nature communications

Article

https://doi.org/10.1038/s41467-025-58106-9

Emergence and global spread of a dominant multidrug-resistant clade within *Acinetobacter baumannii*

Received: 5 April 2024

Accepted: 12 March 2025

Published online: 21 March 2025

Check for updates

Shengkai Li[1,2,10], Guilai Jiang[1,3,10], Shengke Wang[2,10], Min Wang[4,10], Yilei Wu[1,5,10], Jinzhi Zhang[6], Xiao Liu[1,3], Ling Zhong[1], Min Zhou[7], Shichang Xie[1,8], Yi Ren[8], Ping He[7], Yongliang Lou[2] ✉, Heng Li[1,3] ✉, Jimei Du[2] ✉ & Zhemin Zhou[1,3,4,9] ✉

The proliferation of multi-drug resistant (MDR) bacteria is driven by the global spread of epidemic lineages that accumulate antimicrobial resistance genes (ARGs). *Acinetobacter baumannii*, a leading cause of nosocomial infections, displays resistance to most frontline antimicrobials and represents a significant challenge to public health. In this study, we conduct a comprehensive genomic analysis of over 15,000 *A. baumannii* genomes to identify a predominant epidemic super-lineage (ESL) accounting for approximately 70% of global isolates. Through hierarchical classification of the ESL into distinct lineages, clusters, and clades, we identified a stepwise evolutionary trajectory responsible for the worldwide expansion and transmission of *A. baumannii* over the last eight decades. We observed the rise and global spread of a previously unrecognized Clade 2.5.6, which emerged in East Asia in 2006. The epidemic of the clade is linked to the ongoing acquisition of ARGs and virulence factors facilitated by genetic recombination. Our results highlight the necessity for One Health-oriented research and interventions to address the spread of this MDR pathogen.

Bacteria with multidrug resistance (MDR) present a new form of pathogenic challenge in treatment, resulting in the majority of nosocomial infections[1]. *Acinetobacter baumannii* is particularly problematic, with nearly all clinical strains exhibiting a broad range of resistance to multiple drugs, disinfectants, and environmental stresses. In particular, Carbapenem-resistant *A. baumannii* (CRAB) exhibited 68.8–99.8% resistance to all antimicrobials except colistin and tigecycline[2], complicating treatment strategies. The increasing prevalence of CRAB has led to an estimated 326,000 deaths in 2019[3], underscoring the pressing need for effective control measures.

Several international clones (ICs) of *A. baumannii* have been identified through restriction fragment length polymorphism[4] and multi-locus sequence typing (MLST)[5–7]. The most prominent ICs are IC2, associated with the majority of nosocomial infections

[1]MOE Key Laboratory of Geriatric Diseases and Immunology, Cancer Institute, Suzhou Medical College, Soochow University, Suzhou, China. [2]Key Laboratory of Laboratory Medicine, Ministry of Education, School of Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou, China. [3]Jiangsu Province Engineering Research Center of Precision Diagnostics and Therapeutics Development, Soochow University, Suzhou, China. [4]Department of Clinical Laboratory, The Second Affiliated Hospital of Soochow University, Suzhou, China. [5]Department of Life Sciences, Imperial College London, London, UK. [6]Department of Critical Care Medicine, Zhejiang Provincial People's Hospital, Hangzhou, China. [7]Department of Immunology and Microbiology, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [8]Iotabiome Biotechnology Inc., Suzhou, China. [9]National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China. [10]These authors contributed equally: Shengkai Li, Guilai Jiang, Shengke Wang, Min Wang, Yilei Wu. ✉e-mail: lyl@wmu.edu.cn; hli@suda.edu.cn; djm@wmu.edu.cn; zmzhou@suda.edu.cn

internationally[8], and IC1, responsible for infection outbreaks in casualties returned from the Middle East conflicts[9]. Furthermore, IC5 accounts for >50% of infections in Latin America[10,11]. The resistome, virulome, and epidemiology of many ICs have also been investigated[12,13]. However, the detailed population structures and temporal dynamics within each of these ICs remain poorly characterized due to the limited resolution of MLST[14]. Moreover, the widespread presence of recombination in *A. baumannii* has influenced its genomic diversification and ARG distributions[15], complicating phylogenetic and functional analyses. A wealth of genomic data for *A. baumannii* has been gathered through surveillance and molecular epidemiology efforts. Nevertheless, the clinical interpretation of genomic data has been impeded by an incomplete understanding of its genetic context.

In this work, we exploit the genetic landscape of over 15,000 *A. baumannii* from around the world. Our analysis revealed their source and geographic distribution and ARG profiles, indicating different genetic reservoirs for strains of clinical and non-clinical settings. Furthermore, we identified an epidemic super-lineage (ESL) accounting for ~70% of existing clinical *A. baumannii* strains. Further characterization of the population structure of the ESL uncovered a cryptic, emerging Clade 2.5.6 with a global prevalence over the past 16 years. We showed that the emergence of Clade 2.5.6 was driven by the progressive acquisition of ARGs and recombination-mediated changes in virulence factors from diverse *A. baumannii* populations, calling for a One Health approach to research and intervention to manage the spread of this pathogen.

Results

The genetic landscape of *A. baumannii*

The ESKAPE pathogens were each categorized into 1090–8329 distinct sequence types (STs) using MLST schemes (Feb. 2024) from pubMLST or Institut Pasteur[16]. We estimated the genetic diversities of these pathogens based on the relative frequencies of the STs and found that *A. baumannii* displayed a notably lower Simpson diversity index (SDI) of 0.81, in contrast to the SDI range of 0.95–0.99 observed in other ESKAPE members (Fig. 1a). Detailed investigation attributed its low diversity to the prevalence of ST2, which constituted 43.7% of its strains—a significant proportion compared to other ESKAPE pathogens, where the most common STs accounted for only 4.5% (ST235 in *Pseudomonas aeruginosa*) to 12.4% (ST22 in *Staphylococcus aureus*) of strains (Fig. 1b). Notably, if all ST2 strains were removed, *A. baumannii* would have a comparable genetic diversity (SDI: 0.97) to other ESKAPE members. Such exceptional predominance of ST2 was unlikely a sampling bias because it has been reported in 109 countries worldwide, calling for further investigations.

To investigate the genetic landscape of *A. baumannii* at the genomic level, we sequenced 100 clinical strains from Suzhou and Wenzhou, East China, and integrated this data with 15,643 publicly available genomes (Oct. 2022). This compilation spanned strains from 88 countries across 110 years (Table 1 and Supplementary Data 1). MLST analysis using Pasteur's scheme identified 518 STs among these genomes, with ST2 (57.0%) still being the most common. We utilized eBURST to cluster the STs into 371 clonal complexes (CCs), including eleven that had been previously designated as ICs[7,17–22]. Notably, IC2 (CC2) encompassed ST2 and its single-locus variants, constituting ~65% of all genomes.

Among the 8561 strains with isolation sources, only 162 were from non-clinical settings (Supplementary Fig. 1a), suggesting a strong sampling bias toward clinical samples. Interestingly, strains from the IC2 and other ICs only accounted for 29 (17.9%) and 10 (6.2%) of the non-clinical samples (Supplementary Fig. 1b), respectively, indicating distinct genetic contexts between strains in clinical or non-clinical settings.

Phylogeny and resistome of *A. baumannii*

We selected 5824 genomes with pairwise average nucleotide identities (ANIs) of <99.5% to establish a representative set of *A. baumannii* and identified 2266 softcore genes shared by at least 95% (5533/5824) of the representatives (Supplementary Data 2). A maximum-likelihood (ML) tree was constructed for each core gene, and these trees were summarized into a supertree that unveiled the comprehensive phylogenetic structure of *A. baumannii* (Fig. 1c) (see "Methods"). The majority of ICs and CCs showed strong correlations with the supertree, with each primarily linked to a single lineage (Fig. 1c). Remarkably, the top two ICs (IC2 and IC1) were genetically closely related, forming a monophyletic clade in the supertree. This finding reveals a larger-scale population structure above the CCs in *A. baumannii*, which we designated as a "super-lineage" to differentiate it from the lineages formed by CCs.

We also investigated the resistome of *A. baumannii*. The results revealed that each *A. baumannii* strain harbored an average of 15.1 predicted ARGs, conferring resistance to 6.1 drug categories. The most prevalent predicted resistances were aminoglycoside (98.6%), fosfomycin (94.1%), quinolone (87.7%), and sulfonamide (75.4%) (Table 1). In contrast, resistances to tigecycline and colistin were relatively low and were only predicted in 30 and 5 strains, respectively (Supplementary Data 1). Notably, the average number of ARGs per strain and their associated drug types have increased by 50–100% over the past two decades (Fig. 1f), implying strong selective stress imposed by antimicrobial usage. Such a rapid change in population dynamics could be driven by either universal acquisition of ARGs in all lineages or, more plausible, expansion of one or multiple epidemic lineages with high levels of ARGs.

A significant correlation was observed between the international prevalence of IC/CCs and the average number of predicted ARGs per strain ($r = 0.51$) (Fig. 1g), as well as their resistance to carbapenem ($r = 0.42$), minocycline ($r = 0.65$), and sulbactam ($r = 0.38$) (Supplementary Fig. 1c). In addition to the ICs, we predicted three major CCs (≥20 genomes) with averagely ≥15 ARGs and five major CCs with ≥80% carbapenem resistance (Supplementary Data 3). Strains within IC2 encoded an average of 17.1 ARGs, with over two-thirds predicted to resist carbapenem, sulbactam, and minocycline (Table 1). Furthermore, IC1 strains encoded ~15.3 ARGs, with 82.0% resistance to carbapenem, but <20% resistance to minocycline or sulbactam (Table 1). Both IC1 and IC2 were found across 60–66 countries, in contrast to the other nine ICs, which were isolated from only 10 to 35 countries (Fig. 1d and Supplementary Fig. 1d). Consequently, the super-lineage formed by IC2 and IC1 accounted for a total of 10,820 (69%) strains from 79 countries (Table 1), explaining 82.5% of the ARGs in *A. baumannii* and contributing significantly to the escalation of ARGs over the past two decades. Furthermore, 99 of the 100 sequenced clinical isolates belong to IC2, demonstrating its predominance in China (Supplementary Data 4). Henceforth, we refer to IC1 and IC2 collectively as the ESL and focus on exploring its population dynamics below.

High level of recombination in ESL

The ESL accounted for 65% (Africa and North America) to 90% (East Asia and Oceania) of the isolates from every geographic region except South America, where IC5 (CC79) and IC7 (CC25) were predominant (Fig. 1d). High frequencies of recombination were detected within ESL, contributing to >90% of its sequence variations (Supplementary Fig. 2a, b). Notably, the recombination rate was unevenly distributed across the chromosome, dividing it into six distinct blocks with varied recombination frequencies (Fig. 2a and Supplementary Data 5): two high-recombination blocks (HRBs) of 1.23 million bases (MB), two middle-recombination blocks (MRBs) of 0.91 MB, and one large low-recombination block (LRB) exceeding 1.60 MB. Despite the removal of recombinant SNPs, 6–16% of homoplastic SNPs persisted in each block, leading to the formation of topologically distinct phylogenetic trees

https://doi.org/10.1038/s41467-025-58106-9

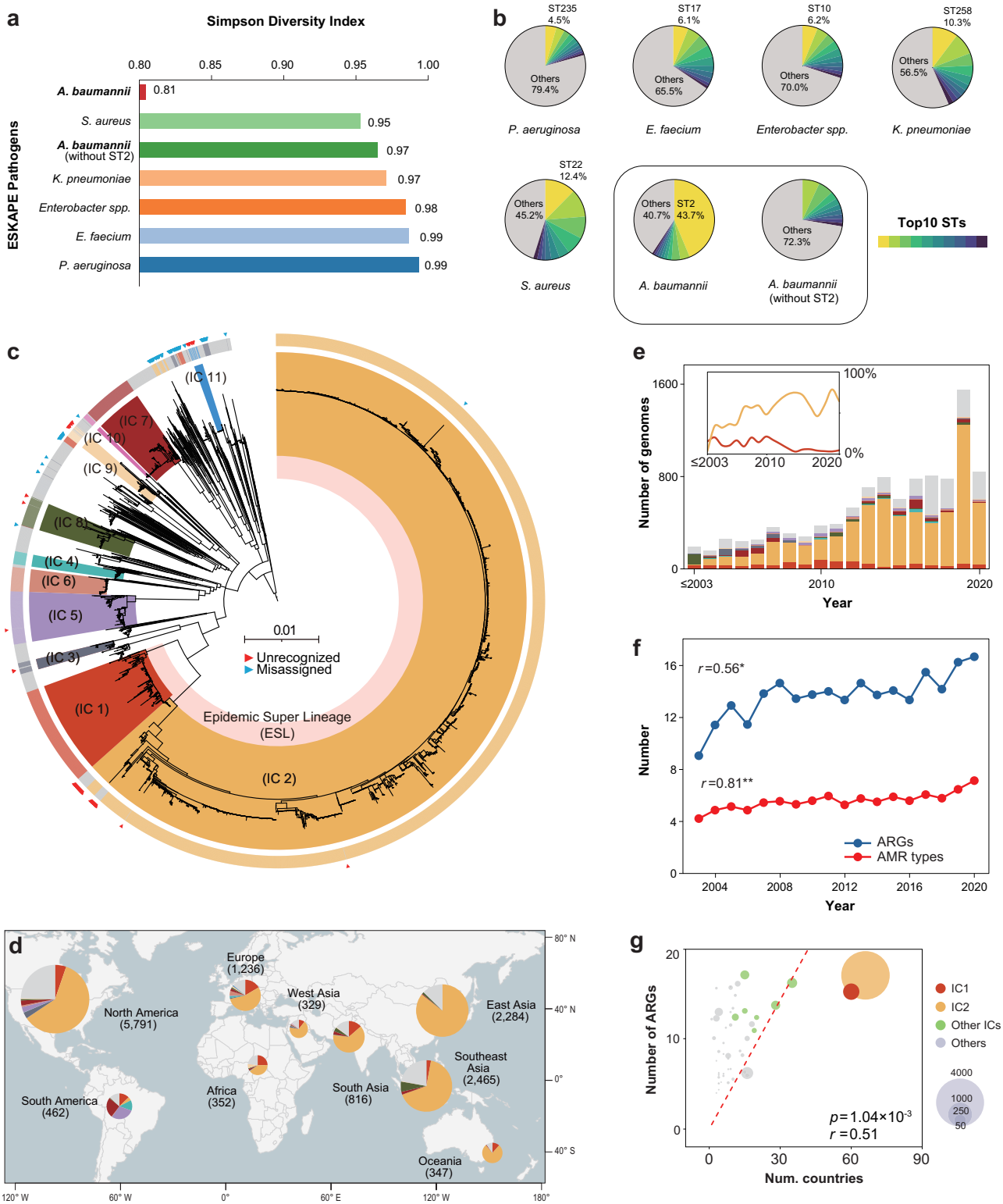

Fig. 1 | Population structure and predicted antimicrobial resistance genotypes of 15,743 *A. baumannii* genomes. a, **b** Genetic diversity of the ESKAPE bacteria by their STs. *A. baumannii* is divided into two groups, with or without ST2. **a** The bar chart shows the Simpson diversity index (SDI) of STs for each ESKAPE member. **b** The pie chart of the top 10 STs and the others for each ESKAPE member. **c** The supertree of *A. baumannii* based on 2266 softcore genes in the 5824 representative genomes. The highlighted clusters in the tree showed the most compatible IC/CC for each of the nine previously designated international clones (ICs) in *A. baumannii*, as shown in the labels. The ICs were predicted using Pasteur's MLST scheme as clonal complexes (CCs), which were shown in the outer ring. Genomes that had inconsistent IC assignments between the supertree and the MLST were marked with triangles, with the false positives ($n = 121$) and the false negatives ($n = 89$) color-coded in blue and red, respectively. **d** Global distributions of the IC1 and IC2 strains, color-coded as in (**a**). **e** Histogram of the yearly isolation frequencies of IC1, IC2, and the others. The inset shows the relative frequencies of IC1 and IC2 each year. **f** The yearly variations of the numbers of AMR genes (blue) and categories of drug resistances (red) between 2003 and 2020. The correlation coefficient of linear regression for each number series was shown nearby. **g** Bubble plots of the numbers of country presences (*X*-axis) versus the numbers of ARG carriages (*Y*-axis) for the IC/CCs. The slope of linear regression (r) is 0.51 and the significance (p) is calculated using the two-tailed student's t-test. The map was modified from open-source data in [https://github.com/antvis/L7].

Table 1 | Summarized information for selected IC/CCs

International Clone (IC)	IC1(CC1)	IC2(CC2)	ESL	Other ICs	Non-ICs	All
Number of Isolates	1094	9726	10,820	1492	3331	15,643
Years of isolation						
Before 2000	21	16	37	12	57	106
2001–2010	265	912	1177	295	463	1935
2011–2020	462	4739	5201	680	1415	7296
2021–	46	1831	1877	75	493	2445
ARGs						
Non OXA-51 family Carbapenemase	81.9%	83.6%	83.4%	69.5%	35.8%	72.0%
Minocycline	18.7%	72.4%	67.0%	23.2%	8.0%	50.2%
Sulbactam	17.2%	93.6%	85.9%	20.9%	5.4%	62.5%
Quinolone	97.2%	99.7%	99.4%	91.0%	47.9%	87.7%
Macrolide	27.2%	69.0%	64.8%	34.3%	17.6%	51.8%
Sulfonamide	90.0%	79.3%	81.3%	79.8%	57.0%	75.4%
Fosfomycin	92.5%	98.2%	97.6%	97.4%	81.5%	94.1%
Aminoglycoside	100.00%	100.00%	100%	100.00%	93.6%	98.6%
num ARGs	15.3	17.1	16.9	14.6	9.1	15.1
num AMR types	6.5	6.5	6.5	6.1	4.4	6.1

(Supplementary Fig. 2c). We found that the mutational phylogeny from the LRB was broadly consistent with the tree reconstructed from the whole core genome (Supplementary Fig. 2d), with a normalized Robinson–Foulds (NRF) distance of 0.1, whereas trees from other regions were all different (NRF distance > 0.8). Given these findings, we utilized the mutational phylogeny from the LRB, which encompassed 66.1% (23,524/35,609) of the mutational SNPs and exhibited the lowest level of homoplasy, as a representative of the ESL's phylogenetic history.

Global dissemination of the ESL

The phylogenetic analysis categorized ESL strains into 59 monophyletic clades (Fig. 3a and Supplementary Data 1). To gain insights into long-term evolutionary patterns, we further merged the closely related clades into 4 and 5 clusters within the IC1 and IC2 lineages, respectively.

Clusters and clades in IC1. IC1 was predominant in Europe and North America, and Cluster 1.4 was also identified in South, West, and Southeast Asia, giving rise to Clades 1.4.4, 1.4.9, and 1.4.10 (Supplementary Fig. 3). Notably, IC1 has only been found twice (GCF_002082825.1 in South Korea and GCF_002999195.1 in China) in East Asia and not present in our 100 Chinese strains. Strains within Asian Clades 1.4.5, 1.4.4, and 1.4.9 acquired additional ARGs, including *tet*(B), which confers resistance to minocycline in *A. baumannii*[23]. These strains subsequently spread to Europe, Africa, and the US, with some being associated with super-spreading events involving casualties from the Middle East conflicts in the 2000s[9]. The isolation frequency of IC1 strains peaked at 20% in the 2000s and has since declined sharply to less than 5% (Fig. 3b). Unfortunately, although some IC1 clusters have been previously dated, we failed to identify significant temporal signals for the entire lineage (Supplementary Fig. 4b, c), precluding the reconstruction of its temporal dynamics.

Global transmission of IC2 and the emergence of Clade 2.5.6. IC2 showed a weak, yet significant signal of SNP accumulation with time in the linear regression test ($r^2 = 0.087$, $p < 10^{-6}$; Supplementary Fig. 4b), allowing our inference of its spatiotemporal dynamics using TreeTime (Fig. 3d and Supplementary Fig. 4a). Moreover, the calculated substitution rate of $8.16 \pm 2.19 \times 10^{-8}$ per year was >63-fold higher than all ten estimated rates in the date-randomization test (Supplementary Fig. 4c), confirming the presence of a significant temporal signal.

Using TreeTime, we estimated the most recent common ancestor (tMRCA) of IC2 to have been present in Europe as Cluster 2.1 prior to 1943 (CI95% 1928–1959) (Fig. 3d). The bacteria were predicted to subsequently spread to Africa, West Asia, and North America, resulting in the formation of Cluster 2.2 before 1945 (CI95% 1932–1959) (Fig. 3e). Cluster 2.3 was predicted to reach East Asia, which later became the primary source for Clusters 2.4 and 2.5 before 1978 (CI95% 1971–1984). Notably, Clade 2.5.6, the most extensive clade within the ESL, was believed to have emerged from Cluster 2.5 in East Asia before 2006 (CI95% 2005–2008) and rapidly disseminated to over 45 countries within the subsequent 16 years (Fig. 3f). The East China strains sequenced in this study revealed the co-circulation of three primary clades of 2.5.6 (57 strains), 2.4.13 (21), and 2.4.6 (10) in China in 2023, highlighting the sustained prevalence of 2.5.6 since its emergence (Supplementary Fig. 6a, c). We also analyzed the effective population sizes (EPSs) for IC2 and its clusters, revealing that Clusters 2.4 and 2.5 experienced rapid growth in their EPSs, leading to a threefold increase in the EPS of the entire lineage (Fig. 3c). This growth was mirrored by a shift in the predominant strain sources, with Cluster 2.3 declining from 40% to 10% between 2008 and 2022, while Cluster 2.5 rose from 0% to 41% (Fig. 3b). These findings underscore the role of Cluster 2.5 as the primary driver of IC2's global expansion.

To further validate our analysis, we conducted additional analyses using BactDating and BEAST2, each with five distinct models, on the IC2 lineage (Supplementary Fig. 5a, b). All results exhibited high consistency with those from TreeTime. Specifically, the tMRCAs averaged 1935 (1908–1953) for BactDating and 1944 (1925–1961) for BEAST2 (Supplementary Fig. 5c). Similarly, all runs consistently placed the origin of Clade 2.5.6 around 2006 (CI 2002–2008) (Supplementary Fig. 5d), corroborating the recent emergence of this clade.

Distinct genetic context found in clinical and community samples

To ease the use of the genotyping scheme, we developed a tool named Capybara (see "Methods") to automatically classify strains into lineages, clusters, and clades based on short sequencing reads, and assessed its efficacy using the 100 strains sequenced in this study (Fig. 4a, Supplementary Data 4), confirming consistent clade assignments.

https://doi.org/10.1038/s41467-025-58106-9

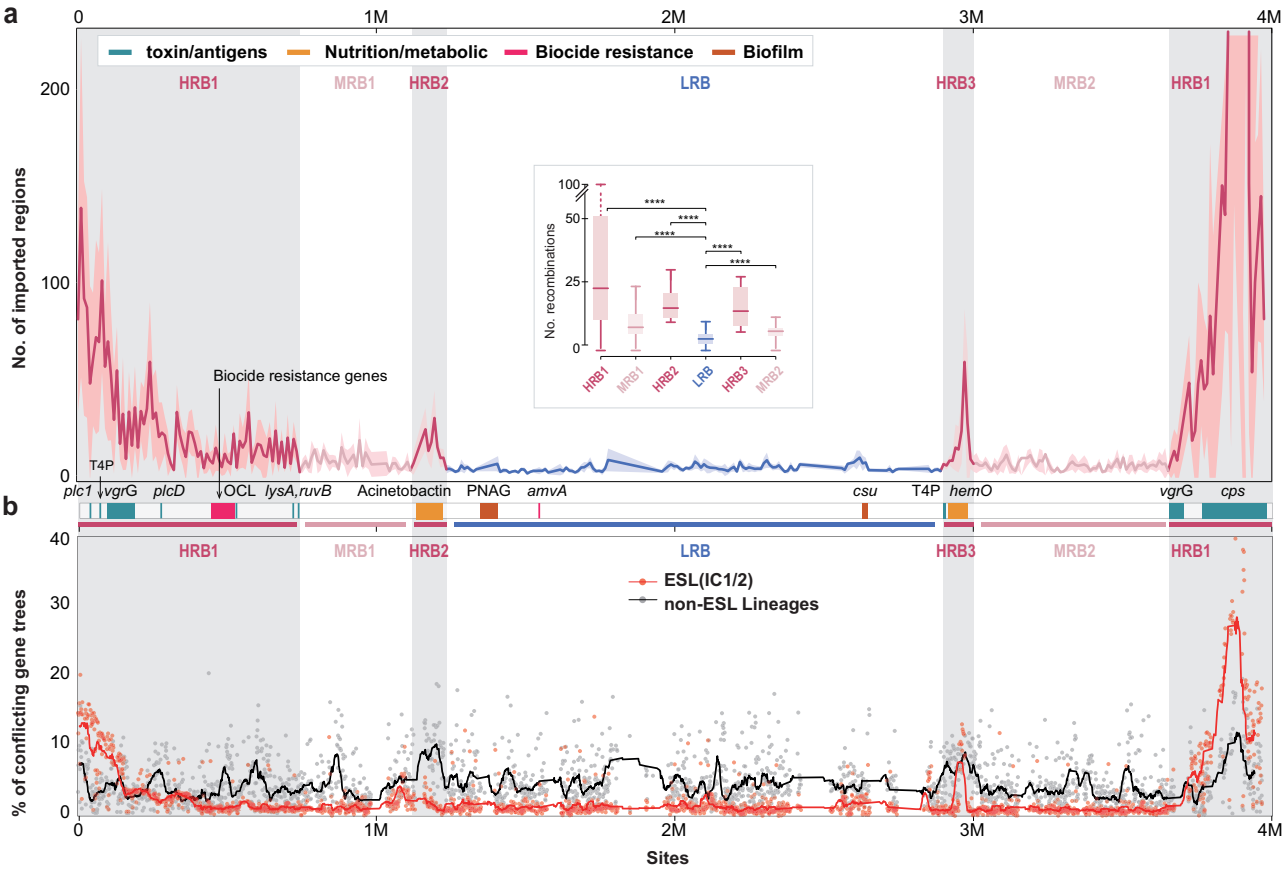

Fig. 2 | Frequencies of recombination events along the chromosome of the reference genome. a Frequencies of imported blocks per 10Kb (*Y*-axis) in windows sliding along coordinates in the chromosome (*X*-axis). The high-recombination blocks (HRBs) were predicted as blocks with >50 importation events (by RecHMM) in any sliding windows of 100 Kb, and the middle-recombination blocks (MRBs) and the low-recombination block (LRB) were predicted as gaps between HRBs. The median values of frequencies of recombination events (HRB1 = 24, HRB2 = 11, HRB3 = 18.5, MRB1 = 3, MRB2 = 13, and LRB = 5. Whiskers, minimum and maximum of the data) in each block were visualized as in the inset, with pairwise significance of differences evaluated using the two-tailed student's *t*-test (****$p < 0.0001$). **b** The scatter plot shows the levels of recombination in the 2266 core genes as measured by percentages of topological conflicts (*Y*-axis) in the 5824 representative genomes. The coordinates of the core genes in the reference genome were shown in the *X*-axis. The black and red dots show the level of recombination with or without the ESL genomes, and the lines show the average values of 10 neighboring core genes. The chromosomal coordinates of some genes encoding virulence factors, environmental adaptations, or biocide resistances were visualized as filled blocks between the two parts and color-coded according to the Key. Source data are provided as a Source Data file.

We then employed Capybara to investigate the genetic diversity of *A. baumannii* in metagenomic samples collected from clinical or community settings. To this end, we analyzed 514 metagenomes, either sequenced in-house or obtained publicly (Supplementary Data 4). *A. baumannii* was detected in 96 (18.7%) metagenomes, comprising 26 community samples and 70 clinical samples from patients hospitalized for over 7 days (Fig. 4b). Genotyping of these samples with Capybara revealed that 87.1% (61/70) of clinical samples belonged to the ESL, with 52 being Clade 2.5.6, while 96.2% (25/26) of community samples were non-ESL (Fig. 4c), consistent with our genome-based analysis that ESL was primarily prevalent in clinical settings.

Recombination as a driving force for emerging lineages in *A. baumannii*

Stepwise acquisition of ARGs in the ESL. Our investigation uncovered a gradual accumulation of carbapenemase genes along the ESL. Notably, the $bla_{OXA\text{-}23}$ gene was present in none to 52% of the strains in Clusters 2.1–2.3, whereas a significantly higher prevalence of 79% to >99% was observed in Clusters 2.4 and 2.5, respectively (Fig. 3d). To assess the role of mobile elements in this incremental acquisition of resistance, we evaluated 168 complete ESL genomes, finding $bla_{OXA\text{-}23}$ in 111 of them (Supplementary Fig. 8). Notably, only 23% (26/111) $bla_{OXA\text{-}23}$ was transferred by plasmids. In contrast, Tn/IS elements account for 93% (103/111) of $bla_{OXA\text{-}23}$ carriage, with the majority being Tn2007 and Tn6166. Furthermore, the instance of $bla_{NDM\text{-}1}$ has increased rapidly over the past 5 years, substantially contributing to carbapenem resistance in Clusters 1.4, 2.4, and 2.5, especially >60% in Clade 2.4.14. Moreover, nearly all ESL strains developed quinolone resistance through *gyrA* and *parC* mutations (Supplementary Data 1). Over 90% of strains in Clusters 2.4 and 2.5, and >60% in Clusters 2.3, contain *mph*(E) and *msr*(E) that confer resistance to macrolides. Additionally, *tet*(B), conferring minocycline resistance, is also nearly ubiquitous across Clusters 2.3–2.5. Many IC2 strains have developed resistance to β-lactamase inhibitors like sulbactam due to the presence of $bla_{TEM\text{-}1}$ gene[24], and the MRCA of Cluster 2.5 acquired the *ftsI*-A515V mutation that has been associated with resistance to sulbactam-durlobactam[25]. As a result, Clade 2.5.6 was predicted to exhibit high resistance to aminoglycosides (100% resistance), fluoroquinolones (100%), β-lactams (100%), carbapenems (99%), sulbactam (98%), sulfonamide (62%), and tetracyclines including minocycline (83%), leaving high susceptibility (>99%) to only tigecycline and polymyxin.

https://doi.org/10.1038/s41467-025-58106-9

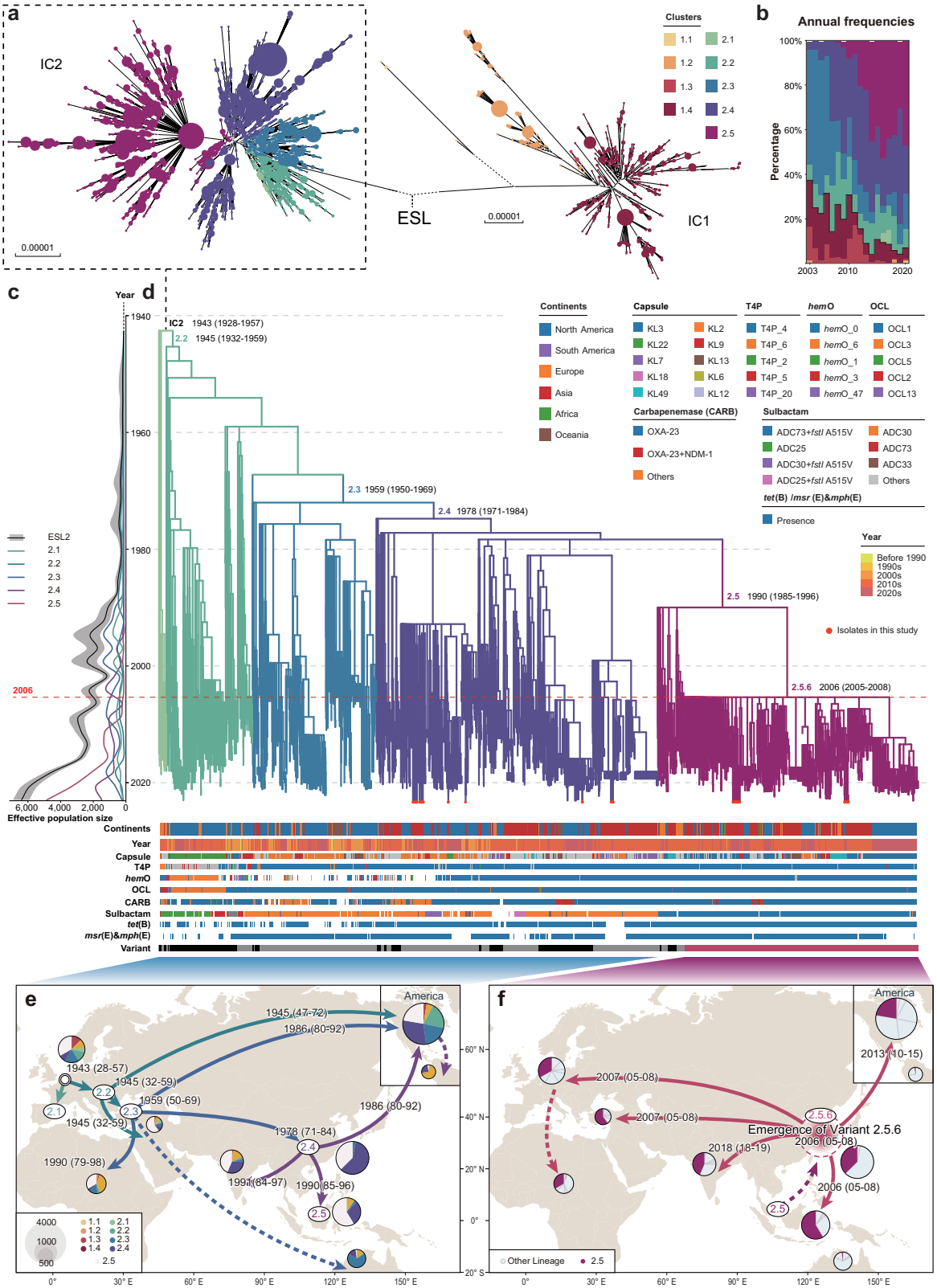

Fig. 3 | The phylogeny and spatiotemporal dynamics of ESL genomes. a The mutational phylogeny of the ESL based on 23,524 SNPs in the 1.6MB LRB (Fig. 2). The nine clusters of the ESL were shown as colored circles on the tree. The sizes of the circles were proportional to the number of associated genomes. **b** The relative frequencies of nine clusters in the ESL for each year between 2003 and 2020. **c** The estimated effective population sizes of the IC2 and all of its Clusters. The dashed line in 2006 shows the year of emergence for Clade 2.5.6. **d** The spatiotemporal tree by TreeTime. The branches were color-coded based on their clusters (as in the Key). The red dots represent the 99 Chinese IC2 strains sequenced in the study. Additionally, the allelic variations of three virulence factors (capsule, type 4 pilus, and *hem*O gene cluster) and four antimicrobial genes conferring resistances to carbapenem, sulbactam, tetracycline, and macrolide were also shown in the bottom, according to the Key. **e** The geographic distribution and major transmission patterns of Clusters 2.1–2.4 in IC2. **f** The geographic distribution and major transmission patterns of Cluster 2.5 with a focus on the rapid dissemination of Clade 2.5.6. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-025-58106-9

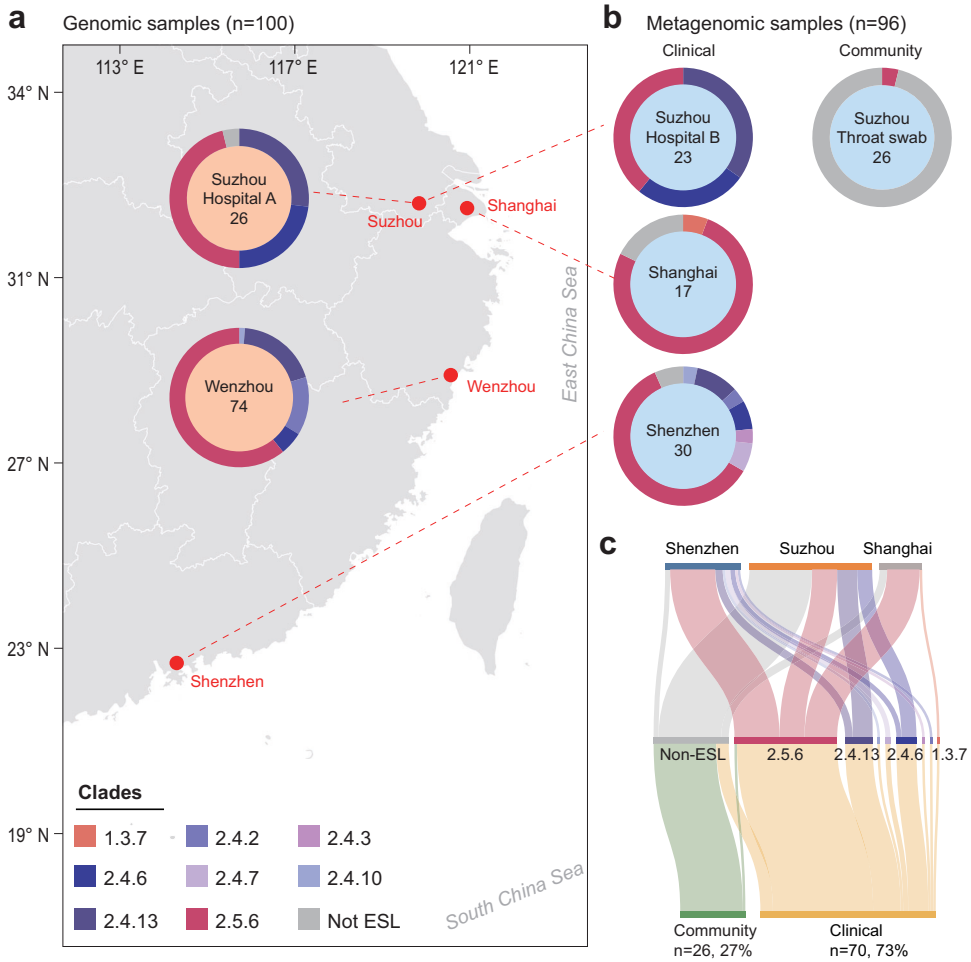

Fig. 4 | The prevalence of clades among *A. baumannii* in East China identified by Capybara. a, **b** The doughnut charts showed the different clades of *A. baumannii* in four different cities: Suzhou, Shanghai, Wenzhou, and Shenzhen. The colors of the inner circle represent the sequence method, **a** for short-reads sequenced data and **b** for metagenomes. **c** The Sankey plot showed the differentiated prevalence of each clade of *A. baumannii* from diverse sources and cities. Non-ESL *A. baumannii* strains were mostly identified in the community in Suzhou. The map was modified from open-source data in https://github.com/antvis/L7.

Virulence/adaptation genes in the HRBs. Our analysis indicates enrichment of virulence genes within the HRBs (Fig. 2a). These include genes related to lipooligosaccharide outer core (OCL), capsule (*cps*), exotoxin (*plc1* and *plcD*), and type 4 pili (T4P) in HRB1, as well as acinetobactin and *hemO* gene clusters in HRB2 and HRB3, which play crucial roles in iron uptake and host adaptation. Moreover, nine out of 14 genes identified by Li et al.[26] as associated with biocide resistances in *A. baumannii* are located in HRB1 (Supplementary Table 1). In contrast, genes linked to biofilm formation, such as the AdeFGH efflux pump (*amvA*), *csu* fimbriae, and β-(1->6)-Poly-N-acetyl-D-glucosamine synthesis, were more prevalent in the LRB. Notably, the type 6 secretion system (T6SS) main island, which encoded structural proteins, was located in the LRB, whereas the *vgrG* island that encoded T6SS effectors was situated in HRB1. These findings suggest a potential interplay between recombination and the evolution of virulence and host adaptation traits in *A. baumannii*.

Evolution of virulence genes in Cluster 2.5 through recombination. To elucidate the evolution of virulence factors in the HRBs, we employed Kaptive and Usearch to genotype the *cps*/OCL regions and other virulence genes across HRBs, respectively. We observed a stepwise acquisition of new virulence gene variants from other IC/CCs. For instance, an acinetobactin allele with <98% identity to the homolog in IC1 was incorporated before the MRCA of IC2. The original *hemO* allele (*hemO*_6) was lost in Clusters 2.2 and 2.3, with some Cluster 2.3 strains acquiring new *hemO* alleles >20 times (Supplementary Fig. 7b). However, none of these alleles persisted long-term until the *hemO*_1 from CC25 or CC164 was acquired before Cluster 2.4 and subsequently stabilized. Furthermore, IC2 underwent serotype conversions (Fig. 3d), such as the complete shift from OCL3 to OCL1 in the lipooligosaccharide after Cluster 2.2 and the change from the previous serotype K7 to a predominant K3 in Cluster 2.5, likely introduced from CC4, which was co-circulating with Cluster 2.5 in the far east (Fig. 3f and Supplementary Fig. 7c). These allelic changes along the IC2 in ESL phylogeny, as depicted in Fig. 5d, demonstrate a stepwise mechanism of gene acquisition.

Enhanced global dissemination of highly recombining IC/CCs. We evaluated the role of recombination at the species level by analyzing the frequency of topological conflicts among randomly selected quartets between gene trees and the supertree (refer to "Methods"). Increased topological conflicts were observed in HRB1 and HRB3 but not HRB2. Intriguingly, HRBs are distinctive for the ESL (Fig. 2b), as other IC/CCs do not show a significant increase in recombination events within these regions. To quantify the impact of recombination on each IC/CC, we introduced the Conflicting Quartet Per Gene Per strain (CQPP) metric. This analysis identified a group of highly recombining clonal complexes (hrCCs) with a CQPP > 1.5, a marked difference from the 0.3 to 0.5 range observed in the ESL (Fig. 5a).

https://doi.org/10.1038/s41467-025-58106-9

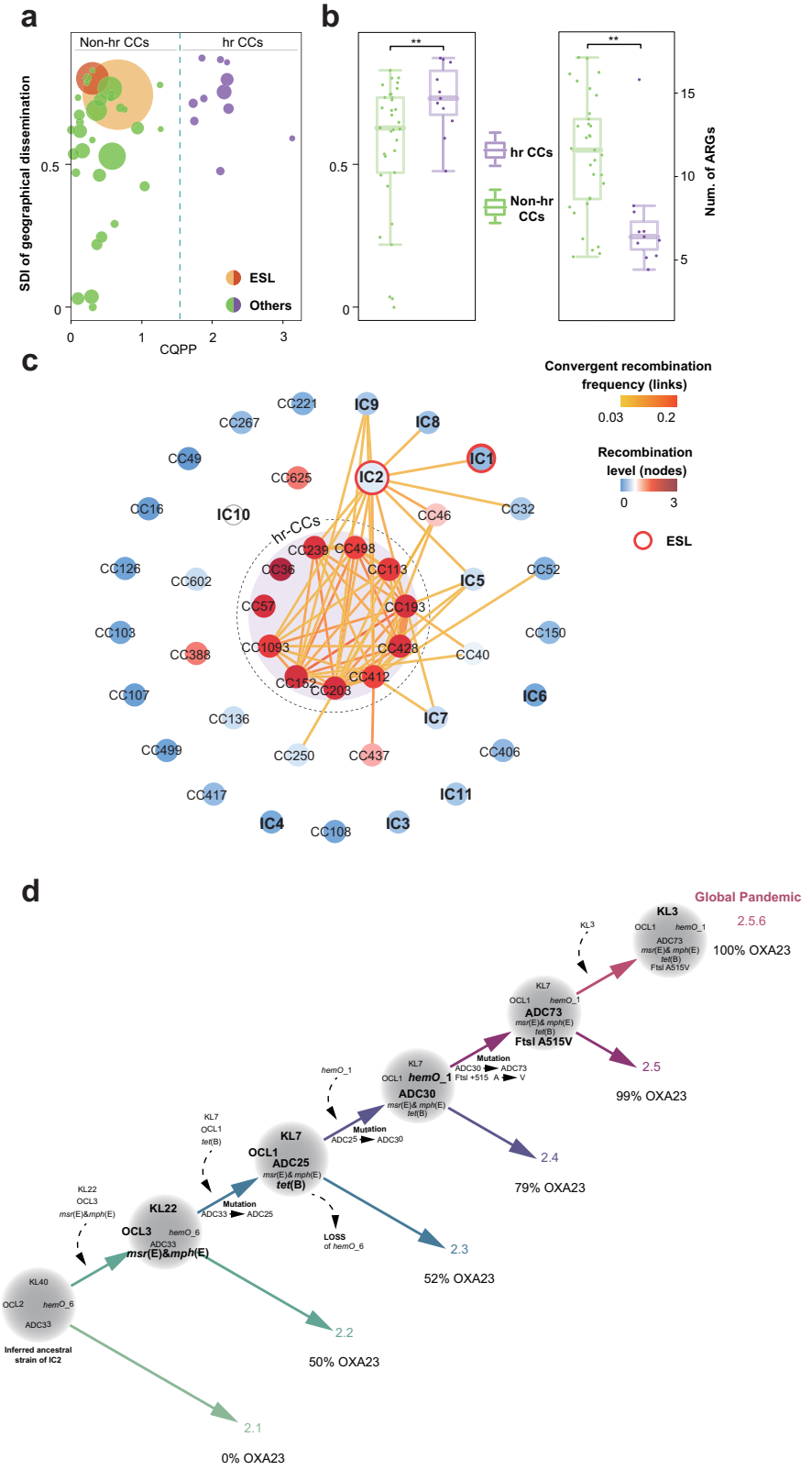

Fig. 5 | Recombination frequencies in *A. baumannii* populations and the model of stepwise acquisition of ARGs and virulence to Clade 2.5.6. a The bubble plot of estimated recombination levels (as conflicting quartets per gene per genome; CQPP; *X*-axis) versus the Simpson's diversity index (SDI; *Y*-axis) of the continental distributions for each IC/CC. The sizes of the bubbles are proportional to the number of genomes in each IC/CC. The dashed line in CQPP of 1.5 specifies the separation of high-recombination CCs (hrCCs, $n = 11$) and the remaining ($n = 30$). **b** The box plots compare SDIs (left) and ARG numbers (right) between hrCCs and other IC/CCs, showing 0.74 for hrCCs vs 0.55 for other IC/CCs in the left panel, and 7.5 for hrCCs vs 15.3 for other IC/CCs in the right panel (**$p < 0.01$; Wilcox-test). Whiskers, minimum, and maximum of the data. **c** The recombination network in *A. baumannii*. The nodes show major IC/CCs color-coded by their associated CQPP values, and the edges show frequencies of convergent recombination between two IC/CCs, measured by numbers of convergent genes per genome. The hrCCs were grouped in the middle and surrounded by a dotted circle. The IC2 and IC1 were highlighted by red halos. **d** The simplified pattern of stepwise evolution in the IC2. Source data are provided as a Source Data file.

Interestingly, these hrCCs do not correlate with higher ARG levels, as they exhibit >50% fewer ARGs (7.5 versus 15.3) compared to other lineages (Fig. 5b). Conversely, hrCCs are characterized by their broad geographic distribution and isolation from various continents, resulting in a higher average continental SDI of 0.74, compared to the SDI of 0.55 for other CCs ($p = 0.022$, Wilcox-test) (Supplementary Table 2). This suggests that the enhanced recombination ability of hrCCs may facilitate the acquisition of region-specific alleles from local populations, promoting their adaptation to a range of global environments. Notably, the reconstructed recombination network (Fig. 5c) further illustrates a pathway linking IC2 to these hrCCs through extensive gene exchange.

In summary, our analyses revealed three HRBs in the chromosome that are responsible for virulence factors involving iron uptakes, stress response, and host adaptation in *A. baumannii*. These regions have experienced ongoing gene conversions in the ESL, allowing the acquisition of diverse alleles from other CCs, and driving the diversification of this epidemic super-lineage.

Discussion

Acinetobacter baumannii, like many ESKAPE pathogens, is marked by the presence of MDR lineages that contribute to 100,000 nosocomial infections annually[27], alongside other lineages of lower clinical significance. The challenge of extracting and interpreting clinically relevant information from its genomes[15,28] is significant, often hindered by an incomplete understanding of the species' genetic context. In this study, we analyzed the genetic landscape of *A. baumannii* across >15,000 genomes, uncovering geographic and antimicrobial resistance patterns among its populations.

While at least eleven ICs have been identified in *A. baumannii*, their definitions have led to some confusion. Our analysis identified 109 CCs with strains from two or more countries, all potentially qualifying as ICs. Moreover, many ICs have not demonstrated a significantly higher carriage of ARGs or broader geographic distribution, casting doubt on their epidemiological significance (Fig. 1g and Supplementary Fig. 1c). For example, IC5 has been proposed as a major clone in some studies[10]. Our genomic data showed that it was rarely isolated outside of South America and encoded relatively lower levels of ARGs. In contrast, our analysis revealed seven non-IC populations that exhibited high levels of predicted ARGs (≥15) or carbapenem genes (≥80%), or both, which could be due to sampling bias but worth further investigation. Notably, IC1 and IC2 stand out for their high ARG carriage and widespread presence in ≥60 countries. Given their clinical importance and global prevalence, as evidenced by various studies[7,19,29], we suggest designating the clade comprising IC1 and IC2 as the ESL, setting it apart from populations with more localized distributions and fewer ARGs.

Our study also uncovered an unexpected trend in the ARG profiles of *A. baumannii*. Contrary to the pattern observed in *Klebsiella pneumoniae*, where multiple carbapenem-resistant populations exist with similar frequencies[30,31], in *A. baumannii*, 91% of the identifiable ARGs and 83.5% of CARBs are concentrated within the predominant ESL. Specifically, we observed a significant global expansion of IC2 since 2007 (Fig. 1e), with its isolation frequency increasing from about one-third to two-thirds post-2019, likely propelled by the clonal expansion of Clade 2.5.6 (Fig. 3). Additionally, based on both genomic and metagenomic data, we showed that the expansion and predominance of the ESL was restricted to clinical environment, and the samples from non-clinical sources exhibited greater genetic diversities (Supplementary Fig. 1a).

Using SNP-based barcodes, we mapped the international transmission of the ESL, classifying it into 59 clades across nine clusters. This approach, previously utilized in clonal pathogens like *Mycobacterium tuberculosis*[32] and *Shigella sonnei*[33], is novel for pathogens with high-recombination rates.

Convergent recombination between two lineages in the dataset will produce spurious similarities between them, thereby scrambling the phylogenetic signal[34]. Meanwhile, recombination events with external, genetically less related organisms could introduce multiple polymorphisms into the population, confounding accurate measures of branch lengths[35]. We estimated the frequencies of externally imported regions in the core genome of the ESL using three distinct approaches of RecHMM, Gubbins, and ClonalFrameML and the frequencies of convergent recombination in each region by the frequencies of incongruent phylogenetic partitions. The analysis revealed the presence of an LRB region with the least frequencies of both external and convergent recombination, suitable for further phylogenetic analysis.

Conversely, we found that the three HRBs are closely associated with virulence factors involved in iron acquisition, stress response, and host adaptation. It is plausible that these blocks have facilitated the acquisition of new virulence alleles from environmental CCs into various clusters in IC2, thereby driving the diversification of this epidemic lineage. Recombination has been primarily considered neutral in bacterial evolution[36]. Reported recombination "hotspots" in bacterial genomes were primarily driven by flanking mobile genetic elements[37] or genes associated with extracellular polysaccharides[38]. Our findings highlighted the role of selection in shaping recombination hotspots in *A. baumannii*, which include not only capsular gene clusters but also many other genes responsible for host/environmental adaptations (Fig. 2a). Notably, HRB2 and HRB3 were both involved in genes responsible for iron uptake, which is essential for the in vivo survival of not only *A. baumannii*[39] but also many other pathogens[40]. Similar recombination hotspots have also been reported in *Campylobacter jejuni*[41], which experienced frequent recombinations in its zinc uptake system genes (*znuABC*), likely associated with adaptation to chicken caeca and other low-zinc environments.

We detected significant temporal signals in IC2 but not IC1 using date-randomization tests, a sensitive approach that allows for high variations of substitution rates[42]. Our results estimated an overall substitution rate of IC2 to $8.16 \pm 2.19 \times 10^{-8}$, substantially lower than previous estimates of $1^{-3} \times 10^{-6}$ substitutions per site year^{-1} based on 24 to 75 IC1 or IC2 isolates[7,43]. We managed to reproduce the same clock rates using their original sets of strains (Supplementary Fig. 9a). However, the estimates of clock rates became lower, with greater uncertainty, when additional isolates from the same lineages were included (Supplementary Fig. 9d). Furthermore, it has been well-acknowledged that the substitution rates reduced when genomes were more divergent[44]. Taken together, our incorporation of all known genetic diversity in the IC2 lineage likely has enabled a more accurate estimate of its emergence.

We also compared three TreeTime results with those from BactDating and BEAST2 runs. All eleven runs generated comparable estimates of the tMRCA for both IC2 and the clade 2.5.6, demonstrating the robustness of the analysis. Notably, we incorporated recombinant events in BactDating analysis, which did not significantly change the date estimates. However, the three pipelines differed in their efficiencies. The TreeTime handles ~8000 IC2 genomes in 30 min. In contrast, it took 1–3 months and 6–10 months for BactDating and BEAST2 on the same dataset, respectively. Thus, we decided to use the TreeTime result, which offers a promising solution for other, large-scale population genetic analyses.

The IC2 lineage experienced a continuous population expansion since its emergence after World War II, coinciding with the beginning time of the widespread use of antimicrobials. We detailed the gradual evolution of IC2, highlighting the ARG acquisitions and recombination events that led to the emergence of Clade 2.5.6 two decades ago. Particularly, we compared the *A. baumannii* genomes from China over the past 24 years and observed a continuous increase in the detection rate of IC2, with only one non-IC2 genome among the clinical isolates

in 2023 (Fig. 4a and Supplementary Fig. 6a). This clade has integrated new virulence gene alleles from other populations and acquired additional ARGs, superimposing on IC2's already substantial resistance profile. Consequently, Clade 2.5.6 exhibits near-total resistance to several antibiotic classes, with high susceptibility retained only to tigecycline and polymyxin. Although instances of resistance to these latter antibiotics were infrequently observed in our dataset, *A. baumannii* strains with such resistance have been documented[45], underscoring the urgent need for controlling the spread of these threatening clades and for developing novel antimicrobial agents.

We developed an integrated bioinformatics pipeline, Capybara, which could be applied to both genomic and metagenomic reads. Our application of this tool to both clinical and community metagenomes revealed a predominance of ESL, particularly Clade 2.5.6, in clinical settings other than communities. Substantially lower frequencies of ICs in non-clinical samples have also been found in genomic data and many other studies[46,47]. Similar findings have also been reported in *K. pneumoniae*[30] and *S. aureus*[48], questioning the efficiency of the "One Health" strategy due to the genetic separation of epidemic lineages causing most infectious diseases and the remaining environmental/animal populations.

We found a group of highly recombing, non-epidemic populations, which were designated as hrCCs. The hrCCs were detected across multiple continents, exhibiting a more extensive geographic distribution and higher continental SDI scores compared to other CCs. This suggests that hrCCs may have adapted to diverse environments by acquiring novel gene alleles, likely linked to the species' well-documented natural competence[28]. Furthermore, they exhibited lower carriages of ARGs (Fig. 5b), thus possibly persistent in non-clinical environments. Similar populations of high recombination have been observed in other bacterial species, including *Salmonella enterica* lineage three[49] and non-epidemic strains of *Vibrio parahaemolyticus*[50]. In contrast, the two ICs within the ESL displayed lower recombination rates, except for a notable pathway linking IC2 to hrCCs through the recombination network (Fig. 5c).

This recombination pathway transferred genes in the non-epidemic populations into the ESL, contributing to the evolution of epidemic lineages. The ESL acquired ARGs and virulence genes through distinct evolutionary processes. ARGs were predominantly accumulated through the integration of transposons and genetic elements, as well as point mutations in core genes. In contrast, virulence genes were largely updated through homologous recombination, particularly in regions characterized by high-recombination rates. It is often that, in ESKAPE pathogens, the strains with the greatest amount of ARGs are not necessarily the most successful in terms of their prevalence[51]. We attribute this dilemma to the high evolutionary costs of expressing ARGs, potentially leading to a competitive disadvantage in environments devoid of selective antimicrobial pressures[52]. Conversely, genes involved in iron uptake, stress response, and immune evasion are crucial for the survival of pathogens, particularly in harsh environments, and thus are essential in the emergence of epidemic lineages. This is exemplified by the rise of IC2 and the subsequent global spread of Clade 2.5.6, which has integrated novel virulence alleles and accumulated additional ARGs (Fig. 5d).

In this study, we also highlighted the significant role of plasmids and transposons in the dissemination of antimicrobial resistance. Notably, by analyzing 162 complete *A. baumannii* genomes, we discovered that the spread of $bla_{OXA\text{-}23}$ genes in IC2 and IC1 was primarily mediated by transposons. In our dataset, 92% (103/111) of the $bla_{OXA\text{-}23}$ genes were associated with transposon (Tn/IS) elements, particularly Tn6166 and Tn2007, whereas only 23% (26/111) were located in plasmids. Furthermore, 69% (18/26) of the plasmid-carrying $bla_{OXA\text{-}23}$ genes were also associated with transposons, underscoring the important role of transposons in the dissemination of $bla_{OXA\text{-}23}$ genes.

Altogether, we propose a gene-centered view of the "One Health" paradigm, where genes, rather than the bacteria themselves, are transferred between different environments through homologous recombination or horizontal genetic transfer. This model might apply to the evolution of other MDR pathogens, emphasizing the critical role of environmental populations in preserving genetic diversity for both ARGs and virulence genes. Such genetic diversity is integral to the selective process that drives the evolution of the ESL and the emergence of successful epidemic clades. Furthermore, the ARGs and virulence genes might also spread into the non-clinical environments through a similar pathway, potentially threatening public health.

A limitation of this work is the absence of phenotypic tests for AMR, which limited our ability to detect novel, undocumented drug resistance mechanisms. Furthermore, the involved strains were predominantly collected in clinical environments over the past 20 years and may not accurately reflect earlier epidemiological patterns. Particularly, the oversampling of recent isolates also reduced the significance of the temporal signal. An equal sampling of the strains along with time would substantially improve the *R* value in the root-to-tip regression.

In conclusion, this study provides a comprehensive overview of the population structure of *A. baumannii*, linking ~70% of current clinical isolates to the extensive ESL, which encompasses the prominent ICs 1 and 2. We have traced the evolutionary trajectory of the ESL and documented the rise and global spread of a novel clade, 2.5.6, within IC2 over the past 16 years. The expansion of IC2 and the emergence of Clade 2.5.6 are linked to the acquisition of ARGs and the recombination-driven evolution of virulence factors, offering insights into the development of epidemic lineages in *A. baumannii*. These findings will support future genomic surveillance initiatives and enhance our understanding of bacterial pathogen diversification and global dissemination.

Methods

Genomes and representative sequences

A total of 6208 assemblies of *A. baumannii* were retrieved from the RefSeq database in Aug. 2022. Additionally, we also downloaded 10,362 sets of short reads from the NCBI SRA database and assembled them using the "assemble" module in EToKi[53]. All genomes were compared with the reference genome [K09-14: GCF_008632635.1; from soil, 2019, Malaysia] using fastANI v1.33[54], and only 15,643 genomes that exhibited ≥95% ANIs to the references were kept in downstream analysis. All genomes were annotated using Prokka v1.14.6[55]. The 7-gene MLST STs of each genome were in silico predicted after comparing the genomic sequences using BLASTn v2.11.0[56] onto the allelic sequences for Pasteur's and Oxford's schemes, both hosted on PubMLST[57]. MLST profiles were then clustered into different CCs with eBurst v1.0.5[58]. Antibiotic resistance genes in each genome were predicted using AMRFinderPlus v3.11.26[59], and the virulence genes were predicted based on BLASTp searches against the VFDB 2022 release[60], with ≥ 90% identity and ≥ 60% coverage, respectively. Furthermore, genes associated with biocide resistance were also predicted by BLASTp searching of genes described in Li et al.[26]. Finally, the types of OCL and *cps* were predicted using Kaptive v2.0.3[61], and the remaining virulence genes were separated into clusters of <99.5% identities using USEARCH[62].

To select a set of representative genomes that retained most of the genetic diversity while removing genetic redundancy, we compared pairwise genetic distances of all genomes using Kssd v1.1[63] and separated genomes into single-linkage clusters of ≥99.5% identities. Only one sequence with the greatest N50 value was chosen for each cluster and was further quality-checked by the presence of >37/40 single-copy core genes using fetchMGs v1.2[64], resulting in the final subset of 5824 representative genomes for *A. baumannii*.

Genomic sequencing of East China isolates

The DNA of 100 isolates from Wenzhou, Zhejiang, and Suzhou, Jiangsu were purified and extracted by Qiagen EZ1 DNA Tissue Kit (Qiagen Sciences, Germantown, MD, USA) according to the manufacturer's instructions. Paired-end libraries with insert sizes of ~300 bp were prepared following Illumina standard genomic DNA library preparation procedure (VAHTS Universal DNA Library Prep kit for Illumina V3) and sequenced on an Illumina NovaSeq 6000 using the S4 reagent kits (v1.5) according to the manufacturer's instructions. The sequencing reads of each isolate were quality-trimmed and assembled into contigs using EToKi. Assembled genomes were submitted to NCBI under BioProject accession number PRJNA1112767. Community metagenomic samples from Suzhou City were publicly accessible with BioProject accession PRJNA1225594. Clinical metagenomic samples from hospital B in Suzhou City were publicly accessible with BioProject accession PRJNA1219970. Metagenomic samples from Shanghai City were publicly accessible with NCBI BioProject accession PRJNA1028672. Metagenomic samples from Shenzhen City were publicly accessible with NCBI BioProject accession PRJNA572371. A detailed list of the sample accession numbers for all samples is available in Supplementary Data 4.

Phylogenetic analysis

Based on the representative genomes, a total of 36,992 pan genes were estimated using PEPPAN v1.0.5[65]. Reference sequences for each pan gene were selected by choosing one allele for each cluster with ≥90% identity using EToKi MLSTdb. The MLSTdb module also identifies potential paralogous genes with ≥90% identities, which were removed from the scheme. We calculated the presence of pan genes in the representative genomes using the DTy[66], which identifies and extracts homologous genes in genomes based on the reference sequences. We then used a divide-and-conquer strategy for ML estimation based on sequence divergence of core genes. Based on the results, we extracted a subset of 2266 softcore genes by selecting each gene that was (1) present in ≥95% of genomes and (2) maintained intact open reading frames in ≥94% of its alleles, thresholds that were derived from our early practice for core genome multi-locus sequence typing schemes[53].

We estimated an ML tree for each of the 2266 core genes (Supplementary Data 2) in the representative genomes using IQTREE[67] and summarized all the gene trees together into a supertree using the cgMLSA pipeline[68]. Briefly, cgMLSA summarized the gene trees into a guide tree using ASTRID[69] and used the guide tree to separate genomes into disjoint subgroups. The supertree for each disjoint subgroup was estimated using ASTRAL[70]. cgMLSA implemented a tag-based approach to summarize the trees of all disjoint subgroups together into the final supertree, which had the branch lengths estimated using ERaBLE[71].

Removal of recombination blocks

We aligned all genomes in the ESL onto a reference genome [MDR-TJ: GCF_000187205.2; from the hospital, 2013, China[72]] using EToKi align, which identified a total of 181,732 SNPs in the 2.35 MB non-repetitive core genome. An ML tree was estimated based on all core SNPs using IQTREE and used to identify recombination in the ESL using all of RecHMM, Gubbins[73], and ClonalFrameML[74]. Both pipelines identified high frequencies of recombination (r/m > 10), with uneven distribution along the chromosome (Supplementary Fig. 2a–c). The RecHMM results were chosen arbitrarily for downstream analyses. We identified the HRBs as continuous regions with >50 recombination events in a sliding window of 100Kb, and the MRBs and LRB were regions between neighboring HRBs. Mutational trees were built for each of the HRBs, MRBs, and LRB (Supplementary Fig. 2d) after the removal of SNPs imported by recombination. The frequencies of homoplastic sites were evaluated by TreeTime[75]. All phylogenies were visualized using ITOL v6[76] or GrapeTree v1.5.0[77].

Phylogeny and prediction of lineages and clades of ESL

After the removal of recombinant SNPs among LRB, we eventually reserved 23,524 SNPs to build the phylogeny of ESL. We visually separated all ESL strains into 59 monophyletic clusters of clades along the phylogeny. Furthermore, to provide long-term evolutionary insights, these clades were further grouped into 4 and 5 clusters in the IC1 and IC2 lineages, respectively. Furthermore, hierBAPS[78] was employed for population assignments of ESL with parameters "max.depth = 2, n.pops = 50", which failed to generate meaningful clusters, possibly due to the high-recombination level of the ESL (Supplementary Fig. 2e). Finally, one synonymous mutation was chosen as the genetic marker for each clade, which was adopted in Capybara for rapid clade assignments (Supplementary Table 3).

Phylogeographical inference for IC2

We evaluated the accumulation of variations in both the ESL and the two of its lineages (IC1 and IC2) using both the TempEst v1.5.3[79] and tested the reliability of the temporal signal by date-randomization test[42] which analyzed multiple date-randomized replicate datasets after randomly reassigning the isolation dates of the genomes. IC2 is the only group that had significant signals (Supplementary Fig. 4b, c). Therefore, we calibrated the temporal tree for IC2 based on the isolation years of its genomes using TreeTime, which finished in 30 min, and further predicted its country-wise transmissions using the "mugration" module in TreeTime (Supplementary Fig. 4a). Furthermore, we estimated the skyline plot of its EPS using Skygrowth version 0.3.1[80]. We also cross-validated the substitution rate with previous research using small datasets (Supplementary Fig. 8). Additionally, we ran BactDating and BEAST2 on the IC2 lineage. Specifically, BactDating incorporated recombinant events in date estimates by loading clonalframeML results using the loadCFML() function. We ran BactDating with alternative models of "arc", "carc", "negbin", "mixedgamma", and "relaxedgamma", and ran BEAST2 with a "GTR" substitution model and alternative clock models of "Relaxed LogNormal", "Relaxed Exponential", "Relax Uniform", "Strict Clock", and "Random Local Clock". For each run, the chain length was set to 1e12 or when the ESS exceeded 200 (Supplementary Fig. 5). The BactDating runs finished in 1–3 months, whereas the BEAST2 runs lasted for 6–10 months.

Estimation of topological conflicts between the gene trees and the supertree

We employed quartets, namely subtrees with only four tips, to quantify the level of topological conflicts between the supertree and each gene tree. As shown in Supplementary Fig. 7a, only three topologies are expected for any quartet, which are: ((A, B), (C, D)), ((A, C), (B, D)), and ((A, D), (B, C)). The supertree and the gene tree share the same quartet if the topologies of the quartets from both trees are the same, or different otherwise. Then, the conflicting percentage of a gene tree, t, is calculated as $p_t = \sum n_c / \sum n_t$, where the n_c is the total number of conflicting quartets and n_t is the total number of all shared quartets between the gene tree and the supertree. However, it is almost impossible to traverse through all quartets because a total of 4.8×10^{13} quartets are expected for a supertree with 5,824 genomes. Therefore, we randomly 100 million quartets in each comparison, and the expected conflicting percentage, $\hat{p}_t$, is calculated as $\hat{p}_t = \sum n'_c / \sum n'_t$, where n'_c and n'_t are the number of conflicting and total quartets in the 100 million random selections.

Evaluation of recombination level for each CC

For each CC, g, the 100 million randomly generated quartets for each gene tree, t, could be separated into two categories: $n'_{t,g}$, the number of quartets that involve at least one genome from g, and the remaining. The recombination level of g in tree t is calculated as $\hat{p}_{t,g} = \sum n'_{c,g} / \sum n'_{t,g}$. For example, the recombination levels of ESL-containing quartets and the others without ESL were calculated for

gene trees and visualized in Fig. 2b. Furthermore, we calculated the overall recombination frequency of each CC, g, as conflicting quartets per gene per genome (CQPP), as:

$$CQPP_g = \frac{\sum_{t \in T} \hat{p}_{t,g}}{N}$$

Where N is the number of genomes in g and $T = \{t\}$.

Identification of the convergent recombination network

We further inferred recent inter-population recombination that results in the convergence of genes between distant CCs. To this end, for each conflicting quartet between the supertree and the gene tree, we have two topologies of q_t for the quartet from the gene tree and q_s for the quartet from the supertree (Supplementary Fig. 7a). We then identified the pair of tips (i.e., 1 and 2 in Supplementary Fig. 7a) with the shortest genetic distance in q_t, which always fell into one clade in q_t but not in q_s (Supplementary Fig. 7a). Such a topological difference is likely a result of recombination, either due to divergent recombination that took the original neighbors of A or B away (3 and 4 in Supplementary Fig. 7a) or by convergent recombination that brought A and B together (5 and 6 in Supplementary Fig. 7a). The convergent recombination is more likely to have occurred because A and B are selected to represent the smallest genetic differences. We summarized all such convergent events together for each pair of genomes, and further calculated the average convergent frequencies between all pairs of CCs, which were visualized as a network in Fig. 5c.

Description of the Capybara pipeline

Capybara is a generalized genotyping pipeline based on an SNP barcode scheme, namely a set of 80 pre-calculated SNPs (Supplementary Table 3) that separate the ESL into clusters and clades. Capybara accepts the files containing sequencing reads as parameters, and (1) Align all reads onto a reference genome [MDR-TJ: GCF_000187205.2] by minimap2[81] and call base variants using SAMtools[82] and bcftools[83]. (2) It then compares the called nucleotides with the set of pre-calculated SNPs and assigns the read sets to the corresponding clusters and clades based on the matching SNPs.

Identification of *A. baumannii* in metagenomic data

MetaPhlAn v4.0[84] was employed to detect the presence of *A. baumannii* in metagenomic reads data, following parameters of "--index mpa_vOct22_CHOCOPhlAnSGB_202212". For sufficient read counts for Capybara, we selected samples with a relative abundance of *A. baumannii* of ≥0.1% for subsequent analysis.

Calculation of the Simpson diversity index of STs

SDI of sequence types (STs) for each bacteria, was calculated as SDI $=1-\sum_{i=1}^{n} p_i^2$, where n was the total number of STs and p_i was the total number of detection rate of the ith ST. The closer the SDI value is to 1, the more evenly distributed the species is in different STs, with no obvious preference.

Reporting summary

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

Data availability

Source data are provided with this paper. The SNP barcode of ESL is available as part of the Capybara pipeline [https://github.com/Zhou-lab-SUDA/Capybara]. 100 clinical samples from hospital A in Suzhou City and the hospital in Wenzhou were deposited in NCBI GenBank under BioProject accession: PRJNA1112767. Community metagenomic samples from Suzhou City were publicly accessible with BioProject accession PRJNA1225594. Clinical metagenomic samples from hospital B in Suzhou City were publicly accessible with BioProject accession PRJNA1219970. Metagenomic samples from Shanghai City were publicly accessible with NCBI BioProject accession PRJNA1028672. Metagenomic samples from Shenzhen City were publicly accessible with NCBI BioProject accession PRJNA572371. A detailed list of the sample accession numbers for all samples are available in Supplementary Data 4. Source data of each Figure and Table can be found in Supplementary Data, Supplementary Table and Source data. Source data are provided with this paper.

Code availability

Code for Capybara is available on GitHub https://github.com/Zhou-lab-SUDA/Capybara which also deposits all additional software packages and tools that are relevant.

References

1. Li, X. et al. China's plan to combat antimicrobial resistance. *Science* **383**, 6690 (2024).
2. Hu, F. et al. Resistance reported from China antimicrobial surveillance network (CHINET) in 2018. *Eur. J. Clin. Microbiol. Infect. Dis.* **38**, 2275–2281 (2019).
3. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
4. Diancourt, L., Passet, V., Nemec, A., Dijkshoorn, L. & Brisse, S. The population structure of Acinetobacter baumannii: expanding multiresistant clones from an ancestral susceptible genetic pool. *PLoS ONE* **5**, e10034 (2010).
5. Higgins, P. G., Dammhayn, C., Hackel, M. & Seifert, H. Global spread of carbapenem-resistant Acinetobacter baumannii. *J. Antimicrob. Chemother.* **65**, 233–238 (2010).
6. Higgins, P. G., Prior, K., Harmsen, D. & Seifert, H. Development and evaluation of a core genome multilocus typing scheme for whole-genome sequence-based typing of Acinetobacter baumannii. *PLoS ONE* **12**, e0179228 (2017).
7. Holt, K. et al. Five decades of genome evolution in the globally distributed, extensively antibiotic-resistant Acinetobacter baumannii global clone 1. *Microb. Genom.* **2**, e000052 - (2016).
8. Wang, X., Qiao, F., Yu, R., Gao, Y. & Zong, Z. Clonal diversity of Acinetobacter baumannii clinical isolates revealed by a snapshot study. *BMC Microbiol.* **13**, 234 (2013).
9. Jones, A. et al. Importation of multidrug-resistant Acinetobacter spp infections with casualties from Iraq. *Lancet Infect. Dis.* **6**, 317–318 (2006).
10. Müller, C. et al. A global view on carbapenem-resistant Acinetobacter baumannii. *mBio* **14**, e0226023 (2023).
11. Castillo-Ramírez, S. Genomic epidemiology of *Acinetobacter baumannii* goes global. *mBio* **14**, e0252023 (2023).
12. Hernández-González, I. L., Mateo-Estrada, V. & Castillo-Ramirez, S. The promiscuous and highly mobile resistome of *Acinetobacter baumannii*. *Microb. Genom.* **8**, 000762 (2022).
13. Banerjee, R. et al. Exploring the resistome and virulome in major sequence types of *Acinetobacter baumannii* genomes: correlations with genome divergence and sequence types. *Infect. Genet. Evol.* **119**, 105579 (2024).
14. Castillo-Ramírez, S. & Graña-Miraglia, L. Inaccurate multilocus sequence typing of *Acinetobacter baumannii*. *Emerg. Infect. Dis.* **25**, 186–187 (2019).
15. Álvarez, V. E. et al. Crucial role of the accessory genome in the evolutionary trajectory of *Acinetobacter baumannii* Global Clone 1. *Front. Microbiol.* **11**, 342 (2020).
16. Denissen, J. et al. Prevalence of ESKAPE pathogens in the environment: antibiotic resistance status, community-acquired infection

and risk to human health. *Int. J. Hyg. Environ. Health* **244**, 114006 (2022).
17. Hamidian, M. & Nigro, S. J. Emergence, molecular mechanisms and global spread of carbapenem-resistant Acinetobacter baumannii. *Microb. Genom.* **5**, e000306 (2019).
18. Graña-Miraglia, L. et al. Origin of OXA-23 variant OXA-239 from a recently emerged lineage of Acinetobacter baumannii International Clone V. *mSphere* **5**, e00801–e00819 (2020).
19. Zarrilli, R., Pournaras, S., Giannouli, M. & Tsakris, A. Global evolution of multidrug-resistant Acinetobacter baumannii clonal lineages. *Int. J. Antimicrob. Agents* **41**, 11–19 (2013).
20. Jean, S. S., Harnod, D. & Hsueh, P. R. Global threat of Carbapenem-resistant Gram-negative bacteria. *Front. Cell. Infect. Microbiol.* **12**, 823684 (2022).
21. Karah, N. et al. Emergence of Acinetobacter baumannii International Clone 10 predominantly found in the Middle East. *bioRxiv* 2023-10 (2023).
22. Hansen, F. et al. Characterisation of carbapenemase-producing Acinetobacter baumannii isolates from Danish patients 2014-2021: detection of a new international clone - IC11. *Int. J. Antimicrob. Agents* **62**, 106866 (2023).
23. Coyne, S., Courvalin, P. & Périchon, B. Efflux-mediated antibiotic resistance in Acinetobacter spp. *Antimicrob. Agents Chemother.* **55**, 947–953 (2011).
24. Krizova, L., Poirel, L., Nordmann, P. & Nemec, A. TEM-1 β-lactamase as a source of resistance to sulbactam in clinical strains of Acinetobacter baumannii. *J. Antimicrob. Chemother.* **68**, 2786–2791 (2013).
25. Hawkey, J. et al. Evolution of carbapenem resistance in Acinetobacter baumannii during a prolonged infection. *Microb. Genom.* **4**, e000165 (2018).
26. Li, L. et al. Systematic analyses identify modes of action of ten clinically relevant biocides and antibiotic antagonism in Acinetobacter baumannii. *Nat. Microbiol.* **8**, 1995–2005 (2023).
27. Antunes, L. C., Visca, P. & Towner, K. J. Acinetobacter baumannii: evolution of a global pathogen. *Pathog. Dis.* **71**, 292–301 (2014).
28. Gallagher, L. A. *Acinetobacter Baumannii: Methods And Protocols* Vol. 1946 (Springer New York, 2019).
29. Khuntayaporn, P. et al. Predominance of international clone 2 multidrug-resistant Acinetobacter baumannii clinical isolates in Thailand: a nationwide study. *Ann. Clin. Microbiol. Antimicrob.* **20**, 19 (2021).
30. Lam, M. M. C. et al. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. *Nat. Commun.* **12**, 4188 (2021).
31. Li, H. et al. KleTy: integrated typing scheme for core genome and plasmids reveals repeated emergence of multi-drug resistant epidemic lineages in Klebsiella worldwide. *Genome Med.* **16**, 130 (2024).
32. Meehan, C. J. et al. Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. *Nat. Rev. Microbiol.* **17**, 533–545 (2019).
33. Hawkey, J. et al. Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, Shigella sonnei. *Nat. Commun.* **12**, 2684 (2021).
34. Didelot, X., Achtman, M., Parkhill, J., Thomson, N. R. & Falush, D. A bimodal pattern of relatedness between the Salmonella Paratyphi A and Typhi genomes: convergence or divergence by homologous recombination? *Genome Res.* **17**, 61–68 (2007).
35. Croucher, N. J., Harris, S. R., Grad, Y. H. & Hanage, W. P. Bacterial genomes in epidemiology—present and future. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120202 (2013).
36. Hedge, J. & Wilson, D. J. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *mBio* **5**, e02158 (2014).
37. Everitt, R. G. et al. Mobile elements drive recombination hotspots in the core genome of Staphylococcus aureus. *Nat. Commun.* **5**, 3956 (2014).
38. Mostowy, R. J. et al. Pneumococcal capsule synthesis locus cps as evolutionary hotspot with potential to generate novel serotypes by recombination. *Mol. Biol. Evol.* **34**, 2537–2554 (2017).
39. Artuso, I. et al. Genomics of Acinetobacter baumannii iron uptake. *Microb. Genom.* **9**, mgen001080 (2023).
40. Murdoch, C. C. & Skaar, E. P. Nutritional immunity: the battle for nutrient metals at the host-pathogen interface. *Nat. Rev. Microbiol.* **20**, 657–670 (2022).
41. Yahara, K. et al. Efficient inference of recombination hot regions in bacterial genomes. *Mol. Biol. Evol.* **31**, 1593–1605 (2014).
42. Duchêne, S., Duchêne, D., Holmes, E. C. & Ho, S. Y. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Mol. Biol. Evol.* **32**, 1895–1906 (2015).
43. Schultz, M. B. et al. Repeated local emergence of carbapenem-resistant Acinetobacter baumannii in a single hospital ward. *Microb. Genom.* **2**, e000050 (2016).
44. Yahara, K. et al. The landscape of realized homologous recombination in pathogenic bacteria. *Mol. Biol. Evol.* **33**, 456–471 (2016).
45. Rolain, J. M., Roch, A., Castanier, M., Papazian, L. & Raoult, D. Acinetobacter baumannii resistant to colistin with impaired virulence: a case report from France. *J. Infect. Dis.* **204**, 1146–1147 (2011).
46. Mateo-Estrada, V. et al. Acinetobacter baumannii from grass: novel but non-resistant clones. *Microb. Genom.* **9**, mgen001054 (2023).
47. Hernández-González, I. L. & Castillo-Ramírez, S. Antibiotic-resistant Acinetobacter baumannii is a One Health problem. *Lancet Microbe* **1**, e279 (2020).
48. Vysakh, P. R. & Jeya, M. A comparative analysis of community acquired and hospital acquired methicillin resistant Staphylococcus aureus. *J. Clin. Diagn. Res.* **7**, 1339–1342 (2013).
49. Wong, V. K. et al. An extended genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid. *Nat. Commun.* **7**, 12827 (2016).
50. Cui, Y., Yang, C., Qiu, H., Wang, H., Yang, R. & Falush, D. The landscape of coadaptation in Vibrio parahaemolyticus. *eLife* **9**, e54136 (2020).
51. Beceiro, A., Tomás, M. & Bou, G. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin. Microbiol. Rev.* **26**, 185–230 (2013).
52. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
53. 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).
54. Jain, C. et al. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
55. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
56. Altschul, S. F. et al. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
57. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **3**, 124 (2018).
58. Francisco, A. P. et al. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinform.* **10**, 152 (2009).
59. 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).

60. Liu, B. et al. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* **47**, D687–D692 (2019).
61. Lam, M. M. C. et al. Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the Klebsiella pneumoniae species complex. *Microb. Genom.* **8**, 000800 (2022).
62. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
63. Yi, H., Lin, Y., Lin, C. & Jin, W. Kssd: sequence dimensionality reduction by k-mer substring space sampling enables real-time large-scale datasets analysis. *Genome Biol.* **22**, 84 (2021).
64. Milanese, A. et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1014 (2019).
65. Zhou, Z., Charlesworth, J. & Achtman, M. Accurate reconstruction of bacterial pan- and core genomes with PEPPAN. *Genome Res.* **30**, 1667–1679 (2020).
66. Zhong, L. et al. Distributed genotyping and clustering of Neisseria strains reveal continual emergence of epidemic meningococcus over a century. *Nat. Commun.* **14**, 7706 (2023).
67. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
68. Achtman, M. et al. EnteroBase: hierarchical clustering of 100 000s of bacterial genomes into species/subspecies and populations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20210240 (2022).
69. Vachaspati, P. & Warnow, T. ASTRID: Accurate Species TRees from Internode Distances. *BMC Genom.* **16**, S3 (2015).
70. Sayyari, E. & Mirarab, S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* **33**, 1654–1668 (2016).
71. Binet, M. et al. Fast and accurate branch lengths estimation for phylogenomic trees. *BMC Bioinform.* **17**, 23 (2016).
72. Gao, F. et al. Genome sequence of Acinetobacter baumannii MDR-TJ. *J. Bacteriol.* **193**, 2365–2366 (2011).
73. 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).
74. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).
75. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).
76. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
77. Zhou, Z. et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* **28**, 1395–1404 (2018).
78. Tonkin-Hill, G. et al. RhierBAPS: an R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res.* **3**, 93 (2018).
79. Rambaut, A. et al. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
80. Volz, E. M. & Didelot, X. Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. *Syst. Biol.* **67**, 719–728 (2018).
81. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
82. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
83. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
84. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).

Acknowledgements

We thank Prof. Mark Achtman for his critical reading and valuable comments. The project was supported by the National Natural Science Foundation of China (No. 32170003, 32370099, 82202465), the Natural Science Foundation of Jiangsu Province (BK20211311), the Provincial-level Talent Program for National Center of Technology Innovation for Biopharmaceuticals (NCTIB2024JS0101), Jiangsu Specially-appointed Professor Project, the Suzhou Top-Notch Talent Groups (ZXD2022003) and the Suzhou Science and Technology Innovations Project in Health Care (SKY2021013).

Author contributions

Z.Z., J.D., and S.L. designed the study. S.L., P.H., S.W., M.Z., H.L., Y.W., and G.J. performed experiments, analyses, and wrote the initial manuscript. S.L., M.W., S.X., and S.W. designed and wrote the pipeline. Y.R., L.Z., and X.L. discussed the results. J.Z. and H.L. performed and revised the results and method content. Z.Z., J.D., and Y.L. directed the project and edited the manuscript.

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/s41467-025-58106-9.

Correspondence and requests for materials should be addressed to Yongliang Lou, Heng Li, Jimei Du or Zhemin Zhou.

Peer review information *Nature Communications* thanks Santiago Castillo-Ramirez, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025