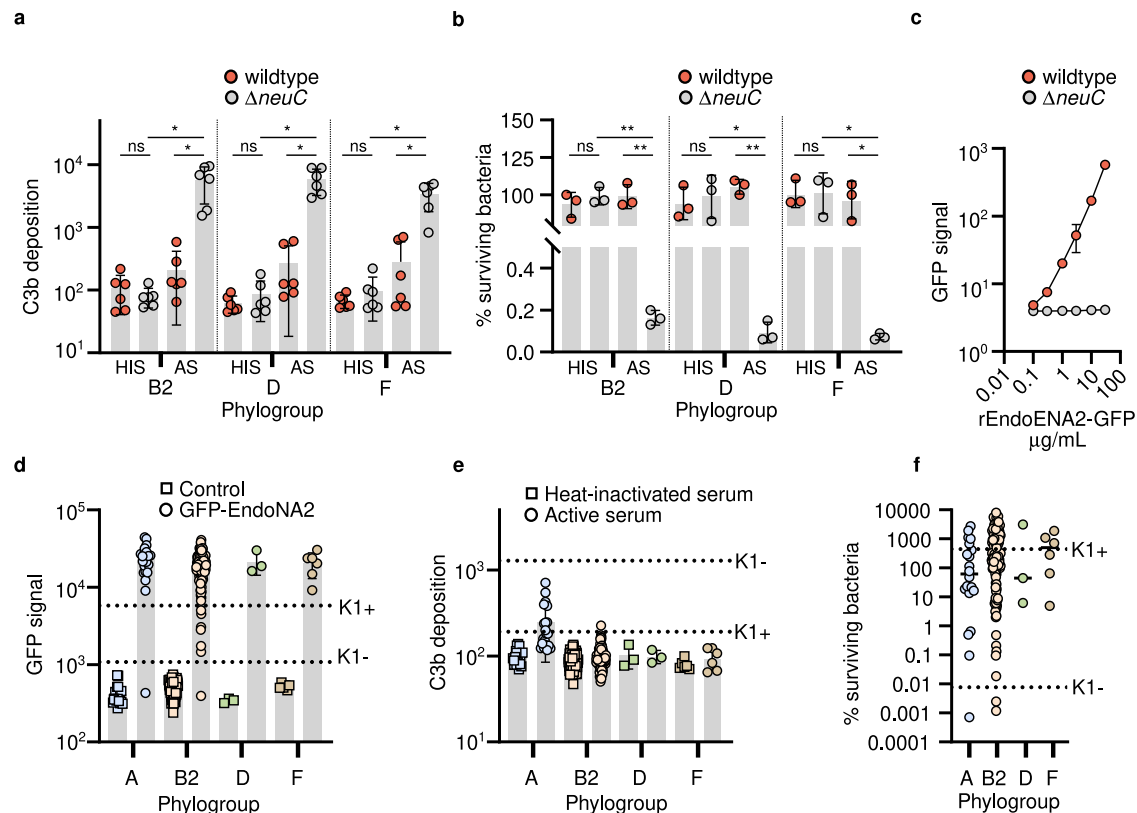
upstream and downstream of the *kpsM* and *intB* genes. *IntB* corresponds to the P4-like integrase gene present and adjacent to the 22 direct-repeat used by the pathogenicity island to be self-mobilised. The position of the tRNA *pheV*, K1-*cps* locus, and type II secretion system is indicated in the plot. The position of the pathogenicity island delimited by the K1-*cps* locus and the truncated tRNA *pheV* is also given. The virulence genes found in the PAI of each lineage are also indicated.

bacteriophage-derived capsule depolymerase EndoE to prevent and treat infections caused by K1 encapsulated *E. coli* in experimental animal models<sup>45,46</sup>. We explored the functional properties of rEndoE and its potential to re-sensitize genetically diverse K1-expressing *E. coli* to human serum.

To assess whether rEndoE was broadly able to re-sensitize genetically distinct K1-expressing *E. coli* to the human innate immune response, we expressed and purified rEndoE (Supplementary Fig. 4a) and examined its capacity to enhance C3b deposition and

serum killing of 21 genetically distinct *E. coli* K1-expressing isolates representative of each major phylogroups (A, B2, D and F). Incubation of human serum with 10 µg/mL of rEndoE enhanced the deposition of C3b onto the surface of each K1-expressing *E. coli* isolate during a 30-min incubation period (Fig. 6a; Supplementary Fig. 4b). Pooled analysis of all isolates revealed that rEndoE lead to a 10-fold increase in C3b deposition (Fig. 6b). Moreover, rEndoE decreased the survival of K1-expressing *E. coli* in human serum during a 180-min incubation period (Fig. 6c, d). Notably, as the enzymatic removal of K1 capsule increased





**Fig. 5 | Expression of K1 capsule correlates with resistance of *E. coli* to killing by human serum.** **a** K1 capsule expression enables *E. coli* strains to resist complement C3b deposition by human serum. Strains were incubated with human serum (AS) or heat-inactivated serum (HIS) for 30 min at 37 °C. Each data point is the mean  $\pm$  s.d. of  $n = 3$  independent experiments. Statistical significance tested by one-way ANOVA with Tukey's post-hoc test (B2,  $\Delta neuC$  HIS vs  $\Delta neuC$  AS  $*p = 0.0337$ , wildtype AS vs  $\Delta neuC$  AS  $*p = 0.0383$ ; D,  $\Delta neuC$  HIS vs  $\Delta neuC$  AS  $*p = 0.0105$ , wildtype AS vs  $\Delta neuC$  AS  $*p = 0.0105$ ; F,  $\Delta neuC$  HIS vs  $\Delta neuC$  AS  $*p = 0.0146$ , wildtype AS vs  $\Delta neuC$  AS  $*p = 0.0252$ ) **b** K1 capsule expression promotes *E. coli* to resist serum-mediated killing. *E. coli* strains were incubated with human serum or heat-inactivated serum for 3 h at 37 °C. Each data point is the mean  $\pm$  s.d. of  $n = 3$  independent experiments. Statistical significance tested by one-way ANOVA with Tukey's post-hoc test (B2,  $\Delta neuC$  HIS vs  $\Delta neuC$  AS  $**p = 0.0019$ , wildtype AS vs  $\Delta neuC$  AS  $**p = 0.0055$ ; D,  $\Delta neuC$  HIS vs  $\Delta neuC$  AS  $*p = 0.0177$ , wildtype AS vs  $\Delta neuC$  AS  $**p = 0.0005$ ; F,  $\Delta neuC$  HIS vs  $\Delta neuC$  AS  $*p = 0.0154$ , wildtype AS vs  $\Delta neuC$  AS  $*p = 0.0141$ ). **c** Concentration-

dependent binding of rEndoNA2-GFP to isogenic K1-*cps*<sup>+</sup> and K1-*cps*<sup>-</sup> ( $\Delta neuC$ ) *E. coli* strains, measured by flow cytometry analysis. Each data point is the mean  $\pm$  s.d. of  $n = 3$  independent experiments. **d** Surface expression of polySia K1 capsule by K1-*cps*<sup>+</sup> clinical *E. coli* isolates, determined by binding of rEndoNA2-GFP to the bacterial surface and subsequent flow cytometry analysis, where mean  $\pm$  s.d. are displayed. **e** C3b deposition on the surface of K1-*cps*<sup>+</sup> clinical *E. coli* isolates after incubation in active human serum, determined by binding of goat anti-human-C3 pAb to the bacterial surface and subsequent flow cytometry analysis, where mean  $\pm$  s.d. are displayed. **f** Survival of K1-*cps*<sup>+</sup> clinical *E. coli* isolates after incubation in human serum, determined by quantification of viable bacteria, where mean  $\pm$  s.d. are displayed. In (c), (d), (e) and (f), *E. coli* isolates are separated by phylogroup, where each data point represents one isolate (the mean quantified from  $n = 3$  independent experiments). Total number of isolates in each phylogroup are  $n = 20$  for A,  $n = 120$  for B2,  $n = 3$  for D and  $n = 6$  for E. Mean values for isogenic K1<sup>+</sup> (wildtype) and K1<sup>-</sup> ( $\Delta neuC$ ) strains are displayed. Source data are provided as a Source Data file.

C3b deposition and serum killing, this demonstrates that K1 capsule provides protection of *E. coli* against the human innate immune response. This observation is noteworthy as the deletion of a gene encoding a major surface component (such as *neuC*) could force a bacterial cell to reconfigure its surface topography, and this could indirectly enhance complement deposition and serum sensitivity. Importantly, our data demonstrates that therapeutic targeting of the K1 capsule can re-sensitize *E. coli* from distinct genetic backgrounds to the human innate immune response.

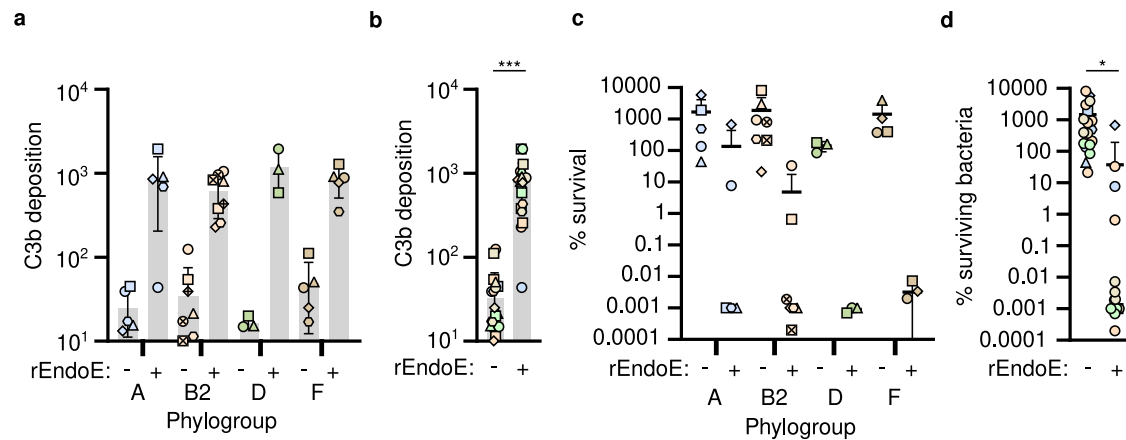
## Discussion

*E. coli* is one of the most clinically important bacterial pathogens worldwide<sup>74</sup>. Its ubiquitous presence in the human intestinal tract<sup>75</sup>, capacity to cause diverse intestinal and extraintestinal infections and the emergence of MDR have made the control of *E. coli* infections a major public health priority<sup>76</sup>. Although there has been a recent focus on the emergence of AMR in *E. coli*<sup>75</sup>, an increased incidence of invasive *E. coli* infections has been reported in many settings<sup>77–80</sup> and this trend is likely to continue due to the expanding elderly population. Thus, the identification of the defining features of enhanced virulence in *E. coli*

and the development of novel therapeutics are desirable to inform future control strategies.

The capsule polysaccharide is one of the major virulence determinants in extraintestinal *E. coli* by providing protection against the human immune system<sup>36–38</sup>. Despite the high diversity of K-antigens in *E. coli*<sup>19</sup>, we show that the genes encoding K1 capsule synthesis are found in ~25% of all bloodstream isolates from two independent longitudinal studies<sup>50,51</sup>. This high prevalence of the K1-*cps* locus in ExPEC isolates suggests that K1 encapsulated *E. coli* are more invasive compared to other *E. coli* encapsulated with other group 2 capsules. Our combined population genomic analysis coupled with in vitro experiments revealed that the capsule was functionally present in four distinct *E. coli* phylogroups providing immune protection irrespective of the genomic background.

We estimated that the K1-*cps* locus was introduced in the *E. coli* population more than 500 years ago, and we could identify later exchange events of the K1-*cps* locus within and between *E. coli* lineages. This included CC95, a non-MDR lineage responsible for most cystitis, pyelonephritis, bacteraemia, and meningitis cases worldwide<sup>20</sup>. The MRCA of CC95 acquired the K1 capsule around the mid-19<sup>th</sup> century



**Fig. 6 | Therapeutic targeting of capsule K1 enhances serum killing of *E. coli*.** **a, b** The effect of rEndoE capsule depolymerase on deposition of complement C3b on K1 encapsulated *E. coli* strains after incubation in human complement-replete serum. Each data point represents C3b deposition for one *E. coli* isolate, representing the mean from  $n = 2$  independent experiments. In **(a)**, each *E. coli* isolate is represented by a different symbol within the phylogroup. Mean  $\pm$  s.d. are displayed. Total number of isolates in each phylogroup are  $n = 5$  for A,  $n = 8$  for B2,  $n = 3$  for D and  $n = 5$  for E. In **(b)**, data points from **a** are pooled for all *E. coli* isolates ( $n = 21$ , with mean  $\pm$  s.d. displayed). Each *E. coli* isolate is coloured by phylogroup (blue = phylogroup A, orange = phylogroup B, green = phylogroup C, brown = phylogroup D) and differentiated by symbol. Statistical significance tested for C3b deposition by paired two-sided Student's *t* test (rEndoE<sup>-</sup> vs. rEndoE<sup>+</sup> \*\*\* $p < 0.0001$ ). **c, d** The

effect of rEndoE capsule depolymerase on deposition of survival of K1 encapsulated *E. coli* strains during incubation in human complement-replete serum. Each data point represents % survival of one *E. coli* isolate, representing the mean from  $n = 2$  independent experiments. In **(c)**, each *E. coli* isolate is represented by a different symbol within the phylogroup. Mean  $\pm$  s.d. are displayed. Total number of isolates in each phylogroup are  $n = 5$  for A,  $n = 8$  for B2,  $n = 3$  for D and  $n = 5$  for E. In **(d)**, data points from **c** are pooled for all *E. coli* isolates ( $n = 21$ , with mean  $\pm$  s.d. displayed). Each *E. coli* isolate is coloured by phylogroup (blue = phylogroup A, orange = phylogroup B, green = phylogroup C, brown = phylogroup D) and differentiated by symbol. Statistical significance tested for C3b deposition by paired two-sided Student's *t* test (rEndoE<sup>-</sup> vs. rEndoE<sup>+</sup> \* $p < 0.0101$ ). Source data are provided as a Source Data file.

and the capsule has remained a distinctive feature of this successful lineage ever since. We recently established the relative invasiveness of *E. coli* lineages by comparing their colonisation capacity with their capacity to cause BSIs and found that CC95 is significantly more pathogenic than other *E. coli* lineages<sup>81</sup>. The success of the CC95 clone in causing human BSIs is likely to be in part driven by the capacity of the K1 capsule to subvert the innate immune response. Other virulence factors are likely to additionally contribute to the capacity of CC95 to cause invasive infections. Of note, 80.3% of K1 encapsulated CC95 isolates carried a PAI downstream of the K1-*cps* locus. This PAI was characterized by the presence of the *ireA* gene and *papGII* gene operon involved in iron-scavenging and bacterial adhesion. Of note, the *papGII* gene operon has been recently identified using a GWAS approach as one of the bacterial traits distinguishing invasive from non-invasive uropathogenic *E. coli*<sup>60</sup>. In addition, this PAI is located in isolates of other lineages and also carried other accessory genes that could benefit bacterial survival or infer virulence properties, including HlyA<sup>63</sup>, CNF1<sup>64</sup>, Sat<sup>65</sup>, Sfa<sup>66</sup>, Ila<sup>67</sup> and aerobactin<sup>68</sup>. Thus, more comprehensive analysis is required to elucidate the function of PAI-encoded virulence genes as molecular determinants of BSIs. This is suggestive that phenotypes encoded on the PAI provide mechanisms for *E. coli* to enhance their survival and virulence capacity to cause BSIs. The fact that capsule with the same polySia K1 chemical composition has evolved independently in virulent lineages of other bacterial pathogens, namely group B *Neisseria meningitidis*, *Moraxella nonliquefaciens* and *Mannheimia haemolytica* A2<sup>82-84</sup>, further suggests that polySia enhances bacterial survival and/or virulence.

The 'origin' of the K1-*cps* encoding K1 encapsulation and 'transfer mechanism' remains unclear. Our data reveal multiple independent introductions of the K1-*cps* locus into *E. coli* populations. This can partially be attributed to horizontal gene transfer (HGT) of the capsule locus between *E. coli* populations. Downstream of the K1-*cps* locus, we observed the presence of a PAI with a non-uniform gene content truncating the *pheV* tRNA gene in all lineages. The gene synteny in this PAI was considered as evidence to infer which lineages acted as donors of the K1-*cps* locus. Furthermore, this PAI can form a circular excision

product that may be exported by larger MGEs<sup>61</sup>. The most likely transfer mechanisms of K1-*cps* are phage transduction and natural competence as the *cps* locus in *E. coli* does not possess an integrase or conjugation machinery.

Surprisingly, our analysis indicated that CC95 has not been the donor clone for most K1-*cps* locus transfers into recipient lineages, despite being the most dominant K1-positive *E. coli* clone and existing for over 250 years. Instead, our analysis indicated transfer of the K1-*cps* from phylogroup A to phylogroup B2 and F, and multiples transfers between lineages of phylogroup B2. The simplest hypothesis to explain why the dominant K1-*cps*+ clone (CC95) does not serve as the donor is the existence of evolutionary barriers preventing HGT. The capsule itself could provide a physical barrier that prevents infection of CC95 donors by phage that can lysogenize and infect the recipient<sup>85</sup>. There could be moments of capsule inactivation or swap that could be extremely deleterious for CC95 as it is a major virulence determinant, which could prevent the long-term persistence of other capsule types in CC95<sup>86</sup>. In addition, or alternatively, CC95 and recipient/donor *E. coli* could be rarely present in sufficient densities in the same environmental niche to permit HGT. This seems unlikely given the global success of the CC95 clone, unless CC95 effectively suppresses most of *E. coli* competitors in the environment. A recent large-scale analysis showed that genes encoding group 2 capsules were detectable in 11 genera within the Enterobacteriaceae belonging to 3 bacterial families<sup>87</sup>, highlighting that transfer of *cps* loci among species. In support of this, searching the NCBI genome catalogue detected the genes of the K1-*cps* locus in *Enterobacter hormaechei*<sup>88</sup> and *Shigella flexneri* (GenBank: CP041620). Additional studies to understand HGT mechanisms of *cps* transfer, and the barriers to transfer, are required.

Our analysis demonstrates that the K1-*cps* can be efficiently acquired by *E. coli*, and that K1 capsule expression provides protection against the human immune system. If K1-*cps* acquisition is so beneficial to *E. coli*, then the fact that 75% of all BSI *E. coli* isolates do not possess K1-*cps* genes suggests that there are costs to a bacterium associated with K1 capsule expression. The immediate costs of K1 capsule production are reduced growth rates, as the production of K1

polySia imposes a substantial metabolic burden<sup>89</sup>, and an increased susceptibility to killing by bacteriophage that targets the K1 capsule<sup>90</sup>. In addition, capsule synthesis may cause a drastic reduction in plasmid conjugation<sup>86</sup>, which could limit the acquisition of new traits required to maintain survival or competitiveness in the local environment, such as antimicrobial resistance genes which are usually disseminated through plasmid conjugation. The fact that K1 capsule expression is thermoregulated<sup>26,91</sup>, downregulated in the urine/bladder environment<sup>92</sup> and switched off in the brain environment<sup>40</sup>, is further suggestive that there are costs associated with K1 expression inside and outside of the mammalian host. These findings highlight the need for future studies that establish the dynamics and functions of K1 capsule expression on the pathway to causing infection.

Much attention and research effort has been directed towards the increasing levels of antimicrobial resistance in *E. coli*<sup>75</sup>. In particular, the rise of MDR isolates belonging to CCI31 lineage has been a global concern<sup>93</sup>. The dominant K1 positive lineage is CC95. This lineage has remained a dominant cause of BSIs for over two decades<sup>50,51,94,95</sup> and a major cause of UTIs<sup>51,60,96</sup>. The dominance of this lineage as a cause of invasive infections is despite CC95 classically being a non-MDR lineage, with the population displaying susceptibility to third generation cephalosporins and carbapenems but strain-dependent resistance to many other antibiotics. Of concern, is the recent report of CC95 isolates in which MDR and K1 capsule expression have converged<sup>97</sup>. As the administration of antibiotics to treat a bacterial infection disturbs the normal gastrointestinal microbiota<sup>98</sup>, can select for colonisation with pathogens<sup>99</sup>, and can increase the risk of infectious and non-infectious diseases<sup>100,101</sup>, there is a strong incentive for the development of anti-bacterial therapies that target specific pathogens without disrupting the microbiota. Here, we have demonstrated that therapeutic targeting of the K1 capsule using a bacteriophage-derived enzyme can sensitize K1 expressing *E. coli* isolates from distinct genetic backgrounds to the human immune response. This K1-capsule depolymerase effectively prevents and treats invasive *E. coli* infections in experimental animal models<sup>45,46</sup>, and similar therapeutic approaches have been effective in enhancing protective immunity against some of the most clinically important bacterial pathogens<sup>102–105</sup>. As our data indicates that around ~25% of *E. coli* causing BSIs are K1 encapsulated, such a therapeutic approach would have a considerable potential to effectively control invasive *E. coli* infections whilst significantly lowering antibiotic usage and its adverse effects.

## Methods

### Ethics

Human serum was prepared from blood obtained from healthy donors, approved by the Regional Ethics Committee and Imperial College Healthcare NHS Trust Tissue Bank (Regional Ethics Committee approval no. 17/WA/O161, Imperial College Healthcare Tissue Bank Human Tissue Authority license no. 12275, and Imperial College Research Ethics Committee no. 19IC5166). All samples were collected after receiving signed informed consent from all participants.

### Bacterial strains

To delete the *neuC* gene present in different *E. coli* lineages, we made use of  $\lambda$ -red recombineering<sup>106</sup> or the thermosensitive allelic replacement using the plasmid pKO3blue<sup>107</sup>. The mutants generated using  $\lambda$ -red recombineering were generated using primers listed in Dataset S5. For mutants generated using thermosensitive allelic replacement, the desired *neuC* chromosomal flanking regions were amplified by PCR using the primers listed in Dataset S5. Then, a fusion PCR from both flanking regions was performed and cloned into <sup>107</sup>pKO3blue, using enzymatic digestion and ligation following standard protocol. The generated plasmids are listed on Dataset S5. To perform the allelic replacement, the desired pKO3blue plasmid was introduced into the correspond strain and the transformants were selected on LB agar

plates supplemented with 20  $\mu$ g/ml chloramphenicol and incubated at 32 °C. Then, one colony was grown at non-permissive temperature, 42 °C, to force the plasmid to integrate into the chromosome and plated on LB agar containing 20  $\mu$ g/ml chloramphenicol and X-gal (5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside). Light blue colonies, which are indicative of plasmid integration, were grown in LB broth at 32 °C. Then, ten-fold serial dilution of the overnight cultures was plated on LB agar containing X-gal and sucrose 5%, which favour plasmid loss. Plates were incubated at 32 °C for 24 h and white colonies, which are indicative that the plasmid is lost, were screened for chloramphenicol sensitivity. A PCR was performed to differentiate wild-type and mutant strains and mutants were verified by sequencing. Isogenic strains constructed and used are shown in Dataset S6. Strains used in population-based screens are shown in Dataset S7. Strains used in rEndoE-based assays are shown in Dataset S8.

### Expression of GFP-endoNA2 (rEndoNA2-GFP) and measurement of K1 capsule expression by *E. coli*

An *E. coli* strain carrying the pFEndoNA2 plasmid was previously generated<sup>108</sup> and accessed through AddGene. The pFEndoNA2 plasmid expresses the GFP-EndoNA2 fusion protein (rEndoNA2-GFP) with a C-terminal 6xHis tag, that includes the catalytically inactive form of the endosialidase from bacteriophage PK1A2. rEndoNA2-GFP was expressed by culture of *E. coli* in LB media at 37 °C with shaking (180 rpm) until absorbance at OD<sub>600</sub> reached 0.6. After addition of 1 mM IPTG, *E. coli* were cultured for a subsequent 24 h at 30 °C with shaking (180 rpm). Bacteria were lysed, and rEndoNA2-GFP was purified using a HiStrap Nickel column (GE Healthcare Life Sciences) and affinity chromatography (AKTA Pure, GE Healthcare Life Sciences). Purified rEndoNA2-GFP was dialysed against PBS at 4 °C.

Levels of K1 capsule expression were quantified by measuring rEndoNA2-GFP binding to *E. coli* using flow cytometry.  $6 \times 10^6$  of logarithmic phase bacteria were incubated on ice for 1 h with 10  $\mu$ g/ml rEndoNA2-GFP. Bacteria were washed in PBS before fixing in 1% (w/v) PFA. Fluorescence was measured by flow cytometry analysis using an Amnis CellStream Benchtop Flow Cytometer. Data was collected using CellStream Acquisition and Analysis Software and analysed using FlowJo v10.8. Gating strategy is shown in Supplementary Fig. 4c.

### Expression of rEndoE

The pQE30.EndoE plasmid<sup>45</sup> was used to express the EndoE protein with a C-terminal 6xHis tag, that includes the catalytically active form of the endosialidase from phage K1E. EndoE was expressed by culture of *E. coli* in LB media at 37 °C with shaking (180 rpm) until absorbance at OD<sub>600</sub> reached 0.35. After addition of 0.5 mM IPTG, *E. coli* were cultured for a subsequent 18 h at 30 °C with shaking (180 rpm). Bacteria were lysed, and EndoE was purified using a HiStrap Nickel column (GE Healthcare Life Sciences) and affinity chromatography (AKTA Pure, GE Healthcare Life Sciences).

### Flow cytometry measurement of complement C3b deposition on the *E. coli* surface

Levels of C3b deposition on *E. coli* following exposure to human serum were measured by flow cytometry.  $6 \times 10^6$  of logarithmic phase bacteria were incubated at 37 °C for 30 min in PBS, 3% pooled active human serum or 3% pooled heat-inactivated human serum. Bacteria were washed twice in PBS, followed by centrifugation at  $2058 \times g$  for 5 min and removal of the supernatant. C3b deposition on the bacterial surface was detected by incubation on ice for 30 min with polyclonal FITC-conjugated goat anti-human-C3b (dilution 1:2000) (MP Biomedical). Bacteria were washed as previously described before fixing in 1% (w/v) PFA. Fluorescence was measured by flow cytometry analysis using an Amnis CellStream Benchtop Flow Cytometer. Data was collected using CellStream Acquisition and Analysis Software and analysed using FlowJo v10.8.

## Susceptibility of *E. coli* to human serum

Bacteria were grown to logarithmic phase and washed twice with gelatin-Veronal-buffered saline plus magnesium and calcium ions (pH 7.35) (GVB++; Sigma Aldrich). Pooled human serum was diluted 1:3 in GVB++ and prewarmed to 37 °C. Bacteria were resuspended in GVB++ and mixed with serum solutions 1:2 to give a final concentration of  $\sim 10^7$  CFU/ml in a total volume of 125  $\mu$ l containing 22% serum. *E. coli* and serum mixtures were incubated at 37 °C for 3 h, and the number of surviving bacteria were quantified by serial dilution and incubation on LB agar overnight. Prewarmed, heat-inactivated (56 °C for 30 min) serum served as a control.

## Bioinformatics

**In silico detection of the K1 capsule locus.** The BSAC, NORM and phenotypically confirmed K1+ collection were screened using an in silico PCR of the assemblies for each of the six *neu* genes. If at least one *neu* gene was detected the isolates were further investigated as putative K1 loci. The K1 *neu* gene alleles from the NORM and BSAC genomes were used to screen 12230 genomes in the Horesh collection, that contains the historical isolates published by Baker et al., using the Horesh et al. BIGSI database to identify further K1 genomes<sup>52</sup>. The putative K1 genomes were screened using an in silico PCR of the assemblies for the entire *kpsF-kpsM* locus. A full length K1 loci was used as a reference database for ARIBA to assemble the putative K1 loci for the genomes which the locus PCR of the assemblies did not return a product. Alleles were assigned for each locus that assembled into a single contig, and panaroo K1 loci gene groups using any difference in the sequence to define a new allele, partial alleles (<95% length) were named as such, only genomes with 14 full length K1 gene sequences or a confirmed K1 phenotype were considered K1. The DNA alleles were translated using EMBOSS transeq and assigned protein alleles.

## Species phylogenetic tree

The entire NORM<sup>50</sup> and BSAC collection<sup>51</sup> (K1 and non-K1), the genotypically confirmed K1 from Horesh et al.<sup>52</sup> and the phenotypically confirmed K1+ collection presented in this work (accession list in Dataset S1), were mapped against *E. coli* strain EC958 (accession HG941718; <https://www.ncbi.nlm.nih.gov/nuccore/HG941718.1>) to create a species-wide phylogeny using iqtree (version 2.2.0.3) with the fast and GTR settings<sup>109</sup>. Phylotypes were assigned using ClermontTyper (version 0.6.3)<sup>110</sup> and corrected based on the phylogenetic clusters, and STs were assigned by mlst <https://github.com/tseemann/mlst> (version 2.9) which makes use of the PubMLST database<sup>111</sup>, and overlaid on the phylogeny in Microreact (version 2.1.8)<sup>112</sup>.

## Lineage analysis

To investigate the age of K1 isolates, the NORM collection isolates belonging to the CC59, and CC95, CC131 and CC10 lineages as defined in ref. 50 were dated using BactDating<sup>113</sup>. Briefly, isolates were mapped to a lineage reference, and gubbins (version 2.4.1)<sup>114</sup> used to remove recombination to generate the BactDating (version 1.1.0) input. The lineages were run in triplicate for multiple bactdating models with MCMC chains of 100,000,000, convergence was assessed using the R (version 4.0) coda package (version 0.19) considering the effective sample size (>200) and Gelman-Rubin diagnostic (-1). The final model was compared to a randomised dates model considering the deviance information criterion.

For each K1-*cps*+ lineage, we obtained a complete reference genome by conducting long-read sequencing and generating a hybrid assembly. This allowed us to confidently characterize the gene synteny present downstream of the K1-*cps* locus. The long-read sequencing and subsequent hybrid assembly was performed as previously described in Arredondo-Alonso et al.<sup>115</sup> to obtain a complete genome sequence from each K1+ lineage present in the NORM collection. We selected the following genomes 30348\_1#322 (CC141), 30224\_1#187 (CC80),

30224\_1#150 (CC144), 30134\_6#117 (CC95), 30345\_1#279 (CC131), 30348\_2#141 (CC1193), 30224\_1#376 (CC62), 30348\_2#375 (CC10) and 30348\_2#365 (CC59). For each genome, we rotated their starting position considering the *dnaA* gene using circlator (fixstart, version 1.5.5) and extracted with bedtools (-getfasta) 50 kbp upstream *kpsM* and 50 kbp downstream the integrase *intB* gene found in the island adjacent to the K1 locus. We annotated the extracted regions with prokka (-genus Escherichia,-species coli) and visualized their gene synteny similarity with clinker (default settings, version 0.0.21).

## K1-*cps* synteny in *E. coli* near-complete genomes

We retrieved all the complete genomes from the RefSeq database (queried in October 2022) annotated as ST95 in Enterobase<sup>116</sup>. The starting position of the chromosomal sequence was rotated considering the *dnaA* gene using circlator (fixstart, version 1.5.5)<sup>117</sup>. The genomes were annotated with prokka (version 1.14.6) considering the *Escherichia* scheme (-genus Escherichia,-species coli)<sup>118</sup>. The pangenome of those ST95 complete genomes was computed with Panaroo (version 1.2.3)<sup>119</sup> in the strict mode (-clean-mode,-no\_clean\_edges) considering the gff files computed by prokka as input. The module 'panaroo-gene-neighbourhood' was used to group the genomes based on the gene graph neighbourhood of the 10 genes (-expand\_no 10) adjacent (upstream and downstream) of the *kpsM* and *kpsF* genes. We focused our analyses on the synteny found by panaroo downstream of the *kpsF* gene and further identified the presence of the pathogenicity island (PAI-V) characterized by the insertion of a -50 kbp region in the *pheV* tRNA gene mediated by a P4-like integrase (*intB* gene) with similarity to the RefSeq amino acid sequence EG12364. The size of the PAI was determined by extracting the size of the genome block surrounding the truncated copies of the *pheV* gene. To visualize the most predominant synteny blocks computed by panaroo, we considered the complete chromosomal sequences from the RefSeq assemblies GCA\_002285755.1 and GCA\_001280405.1 and extracted the nucleotide sequence corresponding to 50 kbp upstream and 70 kbp downstream of the K1-*cps* locus with bedtools (-getfasta, version 2.30.0)<sup>120</sup>. The extracted regions were annotated with prokka (-genus Escherichia,-species coli) and the visualization of their gene clusters similarity was performed with clinker (default settings, version 0.0.21)<sup>121</sup>.

We determined the presence of this PAI in the NORM CC95 isolates by inspecting the tRNA *pheV* gene in the short-read assemblies of the NORM collection using a customized database in abricate (version 1.0.1, <https://github.com/tseemann/abricate>). The presence of a truncated copy of the *pheV* gene in the same short-read contig as for the *kpsF* gene indicated the possible presence of the PAI. We confirmed this by nucleotide searching (blastn) for the PAI-related integrase *intB* gene on the same contig as the other truncated copy of *pheV*. Furthermore, we determined the size of the PAI in the cases where the two copies of the truncated *pheV* gene were present on the same short-read contig. Lastly, the presence on the same short-read contig of an intact *pheV* gene downstream the *kpsF* was considered as evidence for the absence of the PAI. The R library ggtree (version 3.5.1.902)<sup>122</sup> was used to plot the presence/absence of the PAI in the NORM CC95 phylogeny. To screen the virulence gene content present in the NORM CC95 PAI and in the PAIs associated to each of the hybrid assemblies generated from the other K1-*cps*+ lineages (Dataset S4), we used amrfinder (version 3.10.18) with the flags -plus and -organism (indicating *Escherichia* as the taxonomic group)<sup>123</sup>.

## SNP profile of the region 2 of the K1-*cps* locus

To investigate and support the introductions of the K1-*cps* locus, we used parsnp (version 1.2)<sup>124</sup> to compute a pairwise matrix with the SNP differences considering as input a concatenation of the alignments produced by panaroo on the genes corresponding to the region 2 of the locus. Isolates were grouped according to their CC lineage, and the median number of SNPs found within and between CC lineages

reported. To visualize isolates with a similar SNP profile, we embedded the SNP distances computed by parsnp into two dimensions using the t-Distributed Stochastic Neighbour Embedding algorithm as implemented in the Rtsne package (version 0.15) (perplexity = 50, theta = 0.5, is\_distance = TRUE)<sup>125</sup>.

### Statistics and reproducibility

All quantitative data were analysed and graphed using GraphPad Prism 8.4.3 software. All data are displayed as mean  $\pm$  s.d. calculated using the GraphPad Prism 8.4.3 software, unless indicated otherwise. Statistical details of the experiments are provided in the respective figure legends and in each methods section pertaining to the specific technique applied. Bactdating (version 1.10) was used to date the phylogenies of the lineages CC59, CC95, CC131 and CC10. The bactdating models were run in triplicate with MCMC chains of 100,000,000, their convergence was assessed using the R package coda (version 0.19) considering the effective sample size of the parameters (>200) and the Gelman-Rubin diagnostic (-1). The final model was compared to a randomised dates model using bactdating (version 1.10). No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

### Reporting summary

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

### Data availability

The Illumina sequencing data generated in this study have been deposited in the NCBI database under accession codes listed in Dataset S1. The publicly available Illumina sequencing data used in this study have previously been deposited in the NCBI database under accession codes listed in Dataset S1. The accessions of the complete genomes from CC95 retrieved from RefSeq are listed in Dataset S3. The data generated in this study are provided in the Supplementary Information or Source Data file or on Microreact <https://microreact.org/project/cm8f8adPyWoQemAAgvrmi-k1-context>. The Oxford Nanopore Technologies sequencing data used to generate (near-) complete chromosomal sequences for: 30348\_1#322 (CC141), 30224\_1#187 (CC80), 30224\_1#150 (CC144), 30134\_6#117 (CC95), 30345\_1#279 (CC131), 30348\_2#141 (CC1193), 30224\_1#376 (CC62), 30348\_2#375 (CC10) and 30348\_2#365 (CC59) are available through the accessions listed in Dataset S4, and their complete chromosomal sequences shared in the figshare project <https://doi.org/10.6084/m9.figshare.21610674><sup>126</sup>. Source data are provided with this paper.

### Code availability

An Rmarkdown document with the code and files required to generate the figures and results presented here is available at [https://gitlab.com/sirarredondo/k1\\_manuscript](https://gitlab.com/sirarredondo/k1_manuscript).

### References

- Rendueles, O., Garcia-Garcera, M., Néron, B., Touchon, M. & Rocha, E. P. C. Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PLoS Pathog.* **13**, e1006525 (2017).
- Nucci, A., Rocha, E. P. C. & Rendueles, O. Adaptation to novel spatially-structured environments is driven by the capsule and alters virulence-associated traits. *Nat. Commun.* **13**, 4751 (2022).
- Finlay, B. B. & Falkow, S. Common themes in microbial pathogenicity. *Microbiol. Rev.* **53**, 210–230 (1989).
- Bentley, S. D. et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* **2**, e31 (2006).
- Cress, B. F. et al. Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol. Rev.* **38**, 660–697 (2014).
- Buckles, E. L. et al. Role of the K2 capsule in *Escherichia coli* urinary tract infection and serum resistance. *J. Infect. Dis.* **199**, 1689–1697 (2009).
- Hyams, C. et al. *Streptococcus pneumoniae* resistance to complement-mediated immunity is dependent on the capsular serotype. *Infect. Immun.* **78**, 716–725 (2010).
- Campos, M. A. et al. Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infect. Immun.* **72**, 7107–7114 (2004).
- Cole, J. N. et al. M protein and hyaluronic acid capsule are essential for in vivo selection of covRS mutations characteristic of invasive serotype M1T1 group A *Streptococcus*. *MBio* **1**, e00191-10 (2010).
- Brissac, T. et al. Capsule promotes intracellular survival and vascular endothelial cell translocation during invasive Pneumococcal disease. *MBio* **12**, e0251621 (2021).
- Carlin, A. F., Lewis, A. L., Varki, A. & Nizet, V. Group B streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. *J. Bacteriol.* **189**, 1231–1237 (2007).
- Secundino, I. et al. Host and pathogen hyaluronan signal through human siglec-9 to suppress neutrophil activation. *J. Mol. Med.* **94**, 219–233 (2016).
- Thakker, M., Park, J. S., Carey, V. & Lee, J. C. *Staphylococcus aureus* serotype 5 capsular polysaccharide is antiphagocytic and enhances bacterial virulence in a murine bacteremia model. *Infect. Immun.* **66**, 5183–5189 (1998).
- Hyams, C., Camberlein, E., Cohen, J. M., Bax, K. & Brown, J. S. The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect. Immun.* **78**, 704–715 (2010).
- Domenico, P., Salo, R. J., Cross, A. S. & Cunha, B. A. Polysaccharide capsule-mediated resistance to opsonophagocytosis in *Klebsiella pneumoniae*. *Infect. Immun.* **62**, 4495–4499 (1994).
- Spinosa, M. R. et al. The *Neisseria meningitidis* capsule is important for intracellular survival in human cells. *Infect. Immun.* **75**, 3594–3603 (2007).
- Burns, S. M. & Hull, S. I. Loss of resistance to ingestion and phagocytic killing by O(-) and K(-) mutants of a uropathogenic *Escherichia coli* O75:K5 strain. *Infect. Immun.* **67**, 3757–3762 (1999).
- Llobet, E., Tomás, J. M. & Bengoechea, J. A. Capsule polysaccharide is a bacterial decoy for antimicrobial peptides. *Microbiology* **154**, 3877–3886 (2008).
- Whitfield, C. & Roberts, I. S. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol. Microbiol.* **31**, 1307–1319 (1999).
- Riley, L. W. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin. Microbiol. Infect.* **20**, 380–390 (2014).
- Cross, A. S., Gemski, P., Sadoff, J. C., Orskov, F. & Orskov, I. The importance of the K1 capsule in invasive infections caused by *Escherichia coli*. *J. Infect. Dis.* **149**, 184–193 (1984).
- Dewitt, C. W. & Rowe, J. A. Sialic acids (N,7-O-diacetylneuraminic acid and N-acetylneuraminic acid) in *Escherichia coli*. I. Isolation and identification. *J. Bacteriol.* **82**, 838–848 (1961).
- Rodriguez, M. L., Jann, B. & Jann, K. Structure and serological characteristics of the capsular K4 antigen of *Escherichia coli* O5:K4:H4, a fructose-containing polysaccharide with a chondroitin backbone. *Eur. J. Biochem.* **177**, 117–124 (1988).
- Blundell, C. D., Roberts, I. S., Sheehan, J. K. & Almond, A. Investigating the molecular basis for the virulence of *Escherichia coli* K5 by nuclear magnetic resonance analysis of the capsule polysaccharide. *J. Mol. Microbiol. Biotechnol.* **17**, 71–82 (2009).

25. Vann, W. F., Schmidt, M. A., Jann, B. & Jann, K. The structure of the capsular polysaccharide (K5 antigen) of urinary-tract-infective *Escherichia coli* O10:K5:H4. A polymer similar to desulfo-heparin. *Eur. J. Biochem.* **116**, 359–364 (1981).
26. Whitfield, C. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* **75**, 39–68 (2006).
27. Sarff, L. D. et al. Epidemiology of *Escherichia coli* K1 in healthy and diseased newborns. *Lancet* **1**, 1099–1104 (1975).
28. Bingen, E., Bonacorsi, S., Brahimi, N., Denamur, E. & Elion, J. Virulence patterns of *Escherichia coli* K1 strains associated with neonatal meningitis. *J. Clin. Microbiol.* **35**, 2981–2982 (1997).
29. Korhonen, T. K. et al. Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. *Infect. Immun.* **48**, 486–491 (1985).
30. Silver, R. P., Aaronson, W. & Vann, W. F. The K1 capsular polysaccharide of *Escherichia coli*. *Rev. Infect. Dis.* **10**, S282–S286 (1988).
31. Kaijser, B. Immunology of *Escherichia coli*: K antigen and its relation to urinary-tract infection. *J. Infect. Dis.* **127**, 670–677 (1973).
32. Egan, W. et al. Structural studies on the sialic acid polysaccharide antigen of *Escherichia coli* strain Bos-12. *Biochemistry* **16**, 3687–3692 (1977).
33. Barry, G. T. & Goebel, W. F. Colominic acid, a substance of bacterial origin related to sialic acid. *Nature* **179**, 206 (1957).
34. Schnaar, R. L., Gerardy-Schahn, R. & Hildebrandt, H. Sialic acids in the brain: gangliosides and polysialic acid in nervous system development, stability, disease, and regeneration. *Physiol. Rev.* **94**, 461–518 (2014).
35. Villanueva-Cabello, T. M., Gutiérrez-Valenzuela, L. D., Salinas-Marín, R., López-Guerrero, D. V. & Martínez-Duncker, I. Polysialic acid in the immune system. *Front. Immunol.* **12**, 823637 (2021).
36. Opal, S., Cross, A. & Gemski, P. K antigen and serum sensitivity of rough *Escherichia coli*. *Infect. Immun.* **37**, 956–960 (1982).
37. Weiss, J., Victor, M., Cross, A. S. & Elsbach, P. Sensitivity of K1-encapsulated *Escherichia coli* to killing by the bactericidal/permeability-increasing protein of rabbit and human neutrophils. *Infect. Immun.* **38**, 1149–1153 (1982).
38. Abreu, A. G. & Barbosa, A. S. How *Escherichia coli* circumvent complement-mediated killing. *Front. Immunol.* **8**, 452 (2017).
39. Pluschke, G., Mayden, J., Achtman, M. & Levine, R. P. Role of the capsule and the O antigen in resistance of O18:K1 *Escherichia coli* to complement-mediated killing. *Infect. Immun.* **42**, 907–913 (1983).
40. Zelmer, A. et al. Differential expression of the polysialyl capsule during blood-to-brain transit of neuropathogenic *Escherichia coli* K1. *Microbiology* **154**, 2522–2532 (2008).
41. Anderson, G. G., Goller, C. C., Justice, S., Hultgren, S. J. & Seed, P. C. Polysaccharide capsule and sialic acid-mediated regulation promote biofilm-like intracellular bacterial communities during cystitis. *Infect. Immun.* **78**, 963–975 (2010).
42. Birchenough, G. M. H. et al. Postnatal development of the small intestinal mucosa drives age-dependent, regio-selective susceptibility to *Escherichia coli* K1 infection. *Sci. Rep.* **7**, 83 (2017).
43. Orskov, I. & Orskov, F. *Escherichia coli* in extra-intestinal infections. *J. Hyg.* **95**, 551–575 (1985).
44. Moxley, R. A. & Francis, D. H. Natural and experimental infection with an attaching and effacing strain of *Escherichia coli* in calves. *Infect. Immun.* **53**, 339–346 (1986).
45. Mushtaq, N., Redpath, M. B., Luzio, J. P. & Taylor, P. W. Prevention and cure of systemic *Escherichia coli* K1 infection by modification of the bacterial phenotype. *Antimicrob. Agents Chemother.* **48**, 1503–1508 (2004).
46. Mushtaq, N., Redpath, M. B., Luzio, J. P. & Taylor, P. W. Treatment of experimental *Escherichia coli* infection with recombinant bacteriophage-derived capsule depolymerase. *J. Antimicrob. Chemother.* **56**, 160–165 (2005).
47. Pitt, J. Virulence of *Escherichia coli* K1 in adults. *J. Infect. Dis.* **139**, 106–108 (1979).
48. Glode, M. P. et al. Neonatal meningitis due of *Escherichia coli* K1. *J. Infect. Dis.* **136**, S93–S97 (1977).
49. Wilfert, C. M. *E. coli* meningitis: K1 antigen and virulence. *Annu. Rev. Med.* **29**, 129–136 (1978).
50. Gladstone, R. A. et al. Emergence and dissemination of antimicrobial resistance in *Escherichia coli* causing bloodstream infections in Norway in 2002–17: a nationwide, longitudinal, microbial population genomic study. *Lancet Microbe* **2**, e331–e341 (2021).
51. Kallonen, T. et al. Systematic longitudinal survey of invasive *Escherichia coli* in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. *Genome Res.* <https://doi.org/10.1101/gr.216606.116>. (2017)
52. Horesh, G. et al. A comprehensive and high-quality collection of *E. coli* genomes and their genes. *Microb. Genom.* **7**, e000499 (2021).
53. Baker, K. S. et al. The Murray collection of pre-antibiotic era Enterobacteriaceae: a unique research resource. *Genome Med.* **7**, 97 (2015).
54. Janke, B., Dobrindt, U., Hacker, J. & Blum-Oehler, G. A subtractive hybridisation analysis of genomic differences between the uropathogenic *E. coli* strain 536 and the *E. coli* K-12 strain MG1655. *FEMS Microbiol. Lett.* **199**, 61–66 (2001).
55. Vimr, E. R. & Steenbergen, S. M. Mobile contingency locus controlling *Escherichia coli* K1 polysialic acid capsule acetylation. *Mol. Microbiol.* **60**, 828–837 (2006).
56. Kulkarni, R. et al. Roles of putative type II secretion and type IV pilus systems in the virulence of uropathogenic *Escherichia coli*. *PLoS ONE* **4**, e4752 (2009).
57. Schneider, G. et al. The pathogenicity island-associated K15 capsule determinant exhibits a novel genetic structure and correlates with virulence in uropathogenic *Escherichia coli* strain 536. *Infect. Immun.* **72**, 5993–6001 (2004).
58. Middendorf, B. et al. Instability of pathogenicity islands in uropathogenic *Escherichia coli* 536. *J. Bacteriol.* **186**, 3086–3096 (2004).
59. Desvaux, M. et al. Pathogenicity factors of genomic islands in intestinal and extraintestinal *Escherichia coli*. *Front. Microbiol.* **11**, 2065 (2020).
60. Biggel, M. et al. Horizontally acquired *papGII*-containing pathogenicity islands underlie the emergence of invasive uropathogenic *Escherichia coli* lineages. *Nat. Commun.* **11**, 5968 (2020).
61. Sakellaris, H. et al. Regulated site-specific recombination of the *she* pathogenicity island of *Shigella flexneri*. *Mol. Microbiol.* **52**, 1329–1336 (2004).
62. Manson, J. M. & Gilmore, M. S. Pathogenicity island integrase cross-talk: a potential new tool for virulence modulation. *Mol. Microbiol.* **61**, 555–559 (2006).
63. Wang, C. et al. Alpha-hemolysin of uropathogenic *Escherichia coli* induces GM-CSF-mediated acute kidney injury. *Mucosal Immunol.* **13**, 22–33 (2020).
64. Khan, N. A. et al. Cytotoxic Necrotizing factor-1 contributes to *Escherichia coli* K1 invasion of the central nervous system. *J. Biol. Chem.* **277**, 15607–15612 (2002).
65. Guyer, D. M., Henderson, I. R., Nataro, J. P. & Mobley, H. L. Identification of Sat, an autotransporter toxin produced by uropathogenic *Escherichia coli*. *Mol. Microbiol.* **38**, 53–66 (2000).
66. Dobrindt, U. et al. S-Fimbria-encoding determinant *sfa*<sub>5</sub> is located on pathogenicity island III<sub>536</sub> of uropathogenic *Escherichia coli* strain 536. *Infect. Immun.* **69**, 4248–4256 (2001).
67. Rashid, R. A., Tarr, P. I. & Moseley, S. L. Expression of the *Escherichia coli* IrgA homolog adhesin is regulated by the ferric uptake regulation protein. *Microb. Pathog.* **41**, 207–217 (2006).

68. Carbonetti, N. H. et al. Aerobactin-mediated iron uptake by *Escherichia coli* isolates from human extraintestinal infections. *Infect. Immun.* **51**, 966–968 (1986).
69. Zelmer, A. et al. Administration of capsule-selective endosialidase E minimizes upregulation of organ gene expression induced by experimental systemic infection with *Escherichia coli* K1. *Microbiology* **156**, 2205–2215 (2010).
70. Wooster, D. G., Maruvada, R., Blom, A. M. & Prasadarao, N. V. Logarithmic phase *Escherichia coli* K1 efficiently avoids serum killing by promoting C4bp-mediated C3b and C4b degradation. *Immunology* **117**, 482–493 (2006).
71. Pelkonen, S., Aalto, J. & Finne, J. Differential activities of bacteriophage depolymerase on bacterial polysaccharide: binding is essential but degradation is inhibitory in phage infection of K1-defective *Escherichia coli*. *J. Bacteriol.* **174**, 7757–7761 (1992).
72. Nesta, B. & Pizza, M. Vaccines against *Escherichia coli*. *Curr. Top. Microbiol. Immunol.* **416**, 213–242 (2018).
73. Maciejewska, B., Olszak, T. & Drulis-Kawa, Z. Applications of bacteriophages versus phage enzymes to combat and cure bacterial infections: an ambitious and also a realistic application? *Appl. Microbiol. Biotechnol.* **102**, 2563–2581 (2018).
74. Kern, W. V. & Rieg, S. Burden of bacterial bloodstream infection—a brief update on epidemiology and significance of multidrug-resistant pathogens. *Clin. Microbiol. Infect.* **26**, 151–157 (2020).
75. Temkin, E. et al. Estimating the number of infections caused by antibiotic-resistant *Escherichia coli* and *Klebsiella pneumoniae* in 2014: a modelling study. *Lancet Glob. Health* **6**, e969–e979 (2018).
76. Health Organization, W. WHO priority pathogens list for R&D of new antibiotics. (World Health Organization, Geneva, Switzerland). (2017)
77. Rodríguez, I. et al. A 21-year survey of *Escherichia coli* from bloodstream infections (BSI) in a tertiary hospital reveals how community-hospital dynamics of B2 phylogroup clones influence local BSI rates. *mSphere* **6**, e0086821 (2021).
78. Douglas, N. M., Hennessy, J. N., Currie, B. J. & Baird, R. W. Trends in bacteremia over 2 decades in the top end of the Northern Territory of Australia. *Open Forum Infect. Dis.* **7**, ofaa472 (2020).
79. Uk, G. National statistics Annual epidemiological commentary: Gram-negative, MRSA, MSSA bacteraemia and *C. difficile* infections, up to and including financial year 2021 to 2022. (2022).
80. NORM-VET reports. <https://www.vetinst.no/en/surveillance-programmes/norm-norm-vet-report>.
81. Mäklin, T. et al. Strong pathogen competition in neonatal gut colonisation. *Nat. Commun.* **13**, 7417 (2022)
82. Devi, S. J. et al. Identity between polysaccharide antigens of *Moraxella nonliquefaciens*, group B *Neisseria meningitidis*, and *Escherichia coli* K1 (non-O acetylated). *Infect. Immun.* **59**, 732–736 (1991).
83. Puente-Polledo, L., Reglero, A., González-Clemente, C., Rodríguez-Aparicio, L. B. & Ferrero, M. A. Biochemical conditions for the production of polysialic acid by *Pasteurella haemolytica* A2. *Glycoconj. J.* **15**, 855–861 (1998).
84. Read, R. C. et al. The ( $\alpha$ 2 $\rightarrow$ 8)-linked polysialic acid capsule of group B *Neisseria meningitidis* modifies multiple steps during interaction with human macrophages. *Infect. Immun.* **64**, 3210–3217 (1996).
85. Scholl, D., Adhya, S. & Merrill, C. *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Appl. Environ. Microbiol.* **71**, 4872–4874 (2005).
86. Haudiquet, M., Buffet, A., Rendueles, O. & Rocha, E. P. C. Interplay between the cell envelope and mobile genetic elements shapes gene flow in populations of the nosocomial pathogen *Klebsiella pneumoniae*. *PLoS Biol.* **19**, e3001276 (2021).
87. Holt, K. E., Lassalle, F., Wyres, K. L., Wick, R. & Mostowy, R. J. Diversity and evolution of surface polysaccharide synthesis loci in *Enterobacteriales*. *ISME J.* **14**, 1713–1730 (2020).
88. AbuOun, M. et al. A genomic epidemiological study shows that prevalence of antimicrobial resistance in *Enterobacteriales* is associated with the livestock host, as well as antimicrobial usage. *Microb Genom* **7**, 000630 (2021).
89. Roberts, I. S. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* **50**, 285–315 (1996).
90. Scholl, D. & Merrill, C. The genome of bacteriophage K1F, a T7-like phage that has acquired the ability to replicate on K1 strains of *Escherichia coli*. *J. Bacteriol.* **187**, 8499–8503 (2005).
91. Cieslewicz, M. & Vimr, E. Thermoregulation of *kpsF*, the first region 1 gene in the *kps* locus for polysialic acid biosynthesis in *Escherichia coli* K1. *J. Bacteriol.* **178**, 3212–3220 (1996).
92. King, J. E., Aal Owaif, H. A., Jia, J. & Roberts, I. S. Phenotypic heterogeneity in expression of the K1 polysaccharide capsule of uropathogenic *Escherichia coli* and downregulation of the capsule genes during growth in urine. *Infect. Immun.* **83**, 2605–2613 (2015).
93. Petty, N. K. et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc. Natl Acad. Sci. USA* **111**, 5694–5699 (2014).
94. Lipworth, S. et al. Ten-year longitudinal molecular epidemiology study of *Escherichia coli* and *Klebsiella* species bloodstream infections in Oxfordshire, UK. *Genome Med.* **13**, 144 (2021).
95. Royer, G. et al. Phylogroup stability contrasts with high within sequence type complex dynamics of *Escherichia coli* bloodstream infection isolates over a 12-year period. *Genome Med.* **13**, 77 (2021).
96. Yamaji, R., Rubin, J., Thys, E., Friedman, C. R. & Riley, L. W. Persistent pandemic lineages of uropathogenic *Escherichia coli* in a college community from 1999 to 2017. *J. Clin. Microbiol.* **56**, e01834-17 (2018).
97. Forde, B. M. et al. Discovery of  $\alpha$ -mediated colistin resistance in a highly virulent *Escherichia coli* lineage. *mSphere* **3**, e00486-18 (2018).
98. Blaser, M. J. Antibiotic use and its consequences for the normal microbiome. *Science* **352**, 544–545 (2016).
99. Panwar, R. B., Sequeira, R. P. & Clarke, T. B. Microbiota-mediated protection against antibiotic-resistant pathogens. *Genes Immun.* **22**, 255–267 (2021).
100. Young, V. B. & Schmidt, T. M. Antibiotic-associated diarrhea accompanied by large-scale alterations in the composition of the fecal microbiota. *J. Clin. Microbiol.* **42**, 1203–1206 (2004).
101. Baggs, J. et al. Risk of subsequent sepsis within 90 days after a hospital stay by type of antibiotic exposure. *Clin. Infect. Dis.* **66**, 1004–1012 (2018).
102. Loeffler, J. M., Nelson, D. & Fischetti, V. A. Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. *Science* **294**, 2170–2172 (2001).
103. Nelson, D., Loomis, L. & Fischetti, V. A. Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. *Proc. Natl Acad. Sci. USA* **98**, 4107–4112 (2001).
104. Lood, R. et al. Novel phage lysin capable of killing the multidrug-resistant gram-negative bacterium *Acinetobacter baumannii* in a mouse bacteremia model. *Antimicrob. Agents Chemother.* **59**, 1983–1991 (2015).
105. Blundell-Hunter, G. et al. Characterisation of bacteriophage-encoded depolymerases selective for key *Klebsiella pneumoniae* capsular exopolysaccharides. *Front. Cell. Infect. Microbiol.* **11**, 686090 (2021).

106. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
107. Solano, C. et al. Genetic reductionist approach for dissecting individual roles of GGDEF proteins within the c-di-GMP signaling network in *Salmonella*. *Proc. Natl Acad. Sci. USA* **106**, 7997–8002 (2009).
108. Jokilampi, A. et al. Construction of antibody mimics from a non-catalytic enzyme-detection of polysialic acid. *J. Immunol. Methods* **295**, 149–160 (2004).
109. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
110. Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E. & Clermont, O. ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb. Genom.* **4**, e000192 (2018).
111. Jolley, K. A. & Maiden, M. C. J. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinforma.* **11**, 595 (2010).
112. Argimón, S. et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* **2**, e000093 (2016).
113. Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R. & Wilson, D. J. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* **46**, e134 (2018).
114. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
115. Arredondo-Alonso, S. et al. A high-throughput multiplexing and selection strategy to complete bacterial genomes. *Gigascience* **10**, giab079 (2021).
116. Zhou, Z. et al. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. *Genome Res.* **30**, 138–152 (2020).
117. Hunt, M. et al. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* **16**, 294 (2015).
118. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
119. Tonkin-Hill, G. et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
120. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
121. Gilchrist, C. L. M. & Chooi, Y.-H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).
122. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
123. Feldgarden, M. et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728 (2021).
124. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 524 (2014).
125. Krijthe, J. H. Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation. *R package version 0.13*, <https://github.com/jkrijthe> (2015).
126. Arredondo-Alonso, S. et al. Complete chromosomal sequences from the main clonal complexes of *E. coli*. <https://doi.org/10.6084/m9.figshare.21610674> (2023).

## Acknowledgements

We thank David A.B. Dance (University of Oxford) for assistance in collating the strain collection. We thank the Norwegian laboratories for establishing the NORM collection through the Norwegian Surveillance System on Resistant Microbes. This work was financed by the Medical Research Council (MRC, UK) through grants MR/K018396/1, MR/N012542/1 and MR/R009937/1 to P.W.T.; the Biotechnology and Biological Sciences Research Council (BBSRC, UK) grant BB/V006495/1 to A.J.M.; the Wellcome Trust grant 206194 to N.R.T., by the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions (grant No. 801,133 to S.A.-A. and A.K.P.). This work received funding from the Trond Mohn Foundation (grant identifier TMS2019TMT04 to A.K.P., R.A.G., Ø.S., P.J.J., and J.C.). This work was partially supported by the European Research Council (grant No. 742,158 to J.C.), Joint Programming Initiative in Antimicrobial Resistance (JPIAMR Third call, STARCS, JPIAMR2016-AC16/OO039) to S.A.-A. and Engineering and Physical Sciences Research Council (EPSRC, UK) grant EP/X026671/1 to J.R.P.

## Author contributions

These authors contributed equally: George Blundell-Hunter, Zuyi Fu, Rebecca A. Gladstone. S.A.-A. and R.A.G. performed the bioinformatic, genomic analyses and worked on the manuscript writing. E.C.-G., P.J.J., Ø.S., A.K.P., F.C., S.C.-B., M.A.C., M.A.A., M.V., A.C., C.H., N.K., A.M., and J.N.R. sampled isolates. G.B.-H., Z.F., A.F.-S., J.L., and A.J.M. conducted the experimental work. A.J.M., P.W.T., Ø.S., P.J.J., J.C., J.R.P., and N.R.T. acquired the funding. S.A.-A., A.J.M., P.W.T., and J.C. designed the study and contributed to the analyses. S.A.-A., A.J.M., and J.C. wrote the original draft of the manuscript and revised the manuscript with input from all authors. All authors read and approved the final paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-39052-w>.

**Correspondence** and requests for materials should be addressed to Jukka Corander, Peter W. Taylor or Alex J. McCarthy.

**Reprints and permissions information** is available at <http://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) 2023



<sup>1</sup>Department of Biostatistics, University of Oslo, 0317 Oslo, Norway. <sup>2</sup>Parasites and Microbes, Wellcome Sanger Institute, Cambridge, UK. <sup>3</sup>School of Pharmacy, University College London, London, UK. <sup>4</sup>Department of Infectious Disease, Centre for Bacterial Resistance Biology, Imperial College London, London, UK. <sup>5</sup>Great Ormond Street Institute of Child Health, University College London, London, UK. <sup>6</sup>Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway. <sup>7</sup>Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway. <sup>8</sup>University of Missouri Kansas City, Kansas City, USA. <sup>9</sup>Division of Infectious Diseases, Children's Mercy Hospital Kansas City, UMKC School of Medicine, Kansas City, USA. <sup>10</sup>Unidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias, Hospital de Pediatría, Centro Médico Nacional Siglo XXI Instituto Mexicano del Seguro Social, Mexico City, Mexico. <sup>11</sup>Facultad de Medicina, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico. <sup>12</sup>Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit (LOMWRU), Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao PDR. <sup>13</sup>Faculty of Medicine, Chair of Microbiology, Jagiellonian University Medical College, Czyska str. 18, 31-121, Kraków, Poland. <sup>14</sup>British Society for Antimicrobial Chemotherapy, Birmingham, UK. <sup>15</sup>Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK. <sup>16</sup>Laboratory of Pathology and Molecular Biology (LPBM), Gonçalo Moniz Research Institute, Oswaldo Cruz Foundation, Salvador, Brazil. <sup>17</sup>Faculdade de Farmácia, Universidade Federal da Bahia, Salvador, Brazil. <sup>18</sup>London School of Hygiene and Tropical Medicine, London, UK. <sup>19</sup>Helsinki Institute of Information Technology, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland. <sup>20</sup>These authors contributed equally: George Blundell-Hunter, Zuyi Fu, Rebecca A. Gladstone. ✉ e-mail: [jukka.corander@medisin.uio.no](mailto:jukka.corander@medisin.uio.no); [peter.taylor@ucl.ac.uk](mailto:peter.taylor@ucl.ac.uk); [a.mccarthy@imperial.ac.uk](mailto:a.mccarthy@imperial.ac.uk)