High-throughput experimental evolution reveals the mutation-driven landscape of high-level clinical resistance acquired under low-level environmental antibiotic stresses

Hanqing Wang[1], Hui Lu[2], Chao Jiang[3], Lizhong Zhu[1]*, Huijie Lu[1]*

1. College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China
2. ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 311215, China
3. Life Sciences Institute, Zhejiang University, Hangzhou, 310058, China

*To whom correspondence should be addressed:
Huijie Lu (luhuijie@zju.edu.cn)
Lizhong Zhu (zlz@zju.edu.cn)

Abstract

Potential pathogens exposed to low-level environmental antibiotics could develop high-level clinically relevant antibiotic resistance detrimental to the health of the general population. However, the underlying evolutionary landscapes remain poorly understood. We conducted a high-throughput experimental evolution study by exposing an environmentally isolated pathogenic *Escherichia coli* strain to 96 typical antibiotics at 10 μg l^{-1} for 20 days. Antibiotic resistance phenotypic (IC_{90} against 8 clinically used antibiotics) and genetic changes of the evolved populations were systematically investigated, revealing a universal increase in antibiotic resistance (up to 349-fold), and mutations in 2,432 genes. Transposon sequencing was further employed to verify genes potentially associated with resistance. A core set of mutant genes conferring high-level resistance was analyzed to elucidate their resistance mechanisms by analyzing the functions of interacted genes within the gene co-fitness network and performing gene knockout validations. We developed machine-learning models to predict antibiotic resistance phenotypes from antibiotic structures and genomic mutations, enabling the resistance predictions for another 569 antibiotics. Importantly, 14.6% of the 481 key mutations were observed in clinical and environmental *E. coli* isolates retrieved from the NCBI database, and several were over-represented in clinical isolates. Deciphering the evolutionary landscapes underlying resistance exposed to low-level environmental antibiotics is crucial for evaluating the emergence and risks of environment-originated clinical antibiotic resistance.

Introduction

The emergence of antimicrobial resistance (AMR) in the environment is a growing concern in the One Health paradigm[1–5]. Antibiotics are increasingly released into the environment via wastewater discharge and disposal of solid waste, where 40-90% are excreted as parent compounds[6]. Environmental bacteria exposed to these antibiotics at low-level, sub-minimal inhibitory concentrations (sub-MIC) can evolve antibiotic resistance by accumulating diverse low-frequency mutants, which are different from those evolved at clinical concentrations[7–11]. The environmental research community essentially needs to elucidate the contribution of mutation-driven resistance evolution in the environment by crafting analytical measures to pinpoint the responsible selective forces (i.e., antibiotics) driving AMR development[12]. Therefore, exploring the evolutionary landscape under low-level antibiotic exposures is increasingly important at the interface between environmental and clinical microbiology pertaining to antibiotic resistance[4,13].

Traditional low-throughput experimental evolution has provided an incomplete view of the AMR evolutionary landscape, as the number of stress conditions, mutants, and resistance-conferring genes examined were limited[14,15]. High-throughput experimental evolution enables the rapid generation and systematic investigation of population-level antibiotic resistance under a variety of stress conditions, allowing better characterization of the structure- or dose-response relationships among stressors and induced resistance[16,17]. In addition, antibiotic resistance genes (ARGs) were traditionally identified by sequencing and verified through phenotypic assays/knockout experiments, bearing the risks of overlooking critical insights from genetic interactions. Transposon sequencing (Tn-Seq) measures genome-wide gene fitness across large-scale transposon mutagenesis libraries. By combining Tn-Seq with deep resequencing, researchers can map the accumulated mutations onto complex and coordinated interaction networks. The

prevalence and clinical significance of identified mutations in broad environments can be examined by comparative genomics[18]. The above novel approaches could help address critical knowledge gaps in the field of environmental AMR, for example: (1) What is the association between the antibiotic resistance of evolved populations and the chemical structures of antibiotics under low-level exposures? (2) What cellular functions are affected by mutations potentially via new antibiotic resistance mechanisms? (3) Are mutations that arise under low-level antibiotic exposures observed in clinically isolated resistant strains?

This study utilized a soil-isolated opportunistic pathogen *E. coli* strain to perform high-throughput experimental evolution under low-level exposures to 96 antibiotic stressors (**Supplementary Fig. 1**). The antibiotic resistance profiles of evolved populations and their core mutations were characterized. To explore the resistance mechanisms conferred by these core mutant genes, a genome-wide co-fitness interaction network was constructed based on Tn-seq. Chemical fingerprint-based modeling was built to predict antibiotic resistance profiles induced by another 569 antibiotics sharing similar structural features with the 96 stressors. The occurrence of core mutations in broader environmental and clinical isolates was examined by comparative genomics. These findings underscore the importance of AMR developed under low-level antibiotic exposures in environmental settings and its potential spread to clinical contexts.

Results

Antibiotic resistance profiles of evolved *E. coli* QSHQ1 populations

E. coli strain QSHQ1 isolated from soil evolved under exposure to 96 antibiotic stresses at an environmentally relevant concentration of 10 μg l^{-1} for 20 days, which corresponded to approximately 200 generations (**Supplementary Data 1**). These antibiotics can be classified into ten classes and mainly act through four mechanisms, including inhibition of DNA/RNA (DSI), protein (PSI), cell wall (CWSI), and membrane (MSI) synthesis (**Fig. 1a**). All populations survived and displayed varying degrees of growth inhibition compared to control populations without antibiotics (**Supplementary Fig. 2**). Evolved populations on day 20 were tested for antibiotic resistance (indicated by IC_{90}) to eight clinically relevant antibiotics, including azithromycin (AZI), ciprofloxacin (CIP), tetracycline (TET), sulfamethoxazole with trimethoprim (SMZ+TMP), tobramycin (TOB), ceftazidime (CAZ), meropenem (MER), and polymyxin B (PMB). These IC_{90} values were compared with CLSI's breakpoint values for susceptible and resistant bacteria (**Fig. 1b** and **Supplementary Data 2**).

The wild-type (WT) strain was initially susceptible or intermediate to the eight clinically relevant antibiotics. All 96 evolved populations conferred resistance (increased IC_{90}) to at least one of the eight antibiotics (Multiple unpaired *t*-tests, FDR < 0.1, **Fig. 1a**), where resistance to AZI (78.9%) and TOB (18.8%) were the most prevalent. A similar evolution of AZI-resistant strains was recently reported for *Comamonas testosteroni*[10]. Only two evolved populations stressed by marbofloxacin and ceftriaxone displayed significantly decreased IC_{90} values to at least one antibiotic. Specifically, a ciprofloxacin-stressed population displayed the most significant increase in CIP resistance (IC_{90} increased by 345-fold), and a cefazolin-stressed population displayed a 349-fold increase in MER resistance (**Fig. 1b**). Previous studies[19,20] reported 256-fold and >300-fold increase in the MIC

118 for ciprofloxacin of *E. coli* and *S. aureus* cells after < 20 passages, respectively. In
119 fact, the WT strain was susceptible to CIP and MER. Exposure to quinolones was
120 the primary driver of CIP resistance evolution (average IC_{90} fold-change: 24.6 ±
121 14.9), while exposure to aminoglycosides highly contributed to the evolution of
122 TOB resistance (average IC_{90} fold-change: 2.19 ± 0.26) (**Fig. 1c**, **Supplementary**
123 **Figs. 3** and **4**).

124 Populations evolved under each class of antibiotics displayed variations in
125 resistance to the eight antibiotics tested, where 5/8 displayed either significantly
126 positive or negative correlations induced by a specific class of antibiotics (**Fig. 1d**
127 and **Supplementary Fig. 5**). Overall, positive correlations (Pearson's $r > 0.6$, $P <$
128 0.05) were much more abundant than negative ones, where the latter was only
129 observed for macrolides-stressed populations and between their AZI and MER
130 resistance ($r = -0.85$, $P < 0.05$). The 96 antibiotic stressors were further clustered
131 based on the antibiotic resistance profiles they induced (**Fig. 1e** and
132 **Supplementary Fig. 6**). Most aminoglycosides and macrolides (both acting via
133 PSI) were clustered (VI), as well as sulfonamides and tetracyclines (acting via DSI
134 and PSI, respectively, cluster II). Quinolones formed their distinct cluster (cluster
135 III), separated from other antibiotic classes.
136

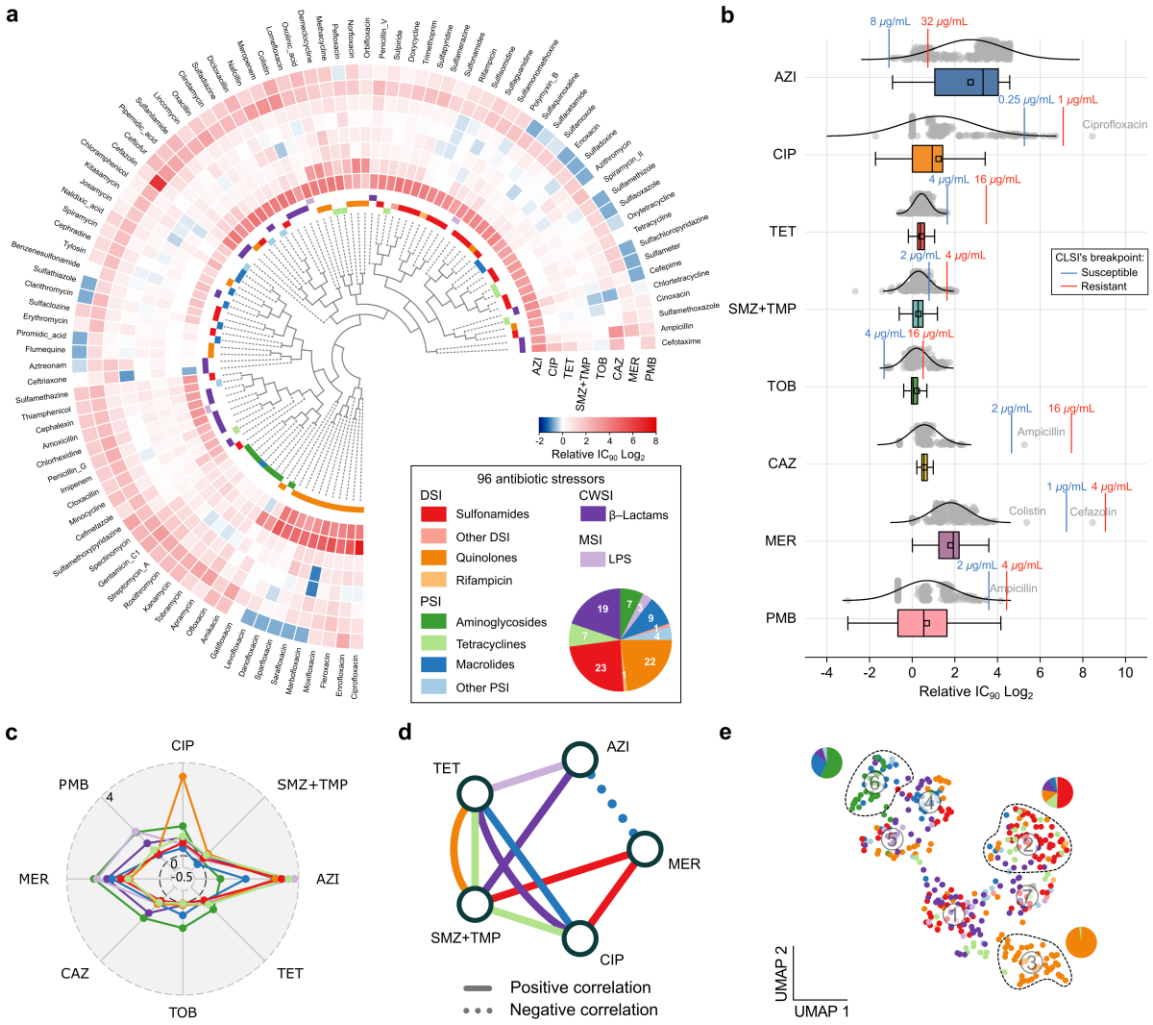

Fig. 1: Antibiotic resistance profiles of evolved populations. **a,** Antibiotic resistance profiles of the evolved populations. Antibiotic resistance changes were indicated by $\log_2$-transformed fold changes in IC_{90} values (relative to the control populations without antibiotic stress). The tree was constructed based on Euclidean distances. Antibiotics stressors: 96; Stressor classes: 10 (different colors of the inner circle); Stressor groups: 4 (DSI, CWSI, PSI, and MSI); Antibiotic resistance assay: 8 (heatmap circles); Biological replicates: 4. **b,** Distribution of antibiotic resistance changes in evolved populations. Grey dots: $\log_2$-transformed fold changes in IC_{90} values of all independently evolved populations; Curves: normal distribution approximation; Colored sticks: susceptible (blue) and resistant (red) breakpoints for the tested antibiotic resistance according to CLSI guidelines[21]. Box plot: center

line, median; small square, mean; left and right sides of the box, upper and lower quartiles; whiskers, 1.5× interquartile range. **c,** Averaged resistance changes (log_2-transformed fold changes in IC_{90} values) of evolved populations stressed by the same class of antibiotics. **d,** Positive and negative correlations between different antibiotic classes according to the antibiotic resistance profiles (Pearson's correlation coefficient $|r| > 0.6$, $P < 0.05$). **e,** Uniform manifold approximation and projection (UMAP) embedding of evolved populations based on their resistance to eight antibiotics. **c**, **d**, and **e** are assigned the same color code as **a** for antibiotic classes.

159 **Phenotypic convergence and divergence of antibiotic resistance evolution**

160 Pairwise antibiotic stressors induced convergent (Spearman's correlation
161 coefficient $\rho > 0.2$, 76%) and divergent ($\rho < -0.2$, 4.6%) antibiotic resistance
162 phenotypes among the 96 antibiotics (**Figs. 2a** and **2b**). Convergent phenotypes
163 (a phenomenon whereby developing similar phenotypes) were predominant and
164 occurred more frequently within the same class of stressors (85.3% vs. 73.9%, **Fig.**
165 **2c**). Divergent phenotypes (orthogonal), otherwise, were more found across
166 different antibiotic classes than within the same class (5.1% vs. 2.3%). Overall, the
167 frequency of convergence was much higher than the previously reported value
168 (60%) for *P. aeruginosa* strains that evolved under high-level antibiotic exposures[22].
169 We hypothesize that the survival of more diverse resistant subpopulations under
170 low-level antibiotic exposures likely contributes to widespread elevated population
171 antibiotic resistance, as these subpopulations may possess similar off-target
172 mutations[23]. However, the consistency of the mutation similarity and convergent
173 phenotypes await confirmation via deep sequencing. Furthermore, stressors within
174 aminoglycosides and lipopeptides (LPS) induced the most convergent resistance
175 phenotypes, while quinolones induced the most divergent resistance patterns
176 compared to other antibiotic classes (**Fig. 2d**). Convergent resistance phenotypes
177 were widespread across classes. Divergent phenotypes were more frequently
178 observed between aminoglycosides-quinolones and aminoglycosides-
179 tetracyclines, e.g., kanamycin and sparfloxacin ($\rho = -0.81$; $P = 0.0149$) than other
180 pairwise combinations (**Fig. 2e** and **Supplementary Fig. 7**). Our analyses of these
181 stressor-stressor pairs corroborated the previous findings of resistance evolution
182 leading to phenotypic convergence and identified new population-level convergent
183 resistance evolution in sub-MIC antibiotic selections[22].

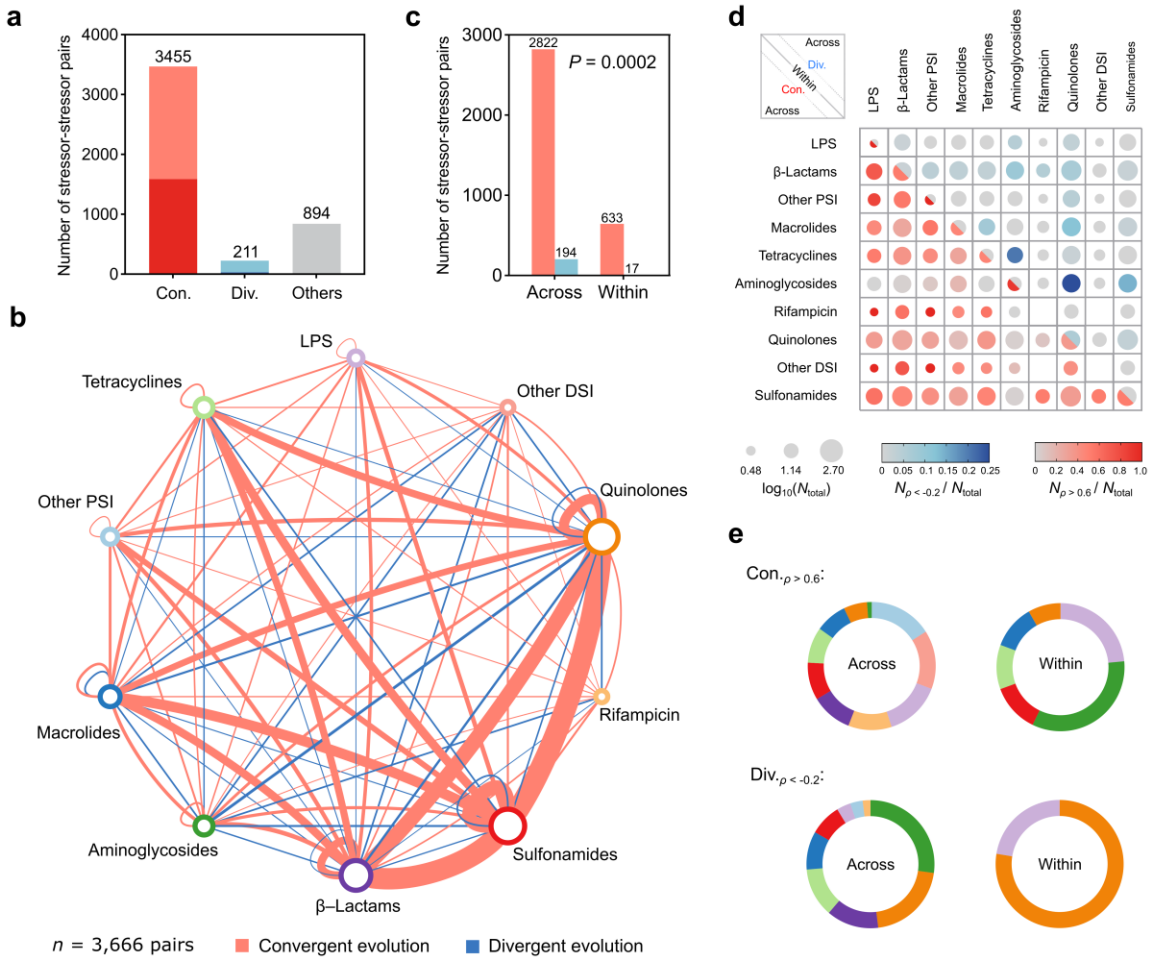

Fig. 2: Convergent and divergent evolution relationships based on antibiotic resistance profiles induced by low-level antibiotic exposures. **a,** Total number of convergent (con.) and divergent (div.) evolution relationships. Strong convergent or divergent evolution is presented in dark red or blue, respectively (Spearman's rank correlation $|\rho| > 0.6$); light red or blue suggests weak convergent or divergent evolution ($0.2 < |\rho| < 0.6$). **b,** Convergent and divergent evolution relationships between stressors belonging to different (across) or same (within) antibiotic classes. Node size reflects the number of stressors within each antibiotic class. Edge thickness reflects the number of convergent or divergent evolution pairs ($|\rho| > 0.2$). **c,** Convergent and divergent resistance evolution occurred between stressors across and within classes. Convergent: red; divergent: blue. *P* value: chi-square

test significance. **d,** Frequency of convergent evolution (strong, $\rho > 0.6$, red) and divergent evolution ($\rho < -0.2$, blue) relationships among antibiotic classes. Dot size reflects the number of stressor pairs within (diagonal) or across (upper and lower triangles) any two antibiotic classes. **e,** Proportions of convergent and divergent evolution relationships 'across' and 'within' antibiotic classes. **e** is assigned the same color code as **b** for antibiotic classes.

203 **Mutations in evolved populations**

204 For each stress condition, the four evolved cultures were pooled for deep
205 resequencing (coverage: 1,642 ± 466, n = 97, including one control group without
206 antibiotic exposure). A total of 56,744 mutation events were detected by filtering
207 the mutations in the control group, averaging 591 mutations per condition
208 (**Supplementary Fig. 8** and **Supplementary Data 3**), causing 4,750 amino acid
209 changes (AACs) involved in 2,432 protein-coding genes (**Fig. 3a**). These
210 mutations were classified into single nucleotide polymorphisms (non-synonymous
211 SNPs, 3,179), insertions (379), and deletions (1,428). Here, we confirmed the
212 consistency of convergent phenotypes with similar mutational backgrounds (**Fig.
213 3b**). Additionally, all antibiotic stressors selected more non-synonymous than
214 synonymous substitutions (dN/dS > 1, positive selection), and there were negative
215 correlations between dN/dS and the number of point mutations (**Fig. 3c**). Overall,
216 DSIs induced more point mutations than other stressors at the same dN/dS level,
217 consistent with their mechanisms of action.

218 The mutation counts of the 2,432 genes exhibit a power-law distribution with
219 a long tail, where the top ten mutant genes account for 6% of the total mutations
220 (**Fig. 3d**). In order to find mutation hotspots along the chromosomal DNA, we
221 performed mutation enrichment analysis in 5-kb sliding windows. The top three
222 mutation hotspots were flagged for the seven classes of antibiotic stressors (**Fig.
223 3e**). Some sites with frequent mutations included, e.g., the acid stress response
224 gene *evgS* and the multidrug efflux pump gene *emrY*. Interestingly, *kdpD* mutation
225 frequency was relatively high in both macrolides and aminoglycosides-stressed
226 populations. Recent studies report that KdpD promotes resistance to osmotic,
227 oxidative and antimicrobial stresses[24]. *yeaY* with frequent mutation under β-lactam
228 stressors encodes a putative outer membrane lipoprotein[25]. *rcsC* flagged under
229 LPS stress is a known ARG that exhibited increasing expression in response to

polymyxin B. Polymyxin E can activate the RcsC/YojN/RcsB two-component system, responding to alterations in the bacterial membrane[26].

Unsupervised clustering UMAP embedding was used to test the hypothesis that antibiotics with the same mechanism of action would enrich mutations leading to similar AACs (**Fig. 3f**). The hypothesis was largely verified with one exception-tetracyclines (PSI), which enriched AACs similar to those enriched under sulfonamides and quinolones stresses (DSI) rather than aminoglycosides and macrolides (PSI). This aligned well with the observed antibiotic resistance phenotypes stressed by these antibiotics. Tetracyclines could likely induce DNA damage due to their strong binding affinity to DNA and their ability to generate free radicals in close proximity to the DNA backbone[27,28]. Among the 2,718 highly variable AACs, 991 were enriched by at least one of the four action mechanisms of stressors (Fisher's exact test $P < 0.05$). DSI antibiotics contributed the most to the overall enrichment (722 hits), and the majority of overlapping hits were observed between PSI and CWSI antibiotics (**Fig. 3g**).

We next used Gene Ontology (GO) enrichment analysis to understand the co-occurrence of mutant gene-associated biological processes for the seven classes of antibiotics (**Fig. 3h**). 'DNA metabolic process' was represented among the top hits by sulfonamides and quinolones (DSI), and 'outer membrane' was the most targeted gene category by LPS stressors (MSI). Meanwhile, tetracyclines caused the most mutant genes in the 'transporter' category, including but not limited to those related to carbohydrate transport and metal ion transport. Methylation of the 16S rRNA is an emerging mechanism of high-level resistance to PSI antibiotics via ribosomal methyltransferases[29,30]. Mutant genes associated with this process were significantly more prevalent in the presence of macrolides and aminoglycosides compared with other antibiotics.

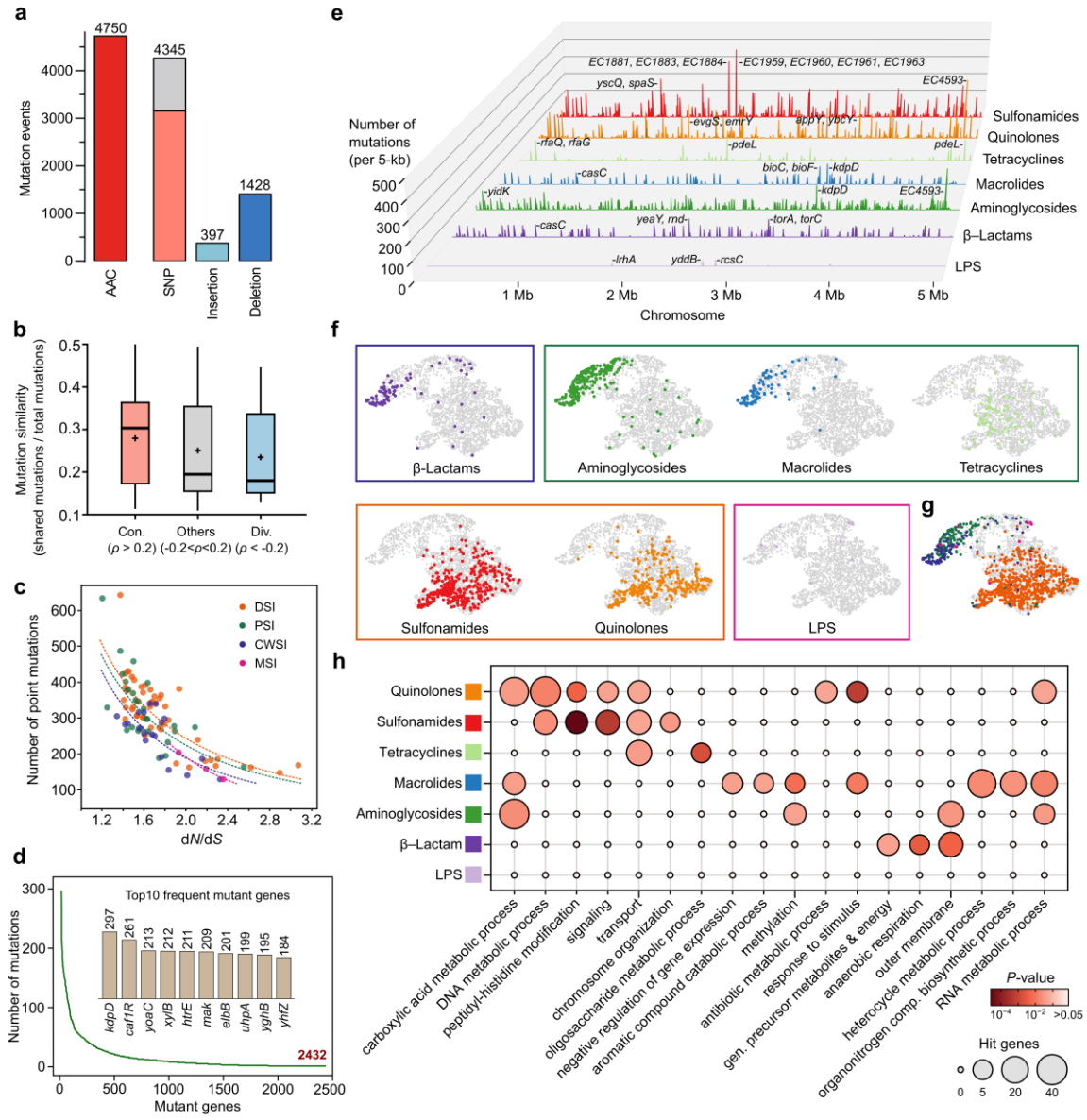

Fig. 3: Mutations detected in evolved populations. **a,** Distribution of mutation events. SNPs are classified into non-synonymous (red) and synonymous (grey) ones. **b,** Mutation similarity in evolved populations exposed to stressors exhibiting convergent or divergent evolutionary relationships. Box plots: center line, median; '+', mean; limits, first and third quartiles; upper and lower whiskers, the highest and lowest values within 1.5× the interquartile range from the hinge. Mutation similarity levels are significantly different for evolved populations exhibiting convergent, other, and divergent relationships (F = 44.59, $P < 0.0001$; one-way ANOVA,

Tukey's honestly significant difference). **c,** Relationships between non-synonymous to synonymous mutation ratio (d*N*/d*S*) and the number of point mutations. Dashed curves represent non-linear regressions performed for the four groups of stressors (R^2 = 0.59-0.99, all $P < 0.05$). **d,** The distribution of mutation counts for a total of 2,432 mutant genes. The top ten mutant genes harboring the most mutations are displayed. **e,** Mutation hotspots in the chromosomes of evolved populations under different antibiotic stresses. The top three mutation hotspots are highlighted for each antibiotic class. **f,** UMAP clustering of highly variable AACs (n = 2,718) occurred under the stresses of seven antibiotic classes, which are further grouped based on the action mechanisms of antibiotics (four groups, rectangle frames). **g,** UMAP clustering of highly variable AACs among the four groups of stressors. **h,** GO biological processes associated with mutant genes for different classes of antibiotic stresses.

Mutant resistance-relevant genes and co-fitness interaction network

Recognizing that not all mutant genes directly contribute to high-level resistance, we employed Tn-Seq to determine gene essentiality and identify potentially resistance-relevant genes (PRGs) across the whole genome. Successful construction of the Tn library using the WT strain enabled high-confidence essentiality and fitness quantification in the presence and absence of eight antibiotics used in MIC assays (**Supplementary Fig. 9**). Overall, the WT strain genome contained 468 essential, 3,910 non-essential, and 367 uncertain genes (**Fig. 4a**). Genes with absolute fitness change $|\Delta W| > 0.15$ for at least one antibiotic were considered as PRGs (n = 2,830, **Fig. 4b** and **Supplementary Data 4**). Their proportions increased from 55% across the whole genome to approximately 73% among mutant genes (**Fig. 4a**). This highlighted that PRGs were indeed selected during experiment evolution under antibiotic stresses.

To map the co-fitness interactions of pairwise PRGs, a 2,830 × 2,830 correlation matrix was calculated based on their conditional fitness profiles. A total of 22,811 co-fitness interactions were identified for 2,152 PRGs (Pearson's $|r| > 0.75$, corrected $P < 0.00002$), including 15,403 positive and 7,408 negative interactions that were used for constructing the high-confident co-fitness interaction network (**Fig. 4c**). Highly connected hub PRGs (Maximal Clique Centrality, MCC score > 1*e*20) were identified, forming two central regions in the interaction network. Spatial analysis of functional enrichment (SAFE) further identified 15 network regions spatially enriched for similar bioprocesses, encompassing 1,093 PRGs in total (high-enrichment PRGs, **Fig. 4d** and **Supplementary Data 5**). In addition to some well-studied resistance-associated bioprocesses such as DNA replication and repair, response to stimulus, adhesion, and metabolism, the SAFE analysis also uncovered the enrichment of bioprocesses previously unreported (e.g., carbon/nitrogen/sulfur utilization) or less

307 known (e.g., metabolism or cell motility[31,32]) to be associated with antibiotic
308 resistance. How these cellular bioprocesses influence resistance evolution under
309 low antibiotic stresses warrants further investigation.

310 Mutant PRGs (n = 1,572) that potentially contribute to the high-level antibiotic
311 resistance of evolved populations, i.e., core mutant PRGs (n = 131, P < 0.01,
312 Fisher's exact test), were identified by comparing those enriched in populations
313 exhibiting the top 10 versus bottom 40 resistance increase (**Supplementary Fig.**
314 **10**). The majority of core mutant PRGs were linked to high-level resistance against
315 TOB, CAZ, MER, and PMB (**Fig. 4e, 4f,** and **Supplementary Data 6**). A
316 representative subset comprising 22 out of the 131 core mutant PRGs was
317 selected for more detailed investigation on their mechanisms of resistance
318 contribution, including 9 hub genes identified in **Fig. 4c**. Some of these
319 representative core mutant PRGs participate in bioprocesses that were spatially
320 enriched by SAFE analysis (high enrichment, **Fig. 4g**), such as response to
321 stimulus (e.g., *ynfB*[33]), signaling (*focC*[34]), and carbohydrate transport (*ycjV*[35]).
322 While other core mutant PRGs contribute to bioprocesses that were not spatially
323 enriched (low-enrichment), such as transcriptional reprogramming genes (e.g.,
324 *rtcR* and *narL*), they may still play roles in the development of AMR. This is likely
325 because compared with high-enrichment PRGs, the low-enrichment ones exhibit
326 higher degrees of co-fitness interactions (**Supplementary Fig. 11**), indicating their
327 potential to interact with broader PRGs and thereby facilitate AMR development.

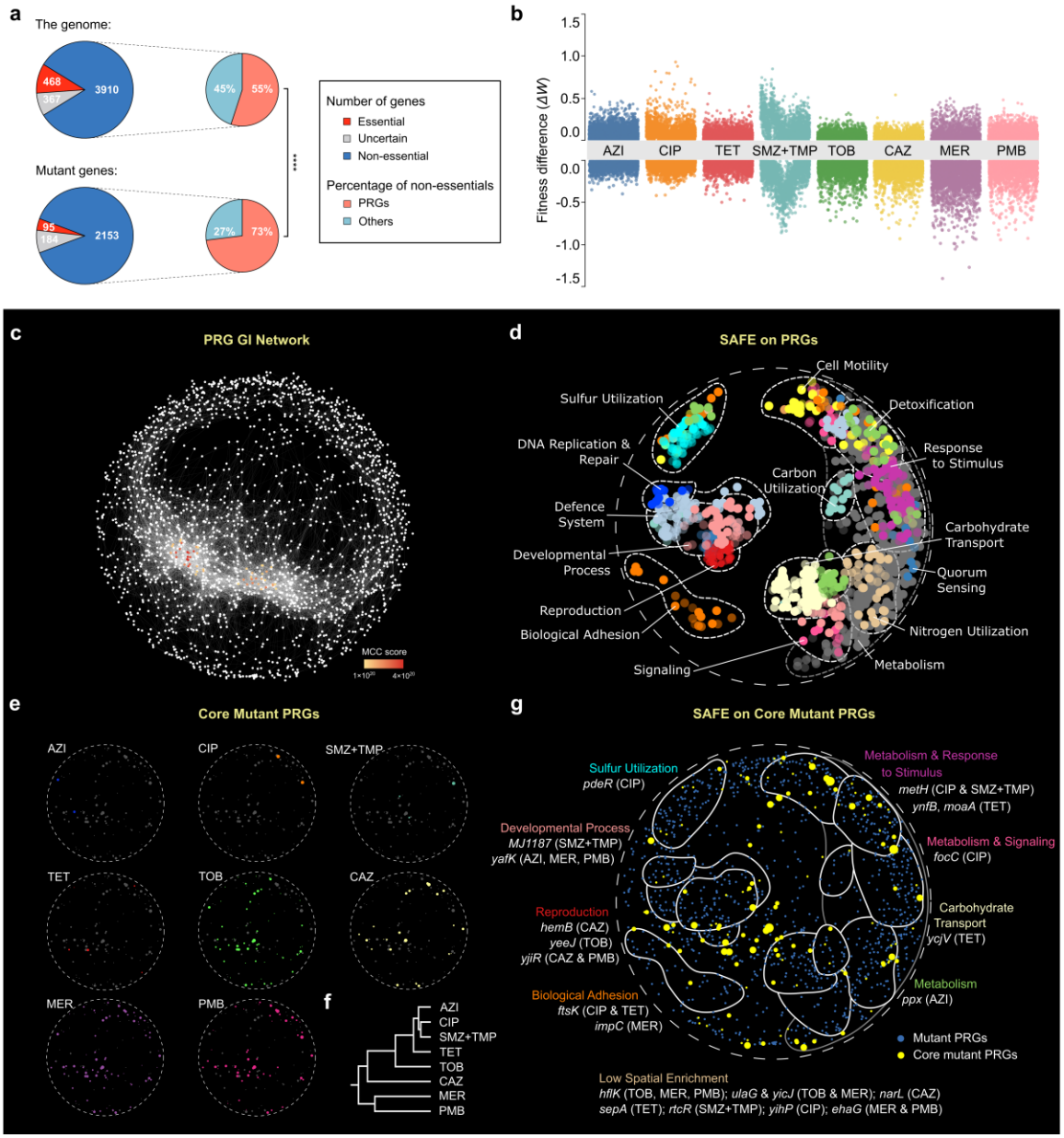

Fig. 4: Gene co-fitness interaction network and spatial analysis of functional enrichment. **a,** Number of essential, non-essential, uncertain genes in the whole genome and mutant genes. ****: chi-square test $P < 1e{-}4$. **b,** Fitness differences (ΔW) of all Tn-Seq library genes for the 8 tested antibiotics. **c,** Gene interaction (GI) network of PRGs (n = 2,152) based on their co-fitness matrix. **d,** Enriched 15 bioprocesses by SAFE analysis on the PRG GI network in **c**. **e,** Core mutant PRGs conferring high-level resistance to the eight antibiotics (nodes, n = 131). Node size

336 reflects the co-occurrence frequency of core mutant PRGs. **f,** Hierarchical
337 clustering of the eight antibiotics based on their associated core mutant PRGs. **g,**
338 Mutant PRGs (blue) and core mutant PRGs (yellow) are highlighted on a
339 schematic representation of the GI network. The 22 labelled representative core
340 mutant PRGs are associated with SAFE high-enrichment bioprocesses (within the
341 circle, $P < 0.05$, corrected for multiple testing) or low-enrichment bioprocesses
342 (outside the circle).
343

344 **Resistance mechanisms of representative core mutant PRGs**

345 Among the 22 representative core mutant PRGs, 16 have reported links to
346 specific resistance mechanisms, including biofilm formation (e.g., *focC* and *pdeR*),
347 efflux pump (e.g., *ycjV*), drug inactivation (e.g., *MJ1187*), target bypass (e.g., *yafK),*
348 etc. We are therefore interested in uncovering the contribution mechanisms of the
349 remaining 6 representative core mutant PRGs with unreported links to resistance
350 (i.e., *metH, ynfB*, *ulaG, ftsK*, *rtcR*, and *hflK*) by constructing their gene interaction
351 modules (module I-VI, respectively, **Fig. 5a**). Within their modules, these PRGs
352 interacted with genes encoding biofilm formation, metabolism, efflux pump, target
353 alteration, membrane permeability, and DNA replication & repair-related proteins.
354 On the one hand, these connections indicated that the 6 PRGs likely contribute to
355 AMR development through similar mechanisms. On the other hand, analyzing the
356 fitness differences (ΔW) of PRG-connected genes in response to antibiotic
357 exposure further refined the proposed resistance mechanisms (**Fig. 5b**). For
358 example, the fimbrial adhesin protein-coding gene *faeJ* and biofilm formation
359 associated gene *vasD*, which interacted with *rtcR* (module V), exhibited significant
360 fitness differences ($|\Delta W| = 0.30 \pm 0.05$ under SMZ+TMP stress). Similarly, the
361 interactions of *hflK* (module VI) with the fimbrial adhesin protein-coding genes
362 *yehB*, *afaA*, and biofilm formation associated gene *suhB* ($|\Delta W| = 0.35 \pm 0.19$ in
363 response to TOB, MER, and PMB stresses) suggested that *hflK* may also
364 contribute to AMR development through affecting biofilm formation.

365 The proposed resistance mechanisms of *rtcR* and *hflK* were further validated
366 via gene knockout experiments. The two knockout strains exhibited increased
367 susceptibility to their respective antibiotics (**Fig. 5c**), prolonged lag phases in
368 growth ($P < 0.05$, Student's *t*-test; **Fig. 5d**), and decreased biofilm formation
369 (approximately 72% lower than that of the WT strain, **Fig. 5e**). These results
370 collectively suggest that $rtcR^{M}$ and $hflK^{M}$ may confer resistance by enhancing

biofilm formation, thereby limiting antibiotics diffusion into the population. The resistance mechanisms of the remaining four representative core mutant PRGs (*metH, ynfB*, *ulaG,* and *ftsK)* could be proposed and validated using similar approaches in future studies.

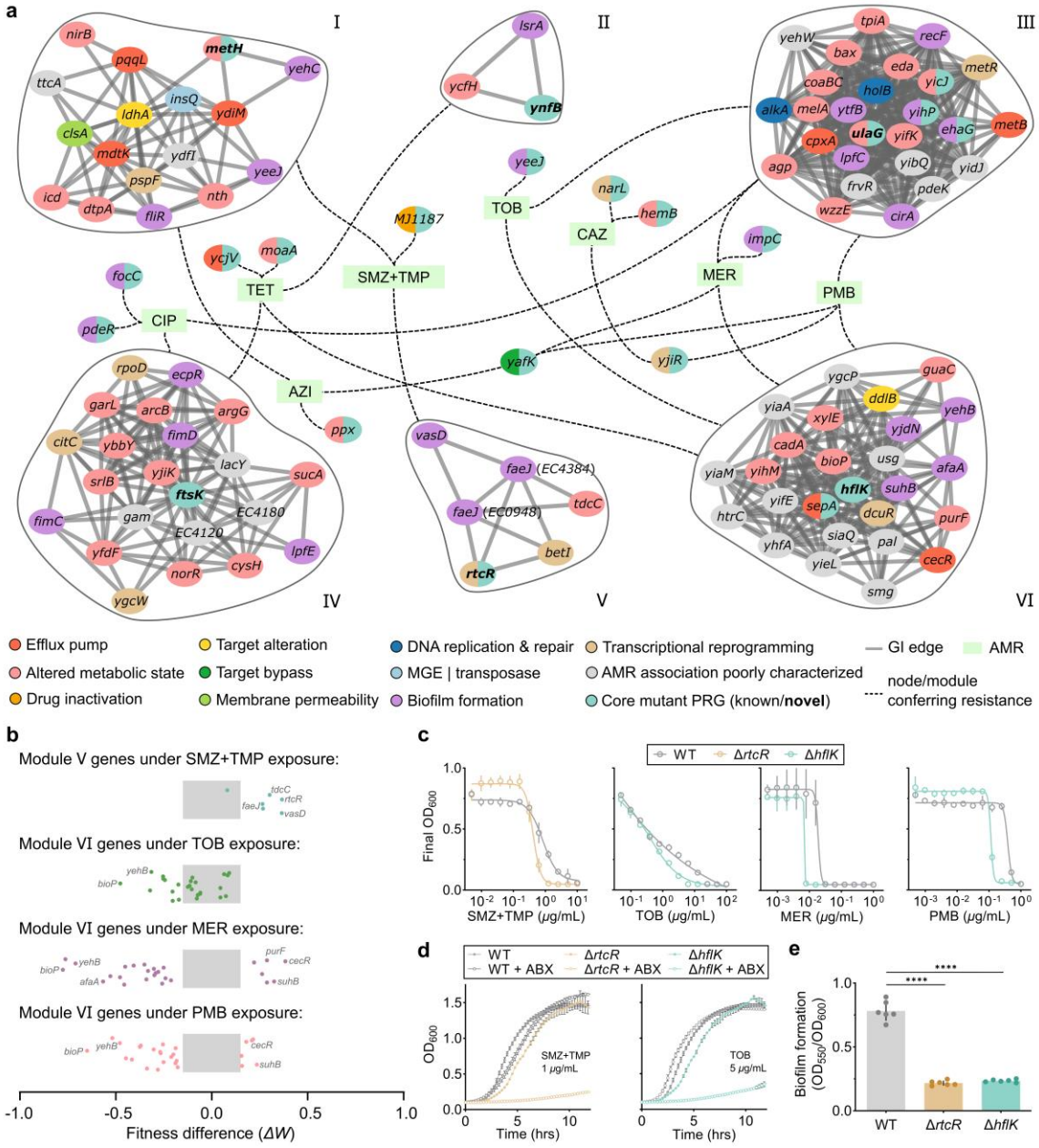

Fig. 5: Gene co-fitness interaction modules of selected core mutant PRGs and verification of their proposed resistance mechanism. **a,** Core mutant PRGs displayed within modules extracted from the gene co-fitness interaction network

379 (in bold, with unknown resistance mechanism) or outside modules (with known
380 resistance mechanisms). **b,** Fitness differences of genes in modules V and VI
381 under specific antibiotic exposures. Grey regions represent abs(ΔW) < 0.15. **c,**
382 Dose–response curves for the three strains exposed to antibiotics at varied
383 concentrations. Error bars represent standard deviations from biological triplicate.
384 **d,** Decreased resistance of Δ*rtcR* and Δ*hflK* strains to SMZ+TMP and TOB,
385 respectively, compared with the WT strain. **e,** Biofilm formation potentials of the
386 WT, Δ*rtcR*, and Δ*hflK* strains determined by the crystal violet assay. $N = 6$ for all
387 strains. ****: $P < 0.0001$.
388

389 **Resistance evolution prediction by machine-learning models**

390 A long-standing challenge in AMR research has been to predict the outcome
391 of resistance evolution based on the chemical structure and/or dose of stress
392 compounds. The 96 antibiotics used in this study are adequately representative of
393 the structure diversity of commonly applied antibiotics. We developed machine
394 learning models that can correlate the chemical structures of antibiotics,
395 represented by extended-connectivity fingerprints (ECFPs), with the antibiotic
396 resistance profiles of evolved strains. Automated generations by genetic
397 algorithms were employed for feature selection and model optimization. The
398 ensembled regression models demonstrated varied modeling performance on the
399 training datasets, with coefficient of determination (R^2) values ranging from
400 approximately 0.3 to 0.8 for the eight antibiotics (**Fig. 6a**). The modeling
401 performance on CIP resistance (R^2 = 0.79 ± 0.01) was better than those for other
402 antibiotics (F = 31.48, P < 1e–4, two-way ANOVA). We then used CIP resistance
403 prediction results as an example to understand the contributions of monomer
404 structures to antibiotic resistance. The top five antibiotics leading to high-level CIP
405 resistance changes ($\log_2$ fold-change from 5.16 to 7.32) were annotated for their
406 atomic groups' contributions to the CIP resistance (**Supplementary Fig. 12)**. The
407 key contributors to CIP resistance changes were the fluorine atom introduced at
408 the C-6 position and the cyclopropyl group introduced at the N-1 position of the
409 parent nucleus.

410 To extend the applicability of the machine learning models, we established the
411 applicability domain based on the Tanimoto similarity network. Among the
412 additional 1,563 antibiotics selected from PubChem, 569 (36.4%) with a Tanimoto
413 similarity >0.2 compared to the 96 antibiotics used in this study were considered
414 applicable (**Fig. 6b** and **Supplementary Data 7**). The best-performing models
415 were used to predict the antibiotic resistance profiles stressed by these applicable

416 antibiotics, resulting in resistance fold changes ($\log_2$ transformed) from −3.32 to
417 5.58 (**Fig. 6c** and **Supplementary Data 8**). To further evaluate the predictive
418 accuracy of the models, we conducted experimental evolution using 16 external
419 antibiotics (within the applicability domain and encompassing eight antibiotic
420 classes, **Supplementary Table 1**). The predicted resistance levels to CIP and CAZ
421 exhibited agreement with their actual values (R^2_{test} > 0.75, **Supplementary Fig.**
422 **13**). These chemical structure-based models demonstrate promising potential for
423 predicting antibiotic evolution under low-level antibiotic stresses.

424 In addition to ECFP modeling, we also explored mutation-based antibiotic
425 resistance prediction. The AAC model was constructed based on the random forest
426 (RF) algorithm, which indeed improved the prediction performance (R^2_{AAC}: 0.73 ±
427 0.05 vs. R^2_{ECFP}: 0.51 ± 0.15 for the training dataset, **Fig. 6a** and **Supplementary**
428 **Fig. 14**). Despite proof-of-concept demonstrations, the applications of AAC-based
429 modeling in predicting resistance phenotypes remain to be validated due to the
430 complexity in intrinsic and extrinsic factors influencing mutation outcomes, e.g.,
431 genetic background and nutrient availability.

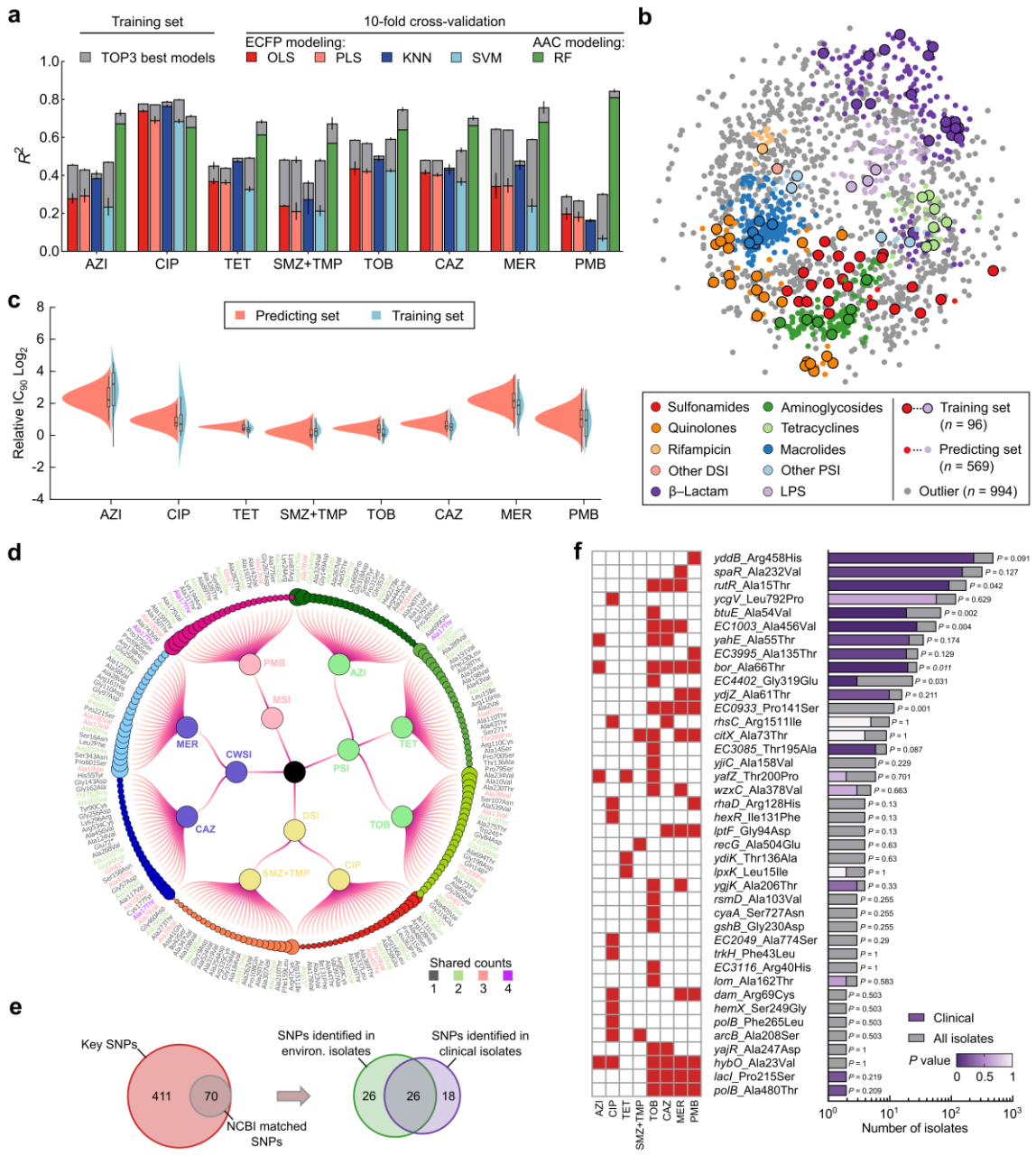

Fig. 6: Performance of machine learning models and prevalence of key point mutations in clinical/environmental isolates. **a,** Performance of the machine learning models. The error bars represent the standard error of mean from the top three best-performing models. **b,** Tanimoto similarity network demonstrating the applicability domain (Tanimoto similarity ≥ 0.2) of ECFP-based models. **c,** Split violin plots showing the normalized distributions of predicted resistance levels for the training (n = 96) and predicting antibiotics (n = 569). **d,** Top 25 most important

SNPs (features) contributing to the accuracy of the AAC-based RF model for each antibiotic. Dot size: importance of features; Shared counts: occurrence of particular SNPs in the circle. **e,** Occurrence of key SNPs (feature-importance SNPs in **d** and SNPs in core mutant PRGs conferring high-level resistance, n = 481 in total) in NCBI-retrieved clinical and environmental isolates. **f,** Left: key SNPs associated with resistance to the eight antibiotics. Right: number of clinical and environmental isolates containing the top 40 key SNPs. Total isolates: 1,263 retrieved from the NCBI Pathogen Detection database. Fisher's exact test P values indicate the overrepresentation of SNPs in clinical isolates and are colored in varying degrees of purple, from highly to not significant (dark to light purple, respectively).

451 **Clinical and environmental prevalence of resistance-relevant mutations**

452 Assessing the prevalence of key mutations in a broader range of clinical or
453 environmental isolates is crucial for evaluating their risks in the 'One Health'
454 paradigm. We focus on two groups of point mutations, comprising a total of 481
455 key SNPs, i.e., feature-importance SNPs contributing to the accuracy of the AAC-
456 based RF models (**Fig. 6d**) and SNPs conferring high-level antibiotic resistance
457 (**Supplementary Fig. 10**). These key SNPs and 1,263 *E. coli* genomes (clinical:
458 577, environmental: 686) retrieved from the NCBI Pathogen Detection database
459 were subjected to comparative genomic analysis (**Supplementary Data 9 and 10**).
460 Overall, 70 of the 481 key SNPs (14.6%) were identified in at least one of the
461 external genomes (**Fig. 6e**), with the highest occurrence in 487 genomes (*yddB*
462 Arg458His, **Fig. 6f**). Their relative prevalence in clinical and environmental isolates
463 was summarized in **Supplementary Fig. 15**. Importantly, 62.9% of the key SNPs
464 were identified in clinical isolates, and three were significantly more prevalent in
465 clinical than environmental isolates, e.g., TetR family of transcription regulator
466 gene *rutR* Ala15Thr and lipoprotein-related gene *bor* Ala66Thr. Furthermore, 74.3%
467 of the key SNPs were identified in environmental isolates, and three were
468 significantly more prevalent, e.g., oxidative stress response gene *btuE* Ala54Val.
469 Notably, co-occurring SNPs (i.e., detected in both clinical and environmental
470 isolates) accounted for 37.1% of the total key SNPs. These results suggested that
471 mutations that significantly contribute to antibiotic resistance are widespread in
472 both clinical and environmental isolates, which may be acquired under sub-MIC
473 antibiotic exposure conditions.
474

Discussion

Despite the growing attention that the development of AMR could be driven by the low-level selective pressure of antibiotics released into the environment, the relationship between AMR phenotypes and genotypes, and whether mutations conferring high-level resistance are prevalent remain key questions to be addressed. In this study, we constructed an unprecedented phenotype-genotype atlas of resistance evolution under 96 low-level antibiotic stress conditions. We used high-throughput experimental evolution and deep-resequencing approaches to determine, at the population level, how bacteria can evolve high-level resistance upon exposure to a wide range of antibiotics at low levels. Tn-Seq further facilitated the identification of PRGs and the elucidation of associated resistance mechanisms. ECFP- and AAC-based machine learning models have shown promise in areas such as predicting resistance phenotypes and identifying feature-importance antibiotic substructures and mutations. We leveraged an extensive collection of publicly available genomes to determine the prevalence of key mutations in a broader range of environmental and clinical settings. Some of the main findings extend beyond our understanding of resistance evolution under low-level antibiotic exposures, shedding insights into the origins and dissemination of antibiotic resistance in the environment.

From a phenotypic perspective, most stressors showed positive correlations in causing consistent antibiotic resistance changes in evolved populations. Cross-resistance was much more prevalent than collateral sensitivity[16], reinforcing and complementing the previous findings of cross-resistance and collateral sensitivity relationships identified at clinical concentrations[36,37]. Some of the newly identified collateral sensitivity relationships (e.g., kanamycin or gentamicin vs. sparfloxacin, and aztreonam vs. ciprofloxacin; $P < 0.05$) offer valuable insights for developing alternative clinical antibiotic dosing strategies. The study is notable for pioneering

502 the establishment of a link between stressor's chemical structures and resistance
503 evolution. For instance, quinolone stressors elicited the highest CIP resistance
504 compared with other antibiotics, and their particular atomic groups, e.g., the C-6
505 fluorine atom and the cyclopropyl group, were found likely to contribute to high-
506 level resistance evolution.

507 From a genotypic perspective, the results provide a comprehensive
508 understanding of how critical mutations contribute to clinically relevant resistance
509 and the evolutionary landscape. Frequently encountered ARGs such as *evgS* and
510 *ermY* were among the most common targets of mutations under antibiotic stresses.
511 Tn-Seq further enabled the construction of the GI network in response to antibiotics.
512 We, therefore, identified a set of core mutant PRGs (n = 131) contributing to high-
513 level resistance evolution through the phenotype-genotype atlas. In addition to the
514 well-characterized ARGs, e.g., *pdeR* involved in biofilm formation[38], *ycjV* involved
515 in efflux pump[35,39], *MJ1187* involved in drug inactivation[40], *yafK* involved in target
516 bypass[41,42], and yjiR involved in transcriptional reprogramming[43], others may
517 confer high-level resistance via unreported mechanisms, e.g., *metH*, *ynfB*, *ulaG*,
518 *ftsK*, *rtcR*, and *hflK*. By locating these understudied core mutant PRGs in the gene
519 co-fitness interaction network and analyzing the functions of interacted genes, their
520 resistance mechanisms were proposed. The resistance contribution mechanism of
521 two core mutant PRGs, namely *rtcR* and *hflK*, through biofilm formation, was
522 experimentally confirmed. Importantly, this demonstrated that the gene co-fitness
523 network generated based on Tn-Seq data highlights valuable interactions that can
524 be explored to uncover resistance mechanisms of potential ARGs. The SAFE
525 analysis further suggests that genome-wide mutations affecting broad biological
526 processes[44,45], such as metabolism and cell motility, can significantly contribute to
527 the evolution of antibiotic resistance[31,32], extending beyond independent, target-
528 gene-related mutations.

529 The ECFP- and AAC-based machine learning models exhibited promising
530 levels of accuracy and scalability in elucidating the intricate relationships between
531 resistance phenotype, genotype, and underlying drivers (i.e., antibiotics),
532 leveraging the power of vast data that are unattainable through traditional low-
533 throughput evolutionary experiments. Feature-importance antibiotic substructures
534 identified may have important implications for the development of next-generation
535 antibiotics and the targeted pollutant removal from contaminated environments[46].
536 The ECFP-based models exhibit broad applicability across a wide range of
537 antibiotics, enabling relatively reliable prediction of antibiotic resistance in *E. coli*
538 strains under low-level antibiotic exposures without the need for intensive, high-
539 throughput experimental evolution. Moreover, the predicted resistance levels can
540 be used as a reference, to some extent, for assessing the risks of environmentally
541 acquired resistance.

542 Key SNPs selected at sub-MICs indeed exhibited both environmental and
543 clinical relevance across a vast collection of *E. coli* isolates, and their environment-
544 clinic co-occurrence rates were up to 37.1%. Some of these key SNPs were even
545 enriched in clinical isolates, suggesting that the variants may already pose health
546 risks. Consequently, antibiotic-contaminated aquatic and soil environments could
547 serve as reservoirs of evolved resistant bacteria carrying clinically relevant, high-
548 resistance-conferring mutations. A few studies have attempted to determine the
549 genetic relationships between environmental and clinical isolates in public
550 databases to infer transmission dynamics[47,48]. Our results made additional
551 contributions to the identification of key mutations and ARGs of concern, as well
552 as the quantification of their prevalence.

553 Disentangling the AMR risk is challenging and requires a holistic
554 understanding of environment-clinic AMR dissemination, including but not limited
555 to its evolution under low-level antibiotic exposures. Some considerations should

be taken into account when extending our results to other studies on environmental resistance. First, environmental pathogens are remarkably diverse, likely resulting in species-specific response and resistance evolution patterns that differ from those of *E. coli*. Second, environmental changes are far more drastic and dynamic in nature, and resistance evolution is an ongoing process. Efforts to predict resistance profiles, either based on chemical structure or mutations, are inevitably constrained by temporal limitations and complex interplay of influencing factors. Third, uncovering the resistance-conferring mechanisms of core mutant PRGs is based on a hypothesis-verification approach. However, even if the hypotheses are experimentally validated, they may not be the primary mechanisms conferring resistance due to the pleiotropy of genes. We are actively investigating the extent to which these lab-scale findings from experimental evolution mirror AMR development in real-world settings, particularly antibiotic-contaminated environments. Nonetheless, these efforts provide valuable scientific support for combating antibiotic resistance within the One Health framework.

572 **Methods**

573 **Bacterial strains, culture conditions, and antibiotics**

574 *E. coli* strain QSHQ1 (GenBank accession number: CP128989) was a
575 plasmid-free opportunistic pathogen that was originally isolated from soil in
576 Hangzhou, China (**Supplementary Fig. 16**). Cells were cultivated with LB medium
577 (150 rpm, 37°C) in Abgene™ 96-well 2.2 ml polypropylene deep-well plates
578 (Thermo Scientific, #AB0661) with breathable sealing film (Corning, #1223S04).
579 All antibiotics used were purchased from Yuanye-Bio, Macklin, and Solarbio
580 (details in **Supplementary Data 1**). Antibiotic stock solutions (dissolved in 10%
581 sterile DMSO) were filter-sterilized and stored at −20°C before use.

582

583 **Experimental evolution under 96 antibiotic stresses**

584 Pre-adaptation: A single colony of *E. coli* strain QSHQ1 was picked from the
585 LB-agar plate and grown in LB medium without antibiotics for 24 h. Cells were sub-
586 cultured daily by 500-fold dilutions in fresh LB medium over a 10-day period,
587 corresponding to approximately 9 generations per day.

588 Experimental evolution: 96 antibiotics were added to the LB-medium at a
589 uniform concentration of 10 μg l^{-1}. Four independent cultures (biological
590 quadruplicate) were grown under each antibiotic stress and were sub-cultured
591 daily by 500-fold dilutions for 20 days (1 ml medium, 96-well plates, 150 rpm, 37°C).
592 The antibiotic concentration was selected to reflect the upper range of
593 environmentally relevant concentrations detected in, e.g., wastewater treatment
594 plants[49]. Four control lineages were cultured under the same condition without
595 antibiotics. Cell growth was monitored by measuring OD_{600} using the Spark
596 microplate reader (Tecan Inc.). After the 20-day experimental evolution, end-point
597 cultures were frozen at −80°C in 30% glycerol for whole-genome sequencing. No
598 growth was observed in the negative control wells throughout the experimental

evolution.

Automated high-throughput MIC assays of evolved populations

The automated high-throughput MIC assays were carried out on the iBioFoundry platform at ZJU-Hangzhou Global Scientific and Technological Innovation Center. The platform consisted of a MultiDrop Combi microplate dispenser (Thermo Scientific, #5840300), a Tecan Freedom Evo 200 Base laboratory automation workstation, a static incubator, and a CLARIOstar Plus microplate reader (BMG LABTECH Inc.), which were operated by two six-axis robotic arms (Thermo Scientific, TS-F7000).

Antibiotic resistance against azithromycin (AZI), ciprofloxacin (CIP), tetracycline (TET), sulfamethoxazole with trimethoprim (SMZ+TMP), tobramycin (TOB), ceftazidime (CAZ), meropenem (MER), and polymyxin B (PMB) were measured for evolved cultures at the end of the experimental evolution. The end-point glycerol-frozen cultures were revived in 1 ml LB medium for 16 h at 37°C with (evolved populations) or without (control) stressing antibiotics. The revived cultures were 2000-fold diluted in freshly prepared MIC plates to a final volume of 130 μl. The 2-fold serial dilutions were performed over 12 steps, starting from the highest concentrations of 400, 25, 100, 200, 100, 200, 25, and 400 mg l^{-1}, respectively, for the eight antibiotics. After 20 h static incubation, the OD_{600} values were measured. The dose-response curves generated from biological quadruplicates were fitted to the Hill equation, and the antibiotic concentration corresponding to an OD_{600} of 0.2 was defined as IC_{90}. In the event of curve overflow (exceeding 100% inhibition) or underfill (below 90% inhibition), the IC_{90} values were assigned as the maximum tested concentration or half of the minimum tested concentration, respectively.

DNA extraction from evolved populations

In the antibiotic resistance assays, iBioFoundry identified the first well with

626 detectable cell growth (OD_{600} > 0.2) in the presence of antibiotics, and pooled cells
627 from up to 32 of these wells together (8 antibiotics × 4 replicates, ~4 ml). Genomic
628 DNA (gDNA) was extracted from pooled cultures using the DNeasy Blood & Tissue
629 kit (Qiagen) in accordance with the manufacturer's instructions. DNA quantity and
630 quality were determined by the Colibri microvolume spectrometer (Titertek-
631 Berthold Inc., Germany). High-quality DNA (A_{260}/A_{280} = 1.8, > 6 μg) was used for
632 whole-genome sequencing.
633

634 **Whole-genome sequencing and mutation analysis**

635 Whole-genome sequencing was performed for the WT strain *E. coli* QSHQ1
636 as well as the 96 evolved populations, following the procedures described
637 previously[10]. Briefly, the WT strain was sequenced using PacBio RS combined with
638 Illumina sequencing (NovaSeq6000). The Illumina data were used to evaluate the
639 complexity of the genomes and correct errors in PacBio long reads. The high-
640 quality complete genome of the WT strain (5,084,285 bp) served as the reference
641 for mutation detection (**Supplementary Fig. 17**). Pooled evolved populations were
642 deep resequenced on the Illumina platform (150 bp×2, ~1,500x coverage,
643 Shanghai BIOZERON Co., Ltd). Mutations detected in the control lineages were
644 excluded from further analysis. Variants with a frequency greater than 0.2% in the
645 population were identified. The non-synonymous (d*N*) and synonymous (d*S*)
646 substitutions were obtained under each stress condition, and the proportion of
647 beneficial mutations was indicated by the *dN/dS* ratio, which was set to 1 for neutral
648 evolution (without antibiotic stress).
649

650 **Transposon sequencing and gene fitness assays**

651 A transposon mutant library for the WT strain was generated by
652 electroporation using the EZ-Tn5 <KAN-2> Tnp transposome kit (Lucigen Inc.).

653 The library consisted of approximately 40,000 insertion mutants, achieving a
654 genome-wide saturation of 97.5%. The library was divided into 9 aliquots, 500x
655 diluted in LB medium and grew for 24 h (4-8 generations) with no antibiotic or one
656 of the eight antibiotics at sub-MIC levels. The gDNA was extracted, and a specific
657 PCR forward primer (GATCCTCTAGAGTCGACCTGCAGG) was designed to
658 amplify the regions flanking the Tn5 insertion (TA) sites. TA sites were sequenced
659 on Illumina NovaSeq6000, and the sequencing data were analyzed using the
660 TRANSIT package v3.2.7[50]. Essential genes were identified based on the Gumbel
661 method by calling long consecutive sequences of TA sites without insertions. For
662 non-essential genes with valid TA sites > 5 and mean insertion counts (N) > 5,
663 roughly 3,500 in total, their gene fitness W was calculated as follows:

670 $$W_i = ln(N_i(t_2)/N_i(t_1))$$

664 where $N_i(t_1)$ and $N_i(t_2)$ were the mean insertion counts of non-essential gene i
665 detected in a particular library before and after antibiotic selection, respectively.
666 The differences in the gene fitness were therefore calculated as $\Delta W = W_{ABX} -$
667 W_{noABX}. Genes with absolute fitness difference values $|\Delta W|$ larger than 0.15 under
668 at least one antibiotic stress condition were identified as potentially resistance-
669 relevant genes (PRGs).

671

672 **Gene co-fitness interaction network and spatial analysis of functional**
673 **enrichment (SAFE)**

674 A co-fitness interaction network for PRGs was constructed as described
675 previously[51]. In brief, a gene fitness matrix was built for a total of 2,830 PRGs,
676 which was further used to calculate Pearson's correlation coefficients (r) for all
677 pairwise genes. A high-confidence co-fitness interaction network was established
678 using Cytoscape, comprising 2,152 PRGs (nodes) and 22,811 interactions (edges)
679 with $|r| > 0.75$ and corrected $P < 0.00002$ (threshold: 0.05, corrected by the number

680 of genes). The plugin CytoHubba was used to calculate the degree and maximal
681 clique centrality (MCC) of all nodes in the network. The plugin ClusterViz was
682 employed to identify distinct modules based on network topology using the FAG-
683 EC algorithm. PRGs were annotated using the Swiss-Prot, GO, CARD, VFDB, and
684 BacMet databases. SAFE highlighted regions of the gene interaction network
685 enriched in specific functional attributes.
686
687 **Knockout of selected PRGs and biofilm formation potential measurement**
688 Two core mutant PRGs, *rtcR* and *hflK* were subjected to knockout
689 experiments to verify their hypothesized involvement in biofilm formation. The
690 single guide RNAs (sgRNAs) were designed using the online design tool
691 (https://en.rc-crispr.com/). The pair of oligos for the two target genes were
692 annealed and ligated to the CB-001 vector (Ubigene Biosciences Co., Ltd.,
693 Guangzhou, China). The CB-001-*rtcR* (or *hflK*) [gRNA] plasmids containing each
694 target sgRNA sequence and donor were transformed into the WT strain by
695 electroporation. Cells were spread onto LB agar and incubated overnight.
696 Transformants were identified by colony PCR and Sanger sequencing. The
697 sequences of sgRNAs and primers for CRISPR were provided in **Supplementary**
698 **Tables 2** and **3**, respectively. Growth curves of the WT and knockout strains were
699 obtained without antibiotics and with two low-level antibiotics (SMZ+TMP: 1 μg/ml,
700 TOB: 5 μg/ml). The biofilm formation potentials of WT and knockout strains were
701 determined based on the microtiter dish assay. In brief, strains were overnight
702 grown in LB medium at 150 rpm and 37°C, and then diluted to OD_{600} = 0.05. The
703 96-well plates were statically incubated at 37°C for 24 h to allow biofilm formation,
704 and the ultimate OD_{600} was determined. For staining, 200 μl of 0.1% crystal violet
705 (CV) was added to each well and incubated for 15 min at 37°C. The wells were
706 then washed three times with water and dried overnight. The stained biofilm was

707 disrupted by adding 200 μl of 95% ethanol solution (95% EtOH, 4.95% distilled
708 H_2O, 0.05% Triton X-100), and OD_{550} was measured afterward. Biofilm formation
709 potential was calculated as OD_{550}/OD_{600}.
710
711 **Machine learning methods**
712 To elucidate the relationships between the chemical structures of antibiotics
713 and the antibiotic resistance profiles they induced, quantitative structure-property
714 relationship (QSPR) models were constructed based on a published study[52]. In
715 brief, molecular fingerprints, including MACCS 166 and ECFPs and about 6,000
716 descriptors, were computed for the 96 stress-inducing antibiotics, which can be
717 fine-tuned with a set of features. The antibiotic resistance profiles of evolved
718 strains were used to train a number of ensemble regression models, including
719 linear regression (OLS), partial least squares (PLS), k-nearest neighbors (KNN),
720 and support vector machines (SVM). Automatic model generation (n = 10,000
721 iterations) by genetic algorithms was performed to find the best combination of
722 features. The performance of a model was evaluated by using statistical
723 parameters, i.e., in-sample coefficients of determination R^2, RMSE, and out-of-
724 sample random 10-fold cross-validation R^2. The visualization strategy for tracing
725 the atomic contribution was previously described[53,54]. To assess the applicability
726 domain of the best-performing models, the structural data of 1,563 antibiotics,
727 excluding the 96 used in experimental evolution, were downloaded from PubChem
728 (https://pubchem.ncbi.nlm.nih.gov/#query=antibiotic). A pairwise Tanimoto
729 similarity matrix was computed for the 1,659 antibiotics in total. An antibiotic was
730 considered as inside domain if its Tanimoto similarity with at least one of the 96
731 antibiotics exceeded 0.2. The best-performing models were employed to predict
732 antibiotic resistance development under stress conditions by the in-domain
733 antibiotics. 16 out of the 569 in-domain antibiotics were randomly selected to

evaluate model prediction performance, following the same procedures used in experimental evolution.

In addition to molecular fingerprint-based modeling, the relationships between 4,750 AACs identified in evolved strains and the resulting antibiotic resistance profiles were also explored using RF models. In brief, the Boole matrix was constructed based on the relative fold-change values and AACs for the 96 antibiotic stressors. The RF models employed 100 individual decision trees, each with a variable range of splits from 10 to 2,000 and a minimum split size of 5 per tree. Bootstrapping was employed to mitigate the inherent overfitting of models, where a 10-fold cross-validation (Shuffle Split) method was used for each machine-learning algorithm.

Prevalence of point mutations based on comparative genomics

Comparative genomic analysis was performed using a well-established pipeline as described previously[31]. Briefly, all *E. coli* genomes with L50 < 20 were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/pathogens/) on 3/24/2023, including 'clinical' (n = 3,376) and 'environmental' (n = 2,545) isolates. Genomes were searched for all key SNPs (n = 481) identified in this study. For each of the 5,921 genomes, Blast was used to identify the genomes hit for the QSHQ1 strain. If the BLAST e-value was 1e–10 or less, and the query genome had a coverage of at least 85%, the CDS sequences from hit clinical (n = 577) and environmental (n = 686) genomes were extracted, and multiple alignment was performed using Muscle. Mutated positions in all genomes in the database that aligned with the QSHQ1 SNPs were tabulated. Mutations that did not appear at least once in the entire pathogen database were removed from consideration; P values for over-representation in clinical (or environmental) isolates were calculated using Fisher's exact test for all remaining mutations.

Data availability

The raw sequence data of Reseq and Tn-Seq analyses are available in the NCBI Sequence Read Archive under the accession number PRJNA987173. All relevant data and models in this study are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

References

1. Programme, U. N. E. Frontiers 2017: Emerging issues of environmental concern. Preprint at (2017).

2. Larsson, D. G. J. & Flach, C.-F. Antibiotic resistance in the environment. *Nat. Rev. Microbiol.* **20**, 1–13 (2021).

3. Martínez, J. L. Antibiotics and antibiotic resistance genes in natural environments. *Science* **321**, 365–367 (2008).

4. Allen, H. K. *et al.* Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* **8**, 251–259 (2010).

5. FAO, UNEP, WHO & OIE), the W. O. for A. H. (WOAH founded as. *One Health Joint Plan of Action (2022–2026): Working together for the health of humans, animals, plants and the environment*. 86 https://www.who.int/publications/i/item/9789240059139 (2022).

6. Polianciuc, S. I., Gurzău, A. E., Kiss, B., Ştefan, M. G. & Loghin, F. Antibiotics in the environment: causes and consequences. *Med. Pharm. Rep.* **93**, 231–240 (2020).

7. Wistrand-Yuen, E. *et al.* Evolution of high-level resistance during low-level antibiotic exposure. *Nat. Commun.* **9**, 1599 (2018).

8. Liang, H., Zhang, J., Hu, J., Li, X. & Li, B. Fluoroquinolone residues in the environment rapidly induce heritable fluoroquinolone resistance in *Escherichia coli*. *Environ. Sci. Technol.* **57**, 4784–4795 (2023).

9. Xing, Y., Wu, S. & Men, Y. Exposure to environmental levels of pesticides stimulates and diversifies evolution in *Escherichia coli* toward higher antibiotic resistance. *Environ. Sci. Technol.* **54**, 8770–8778 (2020).

10. Wang, H., Feng, Y. & Lu, H. Low-level cefepime exposure induces high-level resistance in environmental bacteria: molecular mechanism and evolutionary dynamics. *Environ. Sci. Technol.* **56**, 15074–15083 (2022).

11. Jørgensen, K. M. *et al.* Sublethal ciprofloxacin treatment leads to rapid development of high-level ciprofloxacin resistance during long-term experimental evolution of *Pseudomonas aeruginosa*. *Antimicrob. Agents Ch.* **57**, 4215–21 (2013).

12. Jin, L. *et al.* Integrating environmental dimensions of "One Health" to combat antimicrobial resistance: essential research needs. *Environ. Sci. Technol.* (2022) doi:10.1021/acs.est.2c01651.

13. Theuretzbacher, U. *et al.* Critical analysis of antibacterial agents in clinical development. *Nat. Rev. Microbiol.* **18**, 286–298 (2020).

14. Fait, A. *et al.* Adaptive laboratory evolution and independent component analysis disentangle complex vancomycin adaptation trajectories. *Proc. National. Acad. Sci.* **119**, e2118262119 (2022).

15. Remigi, P., Masson-Boivin, C. & Rocha, E. P. C. Experimental evolution as a tool to investigate natural processes and molecular functions. *Trends Microbiol.* **27**, 623–634 (2019).

16. Maeda, T. *et al.* High-throughput laboratory evolution reveals evolutionary constraints in *Escherichia coli*. *Nat. Commun.* **11**, 5970 (2020).

17. Lukačišinová, M., Fernando, B. & Bollenbach, T. Highly parallel lab evolution reveals that epistasis can curb the evolution of antibiotic resistance. *Nat. Commun.* **11**, 3105 (2020).

18. Palomino, A. *et al.* Metabolic genes on conjugative plasmids are highly prevalent in *Escherichia coli* and can protect against antibiotic treatment. *ISME J.* **17**, 151–162 (2023).

19. D'Lima, L. *et al.* No decrease in susceptibility to NVC-422 in multiple-passage studies with methicillin-resistant *Staphylococcus aureus*, *S. aureus*, *Pseudomonas aeruginosa*, and *Escherichia coli*. *Antimicrob. Agents Chemother.* **56**, 2753–2755 (2012).

20. Pollard, J. E. *et al.* In vitro evaluation of the potential for resistance development to ceragenin CSA-13. *J. Antimicrob. Chemother.* **67**, 2665–2672 (2012).

21. CLSI. Performance standards for antimicrobial susceptibility testing. 33rd ed. Preprint at (2023).

22. Imamovic, L. *et al.* Drug-driven phenotypic convergence supports rational treatment strategies of chronic infections. *Cell* **172**, 121-134.e14 (2018).

23. Stracy, M. *et al.* Minimizing treatment-induced emergence of antibiotic resistance in bacterial infections. *Science* **375**, 889–894 (2022).

24. Alegado, R. A., Chin, C., Monack, D. M. & Tan, M. The two-component sensor kinase KdpD is required for *Salmonella typhimurium* colonization of *Caenorhabditis elegans* and survival in macrophages. *Cell. Microbiol.* **13**, 1618–1637 (2011).

25. Rezuchova, B., Miticka, H., Homerova, D., Roberts, M. & Kormanec, J. New members of the *Escherichia coli* σE regulon identified by a two-plasmid system. *FEMS Microbiol. Lett.* **225**, 1–7 (2003).

26. Groisman, E. A. & Mouslim, C. Sensing by bacterial regulatory systems in host and non-host environments. *Nat. Rev. Microbiol.* **4**, 705–709 (2006).

27. Khan, M. A., Mustafa, J. & Musarrat, J. Mechanism of DNA strand breakage induced by photosensitized tetracycline–Cu(II) complex. *Mutat. Res.Fundam. Mol. Mech. Mutagen.* **525**, 109–119 (2003).

28. Pushparajan, A. R., Ramachandran, R., Reji, J. G. & Kumar, R. A. *Mycobacterium tuberculosis* TetR family transcriptional regulator Rv1019 is a negative regulator of the mfd-mazG operon encoding DNA repair proteins. *FEBS Lett.* **594**, 2867–2880 (2020).

29. Doi, Y., Wachino, J. & Arakawa, Y. Aminoglycoside resistance the emergence of acquired 16S ribosomal RNA methyltransferases. *Infect. Dis. Clin. North Am.* **30**, 523–537 (2016).

30. Darby, E. M. *et al.* Molecular mechanisms of antibiotic resistance revisited. *Nat. Rev. Microbiol.* 1–16 (2022) doi:10.1038/s41579-022-00820-y.

31. Lopatkin, A. J. *et al.* Clinically relevant mutations in core metabolic genes confer antibiotic resistance. *Science* **371**, eaba0862 (2021).

32. Piskovsky, V. & Oliveira, N. M. Bacterial motility can govern the dynamics of antibiotic resistance evolution. *Nat. Commun.* **14**, 5584 (2023).

33. Seo, S. W., Kim, D., O'Brien, E. J., Szubin, R. & Palsson, B. O. Decoding genome-wide GadEWX-transcriptional regulatory networks reveals multifaceted cellular responses to acid stress in *Escherichia coli*. *Nat. Commun.* **6**, 7970 (2015).

34. Nüesch-Inderbinen, M., Treier, A., Stevens, M. J. A. & Stephan, R. Whole genome sequence-based characterisation of Shiga toxin-producing *Escherichia coli* isolated from game meat originating from several European countries. *Sci. Rep.* **13**, 3247 (2023).

35. Moussatova, A., Kandt, C., O'Mara, M. L. & Tieleman, D. P. ATP-binding cassette transporters in *Escherichia coli*. *Biochimica. Et. Biophysica. Acta. Bba. – Biomembr.* **1778**, 1757–1771 (2008).

36. Liu, D. Y. *et al.* Collateral sensitivity profiling in drug-resistant *Escherichia coli* identifies natural products suppressing cephalosporin resistance. *Nat. Commun.* **14**, 1976 (2023).

37. Imamovic, L. & Sommer, M. O. A. Use of collateral sensitivity networks to design drug cycling protocols that avoid resistance development. *Sci. Transl. Med.* **5**, 204ra132 (2013).

38. Wang, N., Gao, J., Xiao, S. & Zhuang, G. Overexpression of pdeR promotes biofilm formation of *Paracoccus denitrificans* by promoting ATP production and iron acquisition. *Front. Microbiol.* **13**, 966976 (2022).

39. Bevers, L. E., Hagedoorn, P.-L., Krijger, G. C. & Hagen, W. R. Tungsten transport protein A (WtpA) in *Pyrococcus furiosus* : the first member of a new class of tungstate and molybdate transporters. *J. Bacteriol.* **188**, 6498–6505 (2006).

40. Bult, C. J. *et al.* Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073 (1996).

41. Kuru, E. *et al.* Fluorescent D-amino-acids reveal bi-cellular cell wall modifications important for *Bdellovibrio bacteriovorus* predation. *Nat. Microbiol.* **2**, 1648–1657 (2017).

42. Voedts, H., Kennedy, S. P., Sezonov, G., Arthur, M. & Hugonnet, J.-E. Genome-wide identification of genes required for alternative peptidoglycan cross-linking in *Escherichia coli* revealed unexpected impacts of β-lactams. *Nat. Commun.* **13**, 7962 (2022).

43. Rodionova, I. A. *et al.* A systems approach discovers the role and characteristics of seven LysR type transcription factors in *Escherichia coli*. *Sci. Rep.* **12**, 7274 (2022).

44. Baryshnikova, A. Systematic functional annotation and visualization of biological networks. *Cell Syst.* **2**, 412–421 (2016).

45. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420–aaf1420 (2016).

46. Zhang, L., Liu, H., Dai, J., Xu, P. & Tang, H. Unveiling degradation mechanism of PAHs by a *Sphingobium* strain from a microbial consortium. *mLife* **1**, 287–302 (2022).

47. Wyres, K. L. & Holt, K. E. *Klebsiella pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr. Opin. Microbiol.* **45**, 131–139 (2018).

48. Pornsukarom, S., Vliet, A. H. M. van & Thakur, S. Whole genome sequencing analysis of multiple *Salmonella serovars* provides insights into phylogenetic relatedness, antimicrobial resistance, and virulence markers across humans, food animals and agriculture environmental sources. *BMC Genom.* **19**, 801 (2018).

49. Geng, J., Liu, X., Wang, J. & Li, S. Accumulation and risk assessment of antibiotics in edible plants grown in contaminated farmlands: A review. *Sci. Total Environ.* 158616 (2022) doi:10.1016/j.scitotenv.2022.158616.

50. DeJesus, M. A., Ambadipudi, C., Baker, R., Sassetti, C. & Ioerger, T. R. TRANSIT - a software tool for Himar1 TnSeq analysis. *Plos Comput. Biol.* **11**, e1004401 (2015).

51. Leshchiner, D. *et al.* A genome-wide atlas of antibiotic susceptibility targets and pathways to tolerance. *Nat. Commun.* **13**, 3165 (2022).

52. Mauri, A. & Bertola, M. Alvascience: A new software suite for the QSAR workflow applied to the blood–brain barrier permeability. *Int. J. Mol. Sci.* **23**, 12882 (2022).

53. Polishchuk, P. G., Kuz'min, V. E., Artemenko, A. G. & Muratov, E. N. Universal approach for structural interpretation of QSAR/QSPR models. *Mol. Inform.* **32**, 843–853 (2013).

54. Riniker, S. & Landrum, G. A. Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods. *J. Cheminformatics* **5**, 43 (2013).

Acknowledgements

This work was supported by the 'Major Program of National Natural Science Foundation of China (Grant No. 22193061 to Lizhong Zhu)', the 'Zhejiang Provincial Natural Science Foundation of China (Grant No. LR23B070002 to Huijie Lu)', the 'General Program of National Natural Science Foundation of China (Grant No. 22276165 to Huijie Lu)', the 'National Key Research and Development Program (Grant No. 2022YFC3704700 to Huijie Lu)', and ZJU Foundation of Institute of Intelligent Biological and Chemical Manufacturing. The authors thank Dr. Haoxuan Liu from Zhejiang University Centre for Evolutionary & Organismal Biology and Dr. Lei Huang from ZJU-Hangzhou Global Scientific and Technological Innovation Center for valuable discussions.

Author contributions

Conceptualization, H.W. and Huijie.L.; Software, H.W. and Hui.L.; Investigation, H.W. and Huijie.L.; Writing-original draft, H.W.; Writing-review & editing, H.W., Huijie.L., C.J. and L.Z.; Supervisions, Huijie.L. and L.Z. All authors approved the final paper.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and material requests should be addressed to Huijie Lu.