Pandemic *Vibrio parahaemolyticus* wave succession by gene loss

Chao Yang[1,#*], Hongling Qiu[2,#], Sarah L. Svensson[1#], Chengpei Ni[3,#], Song Gao[3], Zhizhou Jia[1], Huiqi Wen[1], Li Xie[1], Wenxuan Xu[1], Yujiao Qin[1], Shuzhu Lin[1], Jiancheng Wang[1], Yiquan Zhang[4], Yinghui Li[5], Min Jiang[5], Xiaolu Shi[5], Qinghua Hu[5], Zhemin Zhou[6], Yanjie Chao[1], Ruifu Yang[7], Yujun Cui[7], Jaime Martinez-Urtaza[8,*], Hui Wang[2,*], Daniel Falush[1,*]

[1] The Center for Microbes, Development and Health, CAS Key Laboratory of Molecular Virology and Immunology, Shanghai Institute of Immunity and Infection, Chinese Academy of Sciences, Shanghai, China;

[2] School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China;

[3] Wuxi Center for Disease Control and Prevention, Shenzhen, China;

[4] Department of Clinical Laboratory, Nantong Third People's Hospital, Affiliated Nantong Hospital 3 of Nantong University, Nantong, Jiangsu, China;

[5] Shenzhen Center for Disease Control and Prevention, Shenzhen, China;

[6] Key Laboratory of Alkene-Carbon Fibers-Based Technology & Application for Detection of Major Infectious Diseases, Cancer Institute, Suzhou Medical College, Soochow University, Suzhou, China;

[7] State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China;

[8] Department of Genetics and Microbiology, Facultat de Biociències, Universitat Autònoma de Barcelona, Barcelona, Spain

# These authors contributed equally to the article

* Corresponding authors:

C.Y.: cyang@siii.cas.cn; J.M.-U.: Jaime.Martinez.Urtaza@uab.cat;

H.W.: huiwang@shsmu.edu.cn; D.F.: daniel.falush@siii.cas.cn

Abstract

While the importance of spontaneous mutation and gene acquisition in the emergence and adaptive evolution of microbial pathogens is well recognized, the role of gene loss remains underexplored. Here, by large-scale phylogenomic analysis of 8,684 global isolates, we reconstructed the origin and evolutionary dynamics of the pandemic clone (PC) of *Vibrio parahaemolyticus*, the leading cause of seafood-associated infections worldwide. As has also been observed for the pandemic lineage of *V. cholerae*, there was an 'explosive' burst of gene and genomic island acquisition during the emergence of the clone followed by divergence into distinct, successive sublineages mediating independent waves of cross-country transmission. In *V. parahaemolyticus,* a recent global wave succession event has been driven by gene-loss in the putrescine utilization (Puu) pathway, which confers phenotypic advantages related to environmental adaptation (enhanced biofilms) and human transmission (increased cell adhesion and intestinal colonization, reduced virulence in an animal model and milder clinical symptoms in humans), aligning with the trade-off hypothesis of pathogen virulence evolution. We identify similar lineage-specific Puu-gene-loss across multiple bacterial genera, with effects on biofilm and adhesion replicated in *V. cholerae* and *E. coli*, suggesting convergent evolution and universal mechanisms across bacteria. Our findings provide a paradigm of adaptative evolution by gene loss in a pandemic pathogen.

Introduction

Microbial pathogens have recurrently emerged from environments or animals, with a few leading to disease epidemics or pandemics that pose notable public health challenges[1,2]. Understanding the genetic basis of pathogen emergence and adaptive evolution is critical for disease control and prevention, especially for pandemic pathogens. *Vibrio*, Gram-negative, halophilic bacteria that are ubiquitous in marine and aquatic habitats, are ideal models for investigating pathogen emergence and evolution. Among >100 *Vibrio* species, approximately 12 cause human infections[3]. Two species, *V. cholerae* and *V. parahaemolyticus*, are associated with disease pandemics, with the ongoing seventh cholera pandemic attributed to El Tor (7PET) isolates. *V. cholerae* 7PET was inferred to originate from a pre-pandemic strain that subsequently underwent explosive diversification - gaining multiple genetic elements within a short timeframe (~15 years), including the El Tor type cholera toxin prophage and *Vibrio* seventh pandemic islands (VSP-I and VSP-II)[4,5]. During the seventh pandemic, three successive sub-lineages of 7PET were identified, corresponding to the three independent but overlapping waves of global transmission[5,6], which may be driven by antibiotic resistance[7].

V. parahaemolyticus is the leading cause of seafood associated infections worldwide, with an increase in incidence over the last two decades. Its geographical range has also been expanding, possibly due to global warming[3,8,9]. The emergence of pandemic clone (PC), characterized by serotype O3:K6, substantially changed the *V. parahaemolyticus* epidemiology[10–12]. The O3:K6 PC strain was first isolated in 1995 from a traveler returning to Japan from Indonesia, and in 1996, it caused a disease outbreak in Calcutta, India[10]. Subsequently, it rapidly spread worldwide and came to dominate infections in many regions, along with rapid antigenic divergence (49 serotypes reported)[10–12]. Molecular and epidemiological surveillances has provided evidence that the PC may have originated from a serotype O3:K6 non-pathogenic

strain, after acquisition of multiple marker genes (e.g. virulence regulatory gene sequence *toxRS/new*[13], phage f237 and its unique open reading frame *orf8*[14]), type 3 or 6 secretion systems (T3SS or T6SS) and several *V. parahaemolyticus* genomic islands (VPaI-1 to VPaI-7)[15,16]. However, how these events happened, the driving force and mechanism underlying these events remain unclear. Currently, there is a lack of comprehensive genomic view on the origin and evolution of the PC.

Leveraging a large-scale, global genome dataset of *V. parahaemolyticus*, we comprehensively reconstructed the evolutionary history of the PC. Intriguingly, we identified several similarities between PCs of *V. parahaemolyticus* and *V. cholerae*. However, distinct from *V. cholerae*, we demonstrated that gene-loss in the putrescine utilization pathway has driven the recent global wave succession event by conferring phenotypic advantages, and these phenotypic effects are likely universal among bacteria.

Results

From pre-pandemic (pre-PC) to PC

We assembled a total of 8,684 high-quality (completeness >90% and contamination <5%) genomes of global *V. parahaemolyticus* isolates, including 333 newly sequenced genomes (**Supplementary Table 1**). To characterize population structure and identify PC isolates, we constructed a maximum-likelihood (ML) phylogenetic tree based on core-genome single-nucleotide polymorphisms (SNPs). Based on the ML tree and pairwise SNP-distance, 2,815 were assigned as PC isolates (mean SNP-distance: 437), 39 as pre-PC (close relatives of PC, mean SNP-distance to PC: 15,463), with the remaining being non-PC isolates (mean SNP-distance to PC: 55,554, **Fig. 1a**).

Pre-PC isolates were further split into four sub-groups, designated pre-PC1 to pre-PC4, respectively (**Fig. 1b,c**). Pre-PC1 to pre-PC3 were exclusively identified in Asia, with pre-PC1 and pre-PC2 being clinical isolates, while pre-PC3 were mainly environmental strains. The closest relative of the PC, pre-PC4, was represented mainly (13/16) by environmental isolates from Asia (Japan and China) and the US. The serotypes of most pre-PC isolates (33/39) were predicted to be O3:K6, and sequence types (STs) of pre-PC4b (a subset of pre-PC4) isolates were ST3 except one strain (**Fig. 1c** and **Supplementary Table 2**).

We identified a significant temporal signal among pre-PC3b to PC isolates and inferred the dated phylogeny (**Fig. 1b,2b**). The most recent common ancestor (MRCA) of the PC was estimated at 1983 (95% confidence interval [CI]: 1979-1985), 12 years preceding its first discovery in 1995. The strains are expected to be highly homogeneous at the point of their origin, but early PC strains from 1996 already showed some diversity (an average of 38 non-recombined SNPs), consistent with the long time-interval between origination and first discovery (**Fig. 2b**).

We investigated the distribution of known PC marker genes/islands and inferred the evolutionary history from pre-PC to PC (**Fig. 1b,c** and **Extended Data Fig. 1)**. The PC likely originated from a serotype O3:K6 pathogenic strain encoding T3SS1, T6SS2, phage f237 and

VPaI-3, and subsequently acquired T6SS1 and the *toxRS/new* sequence. Interestingly, we identified 'explosive' evolution defined as gain of multiple genes/islands within a short timeframe, as seen during the origin of *V. cholerae* 7PET. Between 1920 (MRCA of pre-PC4b and PC) and 1983 (MRCA-PC), MRCA-PC gained multiple dispersedly distributed genetic elements, including the well-known virulence factor thermostable direct hemolysin gene *tdh, orf8*, and six out of the seven VPaIs, along with 86 mutations and 80 homologous recombination events. Moreover, *trh* (*tdh*-related hemolysin), VPaI-1 and VPaI-2 were independently gained or lost by pre-PC strains.

We further investigated the source of these marker genes/islands by searching against non-PC genomes (**Extended Data Fig. 2**). All of them were identified in non-PC isolates with varying frequencies, indicating their sources are likely *V. parahaemolyticus*, rather than other *Vibrio* or bacteria. Moreover, the PC appeared to gain phage f237, VPaI-1 and VPaI-7 from non-PC strains, despite being encoded by pre-PC strains. Apart from marker genes/islands, six PC-specific genes were identified, three of which were within the superintegron (**Supplementary Table 3**).

Fine-scale population structure of PC

To reduce the effect of repeat sampling (strains from disease outbreaks) and exclude outlier strains, we generated a representative subset of 942 PC isolates for further analysis (see methods), spanning 20 countries across five continents between 1996 and 2022 (**Extended Data Fig. 3**). Among the representative isolates, we identified 26,377 core-genome SNPs, with 66% attributed to homologous recombination. Two recombination hotspots were identified, including the known O:K antigen region[17] and a novel hotspot (VP1883-1891) encoding a type II toxin-antitoxin system and cold shock protein. There were 281 homoplastic non-recombined SNPs, of which 26 were convergently identified 6-18 times in 16 genes associated with quorum sensing, environmental response and virulence (**Extended Data Fig. 4** and **Supplementary Table 4**).

We constructed an ML tree and dated phylogeny for PC representative strains (**Fig. 2a,b**). By integration with hierarchical clustering (**Extended Data Fig. 5**), we split the PC into four sub-lineages corresponding to independent waves of cross-country transmission (**Fig. 2a-d**), which was also observed in the *V. cholerae* 7PET[5]. Following the nomenclature of *V. cholerae*, we designated these sub-lineages as wave-1 to 4, respectively. The estimated MRCA times for wave-1 to wave-3 were close (1984-1987), while wave-4 emerged more recently in 1998.

We identified 11 serotypes and three STs in representative strains (**Fig. 2b**). While O3:K6 was dominant (79%), serotype conversions were identified, particularly in wave-2 and wave-3. ST3 (96%) remained the predominant ST. Several marker genes/islands (VPaI-1, 4, 5 and *orf8*) were lost in 2%-8% of representative isolates, with the others being conserved (**Extended Data Fig. 6a**). Notably, the genomic island VPaI-7 encoding *tdh* and T3SS2 showed high differentiation between different waves (**Extended Data Fig. 6b**). Additionally, we observed a conserved antimicrobial resistance (AMR) gene profile (**Extended Data Fig. 6a**). Apart from intrinsic ampicillin resistance-related genes, no recently acquired gene was identified in >1% of

representative isolates.

Wave succession and global transmission

Remarkably, we identified a 'wave succession' pattern: wave-1 to wave-4 succeeded each other temporally as the major wave in most countries, except for two South American countries, Peru and Chile, where wave-1 has dominated for over a decade (**Fig. 2c,d**). Wave succession was particularly evident in China, where sampling is sufficient. Similar trends are seen in other countries with ≥10 representative strains, including the US, UK, Canada and Bangladesh, suggesting a global pattern.

Wave-4 frequency has been increasing since its emergence, ultimately becoming predominant from 2012 onward. Since its origin, the effective population size (*Ne*) of the PC increased exponentially until 1998 (**Fig. 2e**). Subsequently, *Ne* fluctuated and showed a decreasing trend after 2010. Population dynamics of waves was consistent with temporal dynamics, with the *Ne* peaking alternatively for different waves. Post-2010, the *Ne* of wave-4 exceeded others and remained stable, while *Ne* of other waves declined rapidly.

We inferred the geographical transmission history of the PC by phylogeographical analysis, focusing on cross-continent events. A total of 67 transmission events were identified, mostly transient and not leading to local colonization (**Extended Data Fig. 7**). Only six 'TC' events (Transmissions resulting in local Colonization that persisted for ≥3 years) were identified (**Fig. 2b** and **Extended Data Fig. 7**). Wave-1 in South America likely sourced from a single TC event from Asia to Peru in 1990 (TC1), much earlier than its first discovery in South America (1997). Subsequently, wave-1 spread from Peru to Chile and backward transmissions were also identified (TC2-4). Additionally, two TC events from Asia to North America were identified (TC5-6) in other waves.

Effects of Puu-gene loss

Puu-gene loss is associated with wave-4 replacement

To investigate the genetic basis of the wave-4 replacement of other waves, we comprehensively identified wave-4-specific variations and examined their functions. A total of 40 variations were identified (**Fig. 3a** and **Extended Data Table 1**), including 29 SNPs (located in 26 genes and 3 intergenic regions), two small indels (located in 2 genes) and most notably, the absence of nine adjacent genes (VP1775-VP1786). The fixation of gene-loss is expected to be adaptative and selection-driven in free-living species such as *V. parahaemolyticus,* which have large effective population sizes. Of the nine genes, five are annotated as part of the putrescine utilization (Puu, VP1775-1781; *puuBCRDA*) pathway, one encodes a Na^+/H^+ antiporter (VP1782), and the other encode hypothetical proteins or DUF3291 domain-containing proteins (VP1783-1786). We focused on the Puu-pathway genes, which degrade the polyamine putrescine into 4-aminobutanoate (GABA) or spermidine (**Fig. 3b**) and have been reported to be involved in putrescine catabolism in *Escherichia coli*[18] and biofilm formation in *Pseudomonas aeruginosa*[19].

Growth and biofilm formation

The *puuC* gene is involved in three sub-pathways in the conversion of putrescine to GABA, and its deletion is expected to completely block the conversion, mimicking the effect of a complete deletion of Puu-genes **(Fig. 3b)**. Therefore, we generated a *puuC* deletion mutant in the wildtype strain RIMD2210633 (wave-3, hereafter RIMD-W3) to explore the function and consequences of loss of Puu-genes in *V. parahaemolyticus*.

As reported in *E. coli*[18], Δ*puuC* exhibited reduced growth when putrescine served as the sole carbon or nitrogen source compared with the wildtype parental strain, while complementation of the deletion strains *in trans* (C-Δ*puuC*) restored growth (**Fig. 3c**). Moreover, Δ*puuC* showed significantly ($p<0.01$) enhanced biofilm formation under rich (alkaline peptone water, APW) or low-nutrient (M9 minimal) media with high (3%) or moderate (1%) salt concentrations after 48 hours of culture (**Fig. 3d** and **Extended Data Fig. 8**).

We further investigated the effect of Puu-gene-loss on extracellular and intracellular metabolites by targeted metabolomics using a liquid chromatograph mass spectrometer (LC-MS). We used a low-nutrient and moderate-salt condition (M9 + 1% NaCl), where the most significant biofilm differences were also observed, to minimize the interference from substances in complex media. The most notable differences were observed during stationary phase, where Δ*puuC* exhibited significant ($p<0.01$) extracellular putrescine accumulation and GABA reduction compared to the wildtype strain. Although similar patterns were also detected for both metabolites intracellularly and during logarithmic phase, the magnitude of the difference and correlation were less pronounced than for extracellular levels in stationary phase (**Fig. 3e**).

We next performed transcriptomic sequencing in stationary phase to identify differentially expressed genes (DEGs) that may underlie the Δ*puuC* mutant phenotypes. We found that genes encoding the SapBCDF (VP1167-VP1170) transporter, which transport putrescine from the intracellular to extracellular environment[20], as well as the extracellular putrescine sensor and signal transduction gene *nspS*[21–23] (VP1526) were significantly up-regulated in Δ*puuC* compared to WT, consistent with its accumulation of extracellular putrescine (**Fig. 3f**, **Extended Data Fig. 9** and **Supplementary Table 5**). The up-regulation of the putrescine transporter SapBCDF suggests suggests active extracellular export of putrescine in the Δ*puuC* strain, likely explaining the unchanged intracellular putrescine levels during the stationary phase compared to the wildtype (**Fig. 3e**).

In line with increased biofilm formation of Δ*puuC*, we identified multiple DEGs between wildtype and Δ*puuC* associated with biofilm or biofilm-related pathways, including c-di-GMP, quorum sensing, CupE fimbriae and Flp pili. We quantified intracellular c-di-GMP levels in wildtype and Δ*puuC* during growth by introducing dual-fluorescence reporter plasmids into wildtype and mutant strains. Consistent with differential expression of several c-di-GMP-related enzymes and higher biofilm formation, c-di-GMP levels were significantly higher in Δ*puuC* compared to wildtype in stationary phase (**Fig. 3g**).

Putrescine has been shown to act as an intraspecies cell-cell communication signal, regulating phenotypes such as biofilm formation and virulence[19,24–26]. To investigate whether extracellular

putrescine drives phenotypic and gene expression changes we observed in the Δ*puuC* mutant, we tested the effects of exogenous putrescine. Upon putrescine addition, both wildtype and Δ*puuC* showed significantly increased biofilm and c-di-GMP levels (**Fig. 3d,g**) and significant gene expression changes (**Extended Data Fig. 9**), suggesting putrescine sensing might play a regulatory role in *V. parahaemolyticus*, as has been described previously[19,24–26]. For Δ*puuC*, exogenous putrescine significantly up-regulated the sensor gene *nspS,* the putrescine importer gene *potC,* and multiple biofilm-related DEGs that we found differentially expressed in Δ*puuC vs.* RIMD-W3 (**Fig. 3f, Extended Data Fig. 9**), suggesting active sensing and regulatory roles for putrescine. However, the impact of exogenous putrescine on gene expression in wildtype was less pronounced than in Δ*puuC,* possibly due to efficient putrescine turnover, as indicated by significant upregulation of Puu-genes upon addition of putrescine (**Extended Data Fig. 9**). Regardless of the degree of impact, several biofilm-related DEGs (Δ*puuC vs.* RIMD-W3), including *opaR* (a known regulator of c-di-GMP and biofilm formation[27]), *cusA* (CupE fimbriae) and VP2422 (Flp pili), were significantly up-regulated in both wildtype and Δ*puuC* strains when we added exogenous putrescine. Conversely, *ompU* (outer member protein, biofilm pathway) and *toxB1* (GTP cyclohydrolase, quorum sensing pathway) were down-regulated (**Extended Data Fig. 9**). Together, these findings suggest that increased extracellular putrescine caused by Puu-gene-loss may act as a cell-cell communication signal that enhances biofilms by regulating biofilm-related genes and promoting c-di-GMP synthesis in Δ*puuC*.

Human-related phenotypes

The most significant category of DEGs (Δ*puuC vs.* RIMD-W3) were those encoding CupE-fimbriae and Flp-pili (**Fig. 3f**). Adding exogenous putrescine further significantly up-regulated all CupE-fimbriae genes and four Flp-pili genes in Δ*puuC*, and also up-regulated *cusA* and VP2422 in wildtype (**Extended Data Fig. 9**). In addition to their role in biofilms, fimbriae and pili are generally related to cell adhesion for pathogens[28]. Motivated by these observations, we performed Caco-2 cell adhesion assays and found significantly enhanced adhesion for Δ*puuC* compared to wildtype (**Fig. 3h**). We also tested whether increased cell adhesion might transform into enhanced intestinal colonization. We constructed bioluminescent wildtype and Δ*puuC* strains and traced their colonization of mice after oral gavage using *in vivo* imaging. At 5 hours post-infection, mice infected with Δ*puuC* exhibited stronger luminescence compared to wildtype (**Fig. 3i**). Bacteria were then rapidly cleared, with no detectable luminescence observed for both wildtype and mutant strains by 10 hours post-infection.

An unexpected observation was the downregulation of the key virulence gene, *tdh,* in Δ*puuC* (2.7-fold, *p*-adj=0.02, **Fig. 3f**), suggesting that altered putrescine metabolism might reduce virulence. Accordingly, we compared the lethality of wildtype and Δ*puuC* mutant strains in mice following intraperitoneal infection. Consistent with *tdh* downregulation, Δ*puuC* exhibited reduced lethality (**Fig. 3j**). These findings together revealed the potential impact of Puu-gene-loss on phenotypes related to human infection, including cell adhesion, intestinal colonization and virulence.

Phenotypes of wave-4 strains

To link the phenotypes related to Puu-gene-loss observed in a Δ*puuC* wave-3 strain with the

behavior of natural wave-4 strains, we next compared 12 natural strains, including six wave-1 to wave-3 (2 strains each) and six wave-4 strains, for the same phenotypes as described above (**Fig. 4a**). Consistent with the above phenotypic differences between wildtype and mutant strains established above, the wave-4 strains exhibited decreased growth when putrescine was provided as the sole carbon or nitrogen source, as well as enhanced biofilms and cell adhesion, compared with wave-1 to wave-3 strains (**Fig. 4a-c**). Moreover, we complemented a wave-4 strain (VPCZ58) with the nine genes (VP1775-VP1786) that are absent in wave-4 strains and identified significantly reduced biofilm formation and adhesion compared to the wildtype strain (**Fig. 4b,c**), providing further evidence that the Puu-genes are responsible for the observed phenotypic differences. Additionally, we analyzed the clinical symptom data of 1,966 PC-related patients reported in our previous study[29], and found that wave-4 was associated with milder symptoms, including significantly lower rates of fever, vomiting and abdominal pain (**Fig. 4d**), consistent with reduced virulence observed in our mouse infection experiment (**Fig. 3j**).

We hypothesized that these phenotypic differences mediated by Puu-gene-loss might confer competitive advantages of wave-4 strains in the environment. Therefore, we performed competition experiments for natural strains in seawater with chitin as a nutrition source, mimicking a possible natural environment. We mixed wave-4 strains with those from other waves in equal (1:1) or different (1:5) ratios. By targeted PCR amplification of a gene with a wave-4-specific SNP allele (VPA0427 A919G), we quantified the relative abundance of wave-4 based on the frequency of the wave-4-specific allele. After 72 hours of competition, we measured a significant increase in the frequency of wave-4-specific alleles for both mixing ratios (**Fig. 4e**), demonstrating the competitive advantage of wave-4 under the tested condition.

Puu-gene-loss in bacteria

Puu-gene-loss could be attributed to selection or genetic drift, and convergent Puu-gene-loss and/or its fixation in specific lineages would provide evidence supporting the selection hypothesis. We analyzed the prevalence of Puu-genes among 108,526 genomes from 36 bacterial species, including *Vibrio*, common pathogens, and probiotics. Puu-genes were generally not conserved among bacteria, with part of the pathway lost in some species (**Fig. 5** and **Extended Data Table 2**). Notably, the absence/pseudogenization of Puu-genes in a subset of strains within species was convergently identified across multiple species (**Fig. 5b** and **Extended Data Fig. 10**). Furthermore, this absence/pseudogenization was fixed or nearly fixed in a specific lineage for species such as *E. coli*, *Shigella boydii, S. sonnei* and *P. aeruginosa*, or convergently occurred in two or more lineages among species such as *Campylobacter coli, Yersinia pestis*, *Bifidobacterium longum* and *Lactobacillus helveticus*. Additionally, Puu-gene-loss was not only identified among *V. parahaemolyticus* PC strains, but also identified in the *V. cholerae* 7PET **(Fig. 5b)**.

To test if Puu-gene-loss has a similar phenotypic effect in other bacteria, we generated *puuC* deletion strains in *V. cholerae* (E7946 background) and *E. coli* (MG1655 background) and tested their biofilm formation and cell adhesion abilities. Consistent with the results in *V. parahaemolyticus,* the Δ*puuC* mutant strains in both species exhibited significantly enhanced

338 biofilms and adhesion compared to their parental wildtype strains, and adding exogenous
339 putrescine further enhanced biofilms (**Fig. 5b**). Together, these results indicate that the Puu-
340 gene-loss is under selection, and the associated phenotypic changes potentially universal across
341 bacterial species.
342
343 **Discussion**
344 Leveraging a large-scale genome dataset of global isolates, we present the most comprehensive
345 genomic view to date on the origin and evolution of the *V. parahaemolyticus* PC. Our findings
346 suggest that the PC likely originated from a serotype O3:K6 pathogenic strain through explosive
347 gain of multiple genes/islands from *V. parahaemolyticus*, rather than from non-pathogenic
348 strains or from other *Vibrio* species[11]. Notably, several characteristics previously considered to
349 be specific to the PC[10–12], including ST3, *toxRS*/new, and VPaI-1 to VPaI-3, were also present
350 in pre-PC strains, while the *orf8* gene remained unique to the PC, thus remaining as a reliable
351 marker. Moreover, we described the fine-scale population structure and spatiotemporal
352 dynamics of the PC, and identified 'wave succession' with potential underlying molecular
353 mechanisms. Together, our findings substantially enhance the current understanding of how
354 epidemic lineages of *V. parahaemolyticus* emerge and evolve.
355
356 During the evolution of the *V. parahaemolyticus* PC, we observed intriguing parallels with the
357 *V. cholerae* 7PET strains and other pathogens. First, we observed explosive gain of
358 genes/islands, including key virulence genes (*tdh* and *ctxAB*) and marker genomic islands (six
359 VPaIs and two VSPs) for the *V. parahaemolyticus* PC, as has been reported for the *V. cholerae*
360 7PET[4,5]. This phenomenon has also been observed during human colonization by other
361 pathogens[30] and in the evolution of metabolic pathways across prokaryotic lineages[31]. Frequent
362 human-mediated cross-regional transmission and adaptation to new environments are
363 considered potential drivers for the *V. cholerae* 7PET[4], which might also apply to the *V.*
364 *parahaemolyticus* PC. Second, for both PCs from pandemic *Vibrio* species, the inferred times
365 of origin (1983 and 1955[4]) and introduction to specific geographical regions (e.g. South
366 America: 1990 and 1989[32]) were much earlier than their first epidemiological observations
367 (1995 and 1961; South America: 1997 and 1991). Two possible explanations for this include
368 variability in the molecular clock rate[33] with higher early mutation rates[34,35], and the possibility
369 of early cryptic disease epidemics, with specific events such as El Niño or warming in coastal
370 conditions triggering a sudden surge in cases and thereby drawing epidemiological attention.
371 While the second explanation does not apply to *V. cholerae* due to the high risk and strict
372 surveillance of cholera, both explanations are applicable to *V. parahaemolyticus*. Third,
373 divergence into successive sublineages mediating waves of infections, identified in both
374 pandemic *Vibrio* and other pathogens such as SARS-CoV-2[36]. While antimicrobial resistance
375 likely drove the wave succession of *V. cholerae* 7PET[7], *V. parahaemolyticus* exhibits low
376 resistance rates to first-line antibiotics, indicating a different driving factor.
377
378 We propose a model of wave succession driven by gene-loss in the Puu-pathway, conferring
379 phenotypic advantages related to environmental adaptation and human-mediated transmission
380 (**Fig. 6**). Puu-gene-loss leads to the intracellular putrescine accumulation, which is subsequently
381 transported extracellularly by the SapBCDF transporter[20]. Extracellular putrescine acts as a

signaling molecule[19,24–26], sensed and transducted to downstram regulatory genes by NspS as reported in *V. cholerae*[21–23], impacting multiple phenotypes including biofilm formation, adhesion, and virulence though the up- or down-regulation of related genes. However, the detailed molecular mechanism underlying these phenotypic changes remains unclear and warrants further studies. Additionally, it is noteworthy that, aside from Puu-gene loss, wave-4 strains carry other unique variations (29 SNPs and 2 indels). These variations may influence additional phenotypes or potentially complement for the effects of Puu-gene loss, and they could also contribute to wave succession.

The importance of biofilm formation in bacterial environmental adaptation and survival under adverse conditions (e.g. antibiotic exposure) is well recognized[37]. Therefore, the enhanced biofilms mediated by Puu-gene-loss may improve the environmental adaptability of wave-4 strains, as suggested by our competition experiment results. Moreover, the ancestral strains of the PC, as well as most non-PC strains, carry the complete Puu-pathway genes, indicating that the loss of Puu-genes occurred recently, possibly in response to environmental changes. Given that elevated putrescine levels have been reported to provide advantages in the acid stress response[38,39], the loss of Puu-genes might have been selected for by recent increases in ocean or coastal acidification.

Human activities are pivotal in pathogen transmission. Increased cell adhesion and intestinal colonization could directly facilitate human-mediated transmission, while reduced virulence, associated with milder clinical symptoms, may lower the likelihood of seeking medical attention, thereby promoting cryptic transmission. Notably, increased adhesion and colonization are not contradictory to reduced virulence, as different infection routes in the experiments (gavage *vs.* intraperitoneal injection) and their underlying mechanisms are known to differ[40].

The effect of the Puu-pathway and putrescine on biofilms is likely a broader phenomenon among bacteria than previously recognized. We have demonstrated that Puu-gene-loss or exogenous putrescine enhances biofilms in three different species. Furthermore, Puu-gene-loss in *P. aeruginosa* also leads to enhanced biofilms[19], and *E. coli* phylogroup B2, characterized by absence of *puuA-D* (**Fig. 5b**) exhibits a stronger biofilm-formation ability than other phylogroups[41,42], providing further evidence for a more universal effect. Additionally, we propose that the increased extracellular putrescine caused by Puu-gene-loss enhances biofilms by upregulating biofilm and biofilm-related genes, as well as promoting c-di-GMP synthesis. This effect of putrescine on c-di-GMP synthesis was also reported in *P. aeruginosa*[19], suggesting a universal phenomenon.

For the first time, we have established links between Puu-gene-loss, cell adhesion and intestinal colonization, which are also likely universal across bacteria. We demonstrated that Puu-gene-loss enhances cell adhesion in three different species and promotes intestinal colonization in *V. parahaemolyticus.* Furthermore, *E. coli* phylogroup B2 strains are often associated with urinary tract infections[43] and show increased persistence in the intestines of human infants[44], both of which is consistent with the hypothesis of increased adhesion and colonization mediated by

426 Puu-gene-loss. Interestingly, the probiotic *E. coli* strain Nissle 1917, known for its strong
427 capability of human gut colonization[45], belongs to phylogroup B2 and lacks Puu-genes (**Fig.**
428 **5b**).
429
430 The convergent loss or pseudogenization of Puu-gene within species (e.g., *C. coli*) and across
431 different bacterial species, along with its lineage-specific fixation (e.g., in *E. coli*), indicate that
432 these events are more likely selection-driven and confer fitness advantages. Puu-gene
433 loss/pseudogenization was uniquely fixed in major pathogenic clonal complexes (CC) of *S.*
434 *boydii* (CC145) and *S. sonnei* (CC152)[46], and recently fixed in *V. parahaemolyticus* wave-4,
435 suggesting potential fitness advantages for human infection or transmission. However, this
436 appears to not always be the case, as Puu-gene loss/pseudogenization has not occurred in major
437 pathogenic lineages of *C. coli* and *P. aeruginosa.* Additionally, Puu-gene loss/pseudogenization
438 appears to be associated with lineage divergence in *E. coli* and *C. coli*, and Puu-gene gain is
439 potentially linked to human adaptation in the probiotic *B. longum,* with Puu-genes
440 predominantly identified in adult human strains and absent in infant or non-human strains.
441 These observations together indicate that while Puu-gene loss/pseudogenization is likely
442 selection-driven, the underlying driving forces vary across different species.
443
444 While the significance of spontaneous mutation and gene acquisition in pathogen evolution is
445 well recognized[47–49], the role of gene loss remains underexplored. Gene loss is pervasive across
446 all life kingdoms[50,51]. In bacteria, it is commonly observed in symbionts and host-dependent
447 pathogens, where is generally considered neutral and driven by genetic drift associated with
448 transmission bottleneck[52,53], while loss of several specific genes has been experimentally
449 validated to be associated with pathogen virulence[54,55]. Free-living bacteria are better models
450 for investigating gene-loss-driven adaptative evolution due to their large effective population
451 sizes, with gene-loss fixation being expected to be adaptative and selection-driven. Some free-
452 living bacteria, such as the abundant open-ocean bacterioplankton *Prochlorococcus*, exhibit
453 extensive gene loss, which has been attributed to accelerated evolution and fitness advantages
454 conferred by genomic and metabolic "streamlining"[56,57]. Moreover, previous studies of gene
455 loss-mediated bacterial adaptation, whether for symbionts, host-dependent pathogens, or free-
456 living species, generally come with the cost of transition from generalists capable of surviving
457 in diverse environments to specialists adapted to specific ecological niches[54–57]. The exact niche
458 of the *V. parahaemolyticus* PC remains unclear but it is more likely a generalist due to
459 widespread distribution in marine and aquatic habitats. If we consider different geographical
460 regions as the macro-niches, the wave-4 strains characterized by Puu-gene-loss have spread
461 globally, without clear evidence supporting its transition from a generalist to specialist strategy.
462 Additionally, to the best of our knowledge, there are no reports of gene loss-mediated adaptation
463 in free-living pathogens. Therefore, our study presents a novel paradigm of pathogen adaptation
464 by gene loss.
465
466 In summary, our study provides the most comprehensive genomic view on the origin and
467 evolution of the *V. parahaemolyticus* PC. We identified several interesting common features
468 between the evolution of pandemic clones in two *Vibrio* species, but the driving forces
469 underlying wave succession events appear distinct. While spontaneous mutation and gene gain-

470 mediated antibiotic resistance may drive the *V. cholerae* 7PET succession[7], we demonstrated
471 that Puu-gene loss, which confers phenotypic advantages related to environmental adaptation
472 and human transmission, has likely driven the recent *V. parahaemolyticus* wave succession. An
473 interesting finding is that *V. parahaemolyticus* wave succession has come along with reduced
474 virulence, which aligns with the trade-off hypothesis of pathogen virulence evolution[58].
475 However, every coin has two sides. Puu-gene-loss could also bring disadvantages, such as
476 decreased growth when putrescine serves as the sole nutrient source, implying the phenotypic
477 advantages should be environment-dependent and might be transient. We also found that Puu-
478 gene-loss is general and frequent among bacteria, selection-driven, with its phenotypic impacts
479 and underly mechanisms likely universal across species. Furthermore, Puu-gene-loss may play
480 key roles in bacterial evolution, including lineage divergence (e.g. *E. coli, C. coli* and *B.*
481 *longum*), human adaptation (e.g. *B. longum*) and the epidemic expansion of specific pathogenic
482 lineages (e.g. *V. parahaemolyticus, S. boydii* and *S. sonnei*). Taken together, our findings
483 provide a paradigm of adaptative evolution by gene loss in pandemic pathogens.
484
485 **Methods**
486 ***V. parahaemolyticus* genome dataset**
487 A total of 8,684 high-quality genomes were included in this study, including 333 newly
488 sequenced and 8,351 publicly available genomes (**Supplementary Table 1**). Public genomes
489 (up to November 2022) were downloaded from National Center for Biotechnology Information
490 GenBank or Sequence Read Archive (SRA) database, with metadata extracted from
491 corresponding BioSample information or the literature. New genomes were obtained from
492 archived strains isolated from multiple countries, including China, Japan, Senegal, Peru and
493 Chile.
494
495 Whole genome sequencing of new strains was performed using Illumina NovaSeq or Miseq
496 platforms. New and public sequencing reads were trimmed using Trimmomatic v0.39[59] and
497 were *de novo* assembled using shovill v1.1.0 pipeline (https://github.com/tseemann/shovill).
498 The quality assessment of genomes was performed using CheckM v1.1.3[60]. Whole genome
499 average nucleotide identity (ANI) against the reference genome of *V. parahaemolyticus*
500 (RIMD2210633, accession numbers: NC_004603.1 and NC_004605.1) was calculated using
501 fastANI v1.1[61]. Only high-quality genomes (completeness >90% or contamination <5%) with
502 ANI >95% were included in further analysis.
503
504 **Phylogenetic and phylogeographical analysis**
505 Genome sequences were aligned against the reference genome using MUMmer v3.23[62] to
506 generate whole-genome alignments. Core-genome (regions present in >99% strains) single-
507 nucleotide polymorphisms (SNPs) were identified using SNP-sites v2.5.1[63]. SNPs located in
508 repetitive regions identified by TRF v4.07b[64] or BLASTN self-search were excluded. Non-
509 repetitive SNPs were used to construct maximum-likelihood phylogenetic trees using FastTree
510 v2.1.10[65] or RAxML-NG v1.0.1[66]. Recombination regions were detected using Gubbins
511 v3.1.3[67]. Homoplastic SNPs were identified using SNPPar v1.0[68]. Recombined and homoplastic
512 SNPs were excluded from the phylogenetic analysis of pandemic clone (PC) and pre-PC
513 isolates. Phylogenetic trees were visualized using ggtree[69]. Hierarchical clustering analysis

514 were performed using fastBAPS v1.0.6[70] to assign strains into subgroups. The transitions in the
515 isolation countries of strains in the phylogenetic tree are considered cross-country events, which
516 are detailed in Extended Data Fig. 7. TC was defined as Transmissions resulting in local
517 Colonization (persisted for ≥3 years).
518
519 **Serotype and MLST typing**
520 *In silico* O- and K-antigen genotyping was performed using *V. parahaemolyticus* Kaptive
521 database[71]. Multi-locus sequence typing (MLST) was performed using mlst
522 (https://github.com/tseemann/mlst) to scan genome sequences against the PubMLST databases.
523
524 **Representative genome dataset of PC**
525 We generated a representative subset of 942 PC isolates. First, to reduce the effect of repeat
526 sampling such as strains from disease outbreaks, we selected the genome with maximum N50
527 from a cluster of strains with ≤6 SNPs from same country and within a year, representing
528 putative outbreaks[29]. Second, we calculated root-to-tip distances of phylogenetic trees based on
529 non-recombined SNPs, and removed putative hypermutators with extreme root-to-tip distances
530 and branch lengths (**Extended Data Fig. 3**).
531
532 **Temporal analysis and population size dynamics**
533 Temporal signals were investigated using TempEst v1.5.3[72] by calculating the correlation
534 between root-to-tip distances and isolation dates (years). For strains with significant temporal
535 signals ($p<0.05$), we re-constructed phylogenetic trees and applied BactDating v1.1.0[73] to infer
536 the dated phylogenetic trees. Effective population size dynamics was inferred using skygrowth
537 v0.3.1[74] based on dated trees.
538
539 **Core-pangenome**
540 All PC representative genomes were re-annotated using Prokka v1.13[75], and the annotation
541 outputs (GFF3 files) were used in Panaroo v1.2.8[76] to identify the core-pangenome and gene
542 presence/absence.
543
544 **Wave-4 specific variations**
545 We analyzed three types of variations, including SNPs, gene presence/absence and unitigs. The
546 SNPs and gene presence/absence data were extracted from the preceding analysis. The unitig
547 matrix was determined using unitig-caller v1.2.1 (https://github.com/bacpop/unitig-caller).
548 Unitig sequences were aligned against assembled genomes using BLASTN to identify
549 corresponding genomic positions and related genes. Wave-4 specific variations were defined as
550 alleles, genes or unitigs present/absent in >95% of wave-4 strains and <5% of non-wave-4
551 strains.
552
553 ***puuC* gene in the reference genome**
554 The *puuC* gene (VP_RS08540/VP1776-7) in the reference genome RIMD221063 was
555 annotated as a pseudogene due to a frameshift caused by insertion of one nucleotide. However,
556 we performed both Sanger sequencing and Illumina re-sequencing of this region in our
557 RIMD2210633 strain. Both sequencing methods confirmed absence of this nucleotide and an

558 intact *puuC* ORF in our strain, with no frameshift detected. Therefore, we hypothesize that the
559 frameshift reported in the reference genome is likely caused by a sequencing error.
560
561 **Puu-genes in bacteria**
562 We investigated the prevalence of Puu-genes in 36 bacterial species, including *Vibrio*, common
563 pathogens, and probiotics. We included a total of 108,526 genomes (62-24,205 genomes per
564 species, **Supplementary Table 6**) downloaded from Genome Taxonomy Database (GTDB,
565 release207)[77]. Puu-related protein sequences in *E. coli*, *P. aeruginosa* and *V. parahaemolyticus*
566 were used as initial queries, which were searched against all genomes of the selected 36 species
567 using TBLASTN to identify the reference genes of each species (**Extended Data Table 2**).
568 Nucleotide and protein sequences of the reference genes for each species were then searched
569 against all genomes of the corresponding species using BLASTN and TBLASTN. We combined
570 BLASTN and TBLASTN results to identify the presence, absence or pseudogenization of genes,
571 with identify >80% and coverage >70% as cutoffs. A gene was classified as present if a
572 BLASTN or TBLASTN match was identified. If only a BLASTN match was identified, the
573 gene was classified as a pseudogene; otherwise, it was considered absent.
574
575 **Bacterial strains and growth conditions**
576 A total of 12 *V. parahaemolyticus* representative strains (pairwise SNP distances: >6) were used
577 for laboratory experiments, including reference strain RIMD2210633 (wave-3) and 11
578 previously or newly sequenced PC strains. Additionally, *V. cholerae* strain E7946 (El Tor,
579 Ogawa) and *E. coli* strain MG1655 (K12) were included. *V. parahaemolyticus* strains were
580 grown routinely in rich broth (LB with 3% NaCl) with shaking or on agar plates at 30°C
581 aerobically. *E. coli* and *V. cholerae* strains were cultured in LB at 37°C.
582
583 **Mutant construction**
584 Deletion mutants of *V. parahaemolyticus* (Δ*puuA* and Δ*puuC*) were constructed by homologous
585 recombination and counterselection with the suicide plasmid pDM4 as previously described[78],
586 with 1kb of homology upstream and downstream of the desired mutation. For *V. cholerae*, a
587 similar approach with pDM4 was used as for *V. parahaemolyticus* (Δ*puuC* and Δ*spuC*), except
588 *E. coli* MFD*pir* was used as donor strain for conjugation. *E. coli* deletion mutants (Δ*puuA* and
589 Δ*puuC*) were constructed using λ red recombination as previously described[79]. *V.*
590 *parahaemolyticus* deletion strains were complemented with the corresponding gene *in trans* on
591 plasmid pMMB207 using the Ptac promoter. All deletion and complement clones were verified
592 by colony PCR. All constructed strains and plasmids are listed in **Supplementary Table 7**.
593
594 **Growth curves**
595 The growth of strains was tested under three conditions: rich/high-salt medium (alkaline
596 peptone water [APW] with 3% NaCl, Beijing Land Bridge technology), modified M9-Put-As
597 medium (putrescine as sole carbon source, M9 medium supplemented with 0.04 M Na_2HPO_4,
598 0.02 M KH_2PO_4,0.42 mM $MgSO_4$, 1mM $CaCl_2$, 3% NaCl, 0.4% $(NH4)_2SO_4$, 0.4% putrescine
599 dihydrochloride, Sangon Biotech) and modified W-Glc-Put medium (putrescine as the sole
600 nitrogen source, M9 medium supplemented with 0.06 M K_2HPO4, 0.03 M KH_2PO_4, 0.42 mM
601 $MgSO_4$, 1mM $CaCl_2$, 3% NaCl, 0.4% glucose, 0.2% putrescine dihydrochloride)[18].

All *V. parahaemolyticus* strains were initially cultured overnight in APW with 3% NaCl at 37°C, followed by 1:100 dilution and growth to logarithmic phase. Log-phage cultures were inoculated into specified media and monitored for growth rate every 15 min at OD_{600} for 24 or 72 hours at 37°C under static conditions using BioTek Synergy H1 (Agilent Technologies).

Biofilm assays

Log-phase cultures were inoculated into different media (APW or M9 with 3% NaCl, or M9 with 1% NaCl) in 96-well plates (Kingmorn biotechnology) and incubated under static conditions at 37°C for 48 hours. Following incubation, culture media was discarded, and 96-well plates were washed with 0.9% NaCl twice, dried at 60 °C for 30 min, and stained with 100 µl 0.1% (w/v) crystal violet solution for 10 min. After staining, the plates were washed with 0.9% NaCl three times. The bound crystal violet was resolubilized with 100 µl 95% ethanol for 10 min and quantified via OD_{595} with BioTek Synergy H1 (Agilent Technologies). Each strain was tested with 30 technical replicates.

RNA extraction and transcriptome analysis

Strains were grown to stationary phase in M9 media with 1% NaCl, with or without exogenous putrescine (2.5 mM putrescine dihydrochloride) and harvested after treatment with stop mix (5% phenol, 95% ethanol). Total RNA was extracted from cells using the hot-phenol method as described previously[80]. RNA was DNase I-digested and rRNA depleted. cDNA libraries were constructed at Novogene Co. Ltd. (Beijing, China) and sequenced on an Illumina NovaSeq platform.

Sequencing reads were trimmed using Trimmomatic v0.39[59] and then mapped to the reference genome using BWA v0.7.17[81]. The number of reads assigned to genes were counted using featureCounts v2.0.6[82]. Differentially expressed genes and pathway enrichment analysis were performed using R packages DESeq2 v1.34.0[83] and ClusterProfiler v4.2.2[84].

Metabolite quantification

The OD_{600} of cultures was monitored until the desired growth phases (log-phase: ~4h, stationary phase: ~13h) at 37°C in M9 medium with 1% NaCl. Then, 10 OD_{600} of culture was collected and metabolites were extracted using 100% methanol. Next, 250 µl of supernatant was transferred to a 1.5 ml tube and mixed with 1 ml 100% LC-MS Grade methanol. The cell pellet was washed once with 5ml ice cold 1X phosphate-buffered saline (PBS) and resuspended in 1ml 100% LC-MS Grade methanol and transferred to a 1.5 ml centrifuge tube. The supernatant and cell pellet with methanol added were incubated in an ultrasonic bath with ice for 15 min at 35 kHz. The mixture was centrifuged twice for 5 min and 10 min at 4 ℃ and 15,000 g and each time the supernatant was transferred to a new tube. 900 µL of the supernatant was concentrated using a vacuum concentrator (SPD130DLX-230, Thermo Scientific) for 3 to 4 hours without heating. The dried samples were kept at -80 ℃ until resuspension using 30 µl of 60/40 methanol/water. The suspended samples were centrifuged for 15 min at 4 ℃ and 15,000 RPM and aliquoted into two portions, 20 µl for spermidine quantification and 10 µl for AQC (aminoquinoly-N-hydroxysuccinimidyl carbamate) derivatization. For AQC

derivatization, 70 µl of borate buffer was added to 10 µl aliquot. The mixture was homogenized using ultrasonic bath for 5 min. 20 µl AQC was added followed by a short vortex. The mixture was kept at room temperature for 1 min, then incubated for 15 min at 55 ℃, with 700 RPM shaking. The sample was centrifuged for 2 min at room temperature at 17,000 g.

Targeted metabolomics to quantify putrescine and GABA was performed using a Liquid Chromatograph Mass Spectrometer (LC-MS, Thermo Fisher Scientific Inc.) by the Mass Spectrometry services platform at the Shanghai Institute of Immunity and Infection. Specifically, the analysis of derived putrescine and GABA was conducted using a UHPLC-ESI-MS/MS system consisting of a TSQ Quantis mass spectrometer (Thermo Scientific, USA) with an electrospray ionization source (ESI) and an Vanquish^{TM} Flex UHPLC System (Thermo Scientific, USA). Data acquisition and processing were performed using Xcalibur 4.5 and Trace Finder software (version 5.1, Thermo Scientific, USA). The UHPLC separation was carried out on a reversed-phase C18 column (XSelect HSS T3 2.5um, 2.1mm X150mm; Waters) with a flow rate of 0.4 mL/min at 37 °C. The analysis was performed using a gradient solvent system consisting of 0.1%FA in water (Phase A) and acetonitrile (Phase B). A 25-min chromatography elution gradient was chosen as follows: 0-5min, 5%-12% B; 5-15min, 12%-35% B; 15-20min, 35%-75% B; 20-22min, 75%-95% B; 22.1-25min, 5%-12% B. The mass spectrometry detection was performed under positive electrospray ionization mode. We determined the retention times of putrescine and GABA by analyzing their standards. Subsequently, we measured the chromatographic peak areas at the retention times corresponding to the putrescine and GABA standards in the samples. We preformed relative quantification by using the detected values of chromatographic peak areas to estimate the relative concentrations of metabolites in the samples.

c-di-GMP assay

To measure intracellular c-di-GMP, we used a previously published riboswitch-based reporter system[85]. Log-phase cultures of strains carrying the reporter were inoculated into M9 medium with 1% NaCl and 50 µg/ml spectinomycin in 96-well plates (six replicate wells per strain). Additionally, six wells of blank control with 200 µl uninoculated medium were included. The plate was incubated in a BioTek Synergy H1 plate reader (Agilent Technologies) at 37°C for 48 hours. OD_{600} and fluorescence (455 nm excitation/489 nm emission for AmCyan and 540 nm excitation/574 nm emission for TurboRFP) were measured every 15 min. The average values of the control wells were subtracted from each measurement (OD_{600}, AmCyan, and TurboRFP) of the strains. The mean values were used to plot growth curves and the whisker length represents standard deviation.

Cell adhesion

Log-phage strain cultures were washed with PBS and then resuspended in cell culture medium without antibiotics. Caco-2 cells grown in 24-well plates were infected with the bacterial suspension at an MOI of 10:1. After 45 min, cells were washed three times with PBS to remove the non-attached bacteria and then lysed with 0.1% saponin for 10 min. The number of cell-associated bacteria was determined by plating serial dilutions of the lysate on LB agar plates. The percentage of cell-associated bacteria was calculated as (recovered CFU/inoculated CFU) × 100%, and the results were presented as adhesion relative to wild-type strain (RIMD-W3).

690
691 **Mouse colonization**
692 Female BALB/c mice aged 6-8 weeks were used to test strain colonization. Mice were
693 randomly divided into four groups, including a control group and three groups inoculated with
694 different strains (three mice each treated group). All mice were fasted for 8 hours prior to gavage.
695
696 We introduced plasmids with *lux* reporters (pWHQ205, encoding luciferase and substrate,
697 constructed in this study) into three strains (RIMD-W3, Δ*puuA* and Δ*puuC*), respectively.
698 Strains carrying *lux* reporters were cultured in LB medium with 3% NaCl supplemented with
699 chloramphenicol (25 μg/ml) to logarithmic phase (OD_{600}=0.6~0.8) at 30°C. Strain
700 concentration was then adjusted to 3×10^{10} CFU/ml with PBS, and each mouse was gavaged
701 with 0.3 ml of the bacterial suspension. At 5 and 10 hours post-infection, the mice were
702 anesthetized with 3% isoflurane for 5 minutes and exposed for 1 minute using the PerkinElmer
703 IVIS Spectrum imaging system (E-Heng Import and Export Co. Ltd).
704
705 **Mouse lethality**
706 Female BALB/c mice aged 4-5 weeks were used for lethality experiment, which were randomly
707 divided into five groups treated by different strains, with each consisting of 10 mice. Log-phage
708 strain cultures were washed three times with PBS and then adjusted to a concentration of 10^9
709 CFU/ml with PBS. Mice were intraperitoneally injected with 100 μl of strain suspension, and
710 the number of mice that died at a specific time point was record. The experimental protocols of
711 colonization and lethality of mice were approved by the animal ethics committee of Shanghai
712 Institute of Immunity and Infection (No. A2023011).
713
714 **Growth competition assay**
715 Log-phage cultures of 12 natural strains, including six wave-1/2/3 and six wave-4 strains
716 (**Supplementary Table 7**), were individually diluted to an OD_{600} of 0.08. Subsequently, the
717 diluted cultures of wave-4 strains were mixed with those of wave-1 to wave-3 strains in equal
718 (1:1) or different (1:5) ratios. These mixtures were then incubated in seawater (collected from
719 Jinshan, Shanghai) supplemented with 0.5% chitin (Sigma) for 72 hours at 30°C under static
720 conditions. Genomic DNA of harvested cultures was extracted using the TIANamp Bacteria
721 DNA kit (Tiangen Biotech). PCR primers VPA0427SNPU (AGCGACGACGCCAGAGAAAG)
722 and VPA0427SNPD (GGTGTCTAAAGTACGGTGCAG) were used to amplify the gene
723 VPA0427, which contains a wave-4 specific SNP allele (A919G).
724
725 The PCR amplification products were used for library construction at Novogene Co. Ltd.
726 (Beijing, China) and sequenced on an Illumina NovaSeq platform. Sequencing reads are
727 trimmed using Trimmomatic v0.39[59] and were then mapped to *V. parahaemolyticus* reference
728 genome using BWA v0.7.17[81]. The frequency of the wave-4 specific allele was calculated to
729 represent the relative abundance of wave-4 strains.
730
731 **Statistical analysis**
732 Datasets were compared using the Mann-Whitney U test for continuous data and Chi-squared tests
733 for categorical data. Log-rank (Mantel-Cox) tests were used for survival curve comparison. p-values

less than 0.05 were considered statistically significant.

Data availability

The sequencing data have been deposited in the NCBI SRA or GenBank database under accession number PRJNA1117214 and PRJNA1062747. The metabolomics data have been deposited in the Metabolights database (accession number: MTBLS11560). Background information of sequenced isolate is listed in Supplementary Table 1-2. Source data are provided with this paper.

Reference

1. Bonneaud, C. & Longdon, B. Emerging pathogen evolution: Using evolutionary theory to understand the fate of novel infectious pathogens. *EMBO Rep* **21**, (2020).
2. Balasubramanian, D., López-Pérez, M., Grant, T. A., Ogbunugafor, C. B. & Almagro-Moreno, S. Molecular mechanisms and drivers of pathogen emergence. *Trends in Microbiology* vol. 30
3. Baker-Austin, C. *et al.* Vibrio spp. infections. *Nat Rev Dis Primers* **4**, 8 (2018).
4. Hu, D. *et al.* Origins of the current seventh cholera pandemic. *Proc Natl Acad Sci U S A* **113**, E7730–E7739 (2016).
5. Mutreja, A. *et al.* Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* **477**, 462–465 (2011).
6. Didelot, X. *et al.* The Role of China in the Global Spread of the Current Cholera Pandemic. *PLoS Genet* **11**, (2015).
7. Weill, F.-X. *et al.* Genomic History of the Seventh Pandemic of Cholera in Africa. Science, , 358: 785-789 (2017).
8. Baker-Austin, C., Stockley, L., Rangdale, R. & Martinez-Urtaza, J. Environmental occurrence and clinical impact of Vibrio vulnificus and Vibrio parahaemolyticus: A European perspective. *Environ Microbiol Rep* **2**, 7–18 (2010).
9. Yang, C. *et al.* Recent mixing of Vibrio parahaemolyticus populations. *ISME J* **13**, 2578–2588 (2019).
10. Nair, G. B. *et al.* Global dissemination of Vibrio parahaemolyticus serotype O3:K6 and its serovariants. *Clin Microbiol Rev* **20**, 39–48 (2007).
11. Espejo, R. T., García, K. & Plaza, N. Insight into the origin and evolution of the Vibrio parahaemolyticus pandemic strain. *Front Microbiol* **8**, 1–6 (2017).
12. Han, D., Yu, F., Chen, X., Zhang, R. & Li, J. Challenges in Vibrio parahaemolyticus infections caused by the pandemic clone. *Future Microbiol* **14**, 437–450 (2019).
13. Matsumoto, C. *et al.* Pandemic spread of an O3:K6 clone of Vibrio parahaemolyticus and emergence of related strains evidenced by arbitrarily primed PCR and toxRS sequence analyses. *J Clin Microbiol* **38**, (2000).
14. Nasu, H. *et al.* A filamentous phage associated with recent pandemic Vibrio parahaemolyticus O3:K6 strains. *J Clin Microbiol* **38**, (2000).
15. Hurley, C. C., Quirke, A. M., Reen, F. J. & Boyd, E. F. Four genomic islands that mark post-1995 pandemic Vibrio parahaemolyticus isolates. *BMC Genomics* **7**, 1–19 (2006).
16. Boyd, E. F. *et al.* Molecular analysis of the emergence of pandemic Vibrio parahaemolyticus. *BMC Microbiol* **8**, 1–14 (2008).

17. Cui, Y. *et al.* Epidemic clones, oceanic gene pools, and eco-LD in the free living marine pathogen vibrio parahaemolyticus. *Mol Biol Evol* **32**, 1396–1410 (2015).
18. Schneider, B. L. & Reitzer, L. Pathway and enzyme redundancy in putrescine catabolism in Escherichia coli. *J Bacteriol* **194**, (2012).
19. Liu, Z., Hossain, S. S., Moreira, Z. M. & Haney, C. H. Putrescine and Its Metabolic Precursor Arginine Promote Biofilm and c-di-GMP Synthesis in Pseudomonas aeruginosa. *J Bacteriol* **204**, (2022).
20. Sugiyama, Y. *et al.* A novel putrescine exporter SapBCDF of Escherichia coli. *Journal of Biological Chemistry* **291**, (2016).
21. Karatan, E., Duncan, T. R. & Watnick, P. I. NspS, a predicted polyamine sensor, mediates activation of Vibrio cholerae biofilm formation by norspermidine. *J Bacteriol* **187**, (2005).
22. Prentice, J. A., van de Weerd, R. & Bridges, A. A. Cell-lysis sensing drives biofilm formation in Vibrio cholerae. *Nat Commun* **15**, (2024).
23. Cockerell, S. R. *et al.* Vibrio cholerae NspS, a homologue of ABC-type periplasmic solute binding proteins, facilitates transduction of polyamine signals independent of their transport. *Microbiology (United Kingdom)* **160**, (2014).
24. Shi, Z. *et al.* Putrescine is an intraspecies and interkingdom cell-cell communication signal modulating the virulence of dickeya zeae. *Front Microbiol* **10**, (2019).
25. Gu, W. *et al.* Cyclic di-GMP interact with putrescine via a PilZ domain receptor YcgR. *bioRxiv* (2023).
26. Banerji, R., Kanojiya, P., Patil, A. & Saroj, S. D. Polyamines in the virulence of bacterial pathogens of respiratory tract. *Mol Oral Microbiol* **36**, (2021).
27. Zamorano-Sánchez, D., Alejandre-Sixtos, J. E., Arredondo-Hernández, A. & Martínez-Méndez, R. OpaR Exerts a Dynamic Control over c-di-GMP Homeostasis and cpsA Expression in Vibrio parahaemolyticus through Its Regulation of ScrC and the Trigger Phosphodiesterase TpdA . *Microbiol Spectr* **11**, (2023).
28. Kline, K. A., Fälker, S., Dahlberg, S., Normark, S. & Henriques-Normark, B. Bacterial Adhesins in Host-Microbe Interactions. *Cell Host and Microbe* vol. 5 Preprint at https://doi.org/10.1016/j.chom.2009.05.011 (2009).
29. Yang, C. *et al.* Outbreak dynamics of foodborne pathogen Vibrio parahaemolyticus over a seventeen year period implies hidden reservoirs. *Nat Microbiol* **7**, 1221–1229 (2022).
30. Datta, M. S. *et al.* Rapid methicillin resistance diversification in Staphylococcus epidermidis colonizing human neonates. *Nat Commun* **12**, (2021).
31. Iwasaki, W. & Takagi, T. Rapid pathway evolution facilitated by horizontal gene transfers across prokaryotic lineages. *PLoS Genet* **5**, (2009).
32. Domman, D. *et al.* Integrated View of Vibrio Cholerae in the Americas. Science **358**, 789-793 (2017).
33. Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proceedings of the National Academy of Sciences* **110**, 577–582 (2013).
34. Rabosky, D. L. & Lovette, I. J. Explosive evolutionary radiations: Decreasing speciation or increasing extinction through time? *Evolution (N Y)* **62**, (2008).
35. Morlon, H., Kemps, B. D., Plotkin, J. B. & Brisson, D. Explosive radiation of a bacterial species group. *Evolution (N Y)* **66**, (2012).
36. Markov, P. V. *et al.* The evolution of SARS-CoV-2. *Nature Reviews Microbiology* vol. 21

(2023).
37. Flemming, H. C. *et al.* Biofilms: An emergent form of bacterial life. *Nature Reviews Microbiology* 14 (2016).
38. Krammer, E. M. & Prévost, M. Function and Regulation of Acid Resistance Antiporters. *Journal of Membrane Biology* **252**, (2019).
39. Romano, A., Ladero, V., Alvarez, M. A. & Lucas, P. M. Putrescine production via the ornithine decarboxylation pathway improves the acid stress survival of Lactobacillus brevis and is part of a horizontally transferred acid resistance locus. *Int J Food Microbiol* **175**, (2014).
40. Hohmann, A. W., Schmidt, G. & Rowley, D. Intestinal colonization and virulence of Salmonella in mice. *Infect Immun* **22**, (1978).
41. Wang, Y. *et al.* Isolation, phylogenetic group, drug resistance, biofilm formation, and adherence genes of Escherichia coli from poultry in central China. *Poult Sci* **95**, (2016).
42. Flament-Simon, S. C. *et al.* Association between kinetics of early biofilm formation and clonal lineage in Escherichia coli. *Front Microbiol* **10**, (2019).
43. Hogins, J., Xuan, Z., Zimmern, P. E. & Reitzer, L. The distinct transcriptome of virulence-associated phylogenetic group B2 Escherichia coli . *Microbiol Spectr* **11**, (2023).
44. Nowrouzian, F. L., Wold, A. E. & Adlerberth, I. Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *Journal of Infectious Diseases* **191**, (2005).
45. Grozdanov, L. *et al.* Analysis of the genome structure of the nonpathogenic probiotic Escherichia coli strain Nissle 1917. *J Bacteriol* **186**, (2004).
46. Chattaway, M. A., Schaefer, U., Tewolde, R., Dallman, T. J. & Jenkins, C. Identification of Escherichia coli and shigella species from whole-genome sequences. *J Clin Microbiol* **55**, (2017).
47. Viana, D. *et al.* A single natural nucleotide mutation alters bacterial pathogen host tropism. *Nat Genet* **47**, (2015).
48. Arnold, B. J., Huang, I. T. & Hanage, W. P. Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews Microbiology* 20 206–218 (2022).
49. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C. & San Millán, Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nature Reviews Microbiology* vol. 19 347–359 (2021).
50. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nature Reviews Genetics* vol. 17 379–391 (2016).
51. Wolf, Y. I. & Koonin, E. V. Genome reduction as the dominant mode of evolution. *BioEssays* **35**, (2013).
52. Pallen, M. J. & Wren, B. W. Bacterial pathogenomics. *Nature* **449**, 835–842 (2007).
53. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology* vol. 10 13–26 (2012).
54. Maurelli, A. T. Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiology Letters* vol. 267 (2007).
55. Bliven, K. A. & Maurelli, A. T. Antivirulence genes: Insights into pathogen evolution through gene loss. *Infection and Immunity* vol. 80 4061–4070 (2012).
56. Dufresne, A., Garczarek, L. & Partensky, F. Accelerated evolution associated with genome

reduction in a free-living prokaryote. *Genome Biol* **6**, (2005).

57. Morris, J. J., Lenski, R. E. & Zinser, E. R. The black queen hypothesis: Evolution of dependencies through adaptive gene loss. *mBio* **3**, (2012).

58. Alizon, S., Hurford, A., Mideo, N. & Van Baalen, M. Virulence evolution and the trade-off hypothesis: History, current state of affairs and the future. *Journal of Evolutionary Biology* vol. 22 (2009).

59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

60. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–1055 (2015).

61. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 1–8 (2018).

62. Delcher, A. L., Salzberg, S. L. & Phillippy, A. M. Using MUMmer to Identify Similar Regions in Large Sequence Sets . *Curr Protoc Bioinformatics* **00**, (2003).

63. Page, A. J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* **2**, e000056 (2016).

64. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).

65. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).

66. Kozlov, A. M. *et al.* RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* (2019).

67. 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).

68. Edwards, D., Duchêne, S., Pope, B. & Holt, K. SNPPar: identifying convergent evolution and other homoplasies from microbial whole-genome alignments. **264017**, 1–33 (2020).

69. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**, (2017).

70. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. W. & Corander, J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res* **47**, 5539–5549 (2019).

71. van der Graaf-Van Bloois, L., Chen, H., Wagenaar, J. A. & Zomer, A. L. Development of Kaptive databases for Vibrio parahaemolyticus O-and K-antigen genotyping. *Microb Genom* **9**, (2023).

72. Rambaut, A., Lam, T. T., Carvalho, L. M. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2**, (2016).

73. Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R. & Wilson, D. J. Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res* **46**, (2018).

74. 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).

75. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069

(2014).
76. Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* **21**, 180 (2020).
77. Parks, D. H. *et al.* GTDB: An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* **50**, D785–D794 (2022).
78. Chimalapati, S., Lafrance, A. E., Chen, L. & Orth, K. Vibrio parahaemolyticus: Basic Techniques for Growth, Genetic Manipulation, and Analysis of Virulence Factors. *Curr Protoc Microbiol* **59**, (2020).
79. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, (2000).
80. Sharma, C. M., Darfeuille, F., Plantinga, T. H. & Vogel, J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* **21**, (2007).
81. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
82. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, (2014).
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).
84. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).
85. Zhou, H. *et al.* Characterization of a natural triple-tandem c-di-GMP riboswitch and application of the riboswitch-based dual-fluorescence reporter. *Sci Rep* **6**, (2016).

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 32270003 and No. 32000008 to C.Y., No. 32170640 and No.32211550014 to D.F., No. 32250610209 to S.S.), Youth Innovation Promotion Association, Chinese Academy of Sciences (No. 2022278 to C.Y.), National Key Research and Development Program of China (No. 2022YFC2304700 to F.D.), Shanghai Rising-Star Program (No. 23QA1410500 to C.Y.), Shanghai Public Health System Construction Three-Year Action Plan (No. GWVI-11.1-43 to H.Q.) and Wuxi Science and Technology Development Fund's 'Light of Taihu Lake' Science and Technology Research Program (Basic Research) (No. K20231033 to C.N., No. K20231045 to S.G.). We are grateful to the labs of Prof. Qiyao Wang, Kim Orth, and Jin He for providing strains and plasmids. We thank Prof. Lei Pan for valuable comments.

Author contributions

C.Y., J.M.-U. and D.F. designed, initiated and coordinated the study. J.M.-U., L.X, Y.L., M.J., X.S. and Q.H. contributed to data collection. C.Y., Z.J., J.W., and W.X. performed bioinformatics analysis. H.Q, S.S., C.N., S.G., Z.J., H.W., Y.Q. and S.L. performed experiments. All authors contributed to interpretation of the data. C.Y. wrote the first draft of the paper and H.Q., S.S., Y.Z., Z.Z., Y.C., J.M.-U., H.W., R.Y., Y.C. and D.F. reviewed and revised the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

958 **Figure legends**

959 **Figure 1. From pre-pandemic (pre-PC) to pandemic clone (PC).** a) Maximum likelihood
960 tree of 8,684 global isolates. Pre-PC and PC isolates are colored in blue and red, respectively.
961 b) Inferred evolutionary history from pre-PC to PC. The status of known PC marker
962 genes/islands in the ancestral strain is indicated by blue ellipses, while subsequent gains or
963 losses are represented by orange and black triangles. The numbers next to tree nodes indicate
964 inferred divergence times and 95% confidence intervals. c) Distribution of known PC marker
965 genes/islands among pre-PC and PC strains.

966

967 **Figure 2. Fine-scale population structure, spatiotemporal and population dynamics of PC**.
968 a) Maximum likelihood tree of 942 representative PC strains. The colors represent different
969 waves. Double slashes indicate artificially truncated branches. b) Dated phylogeny of PC (left)
970 and the geographical, subtyping and distribution of genes in the putrescine utilization (Puu)
971 pathway of strains (right). The colors of tree branches indicate geographic areas, and numbers
972 next to tree nodes indicate inferred divergence times and 95% confidence intervals. TCs
973 indicate transmissions resulting in local colonization. (c) Temporal dynamics of four waves.
974 Bar colors indicate the frequency of different waves. The line represents the number of
975 representative strains. (d) Temporal distribution of representative strains in different countries
976 and regions. The size of circles scales with the number of representative strains. (e) Effective
977 population size dynamics of different waves of the PC.

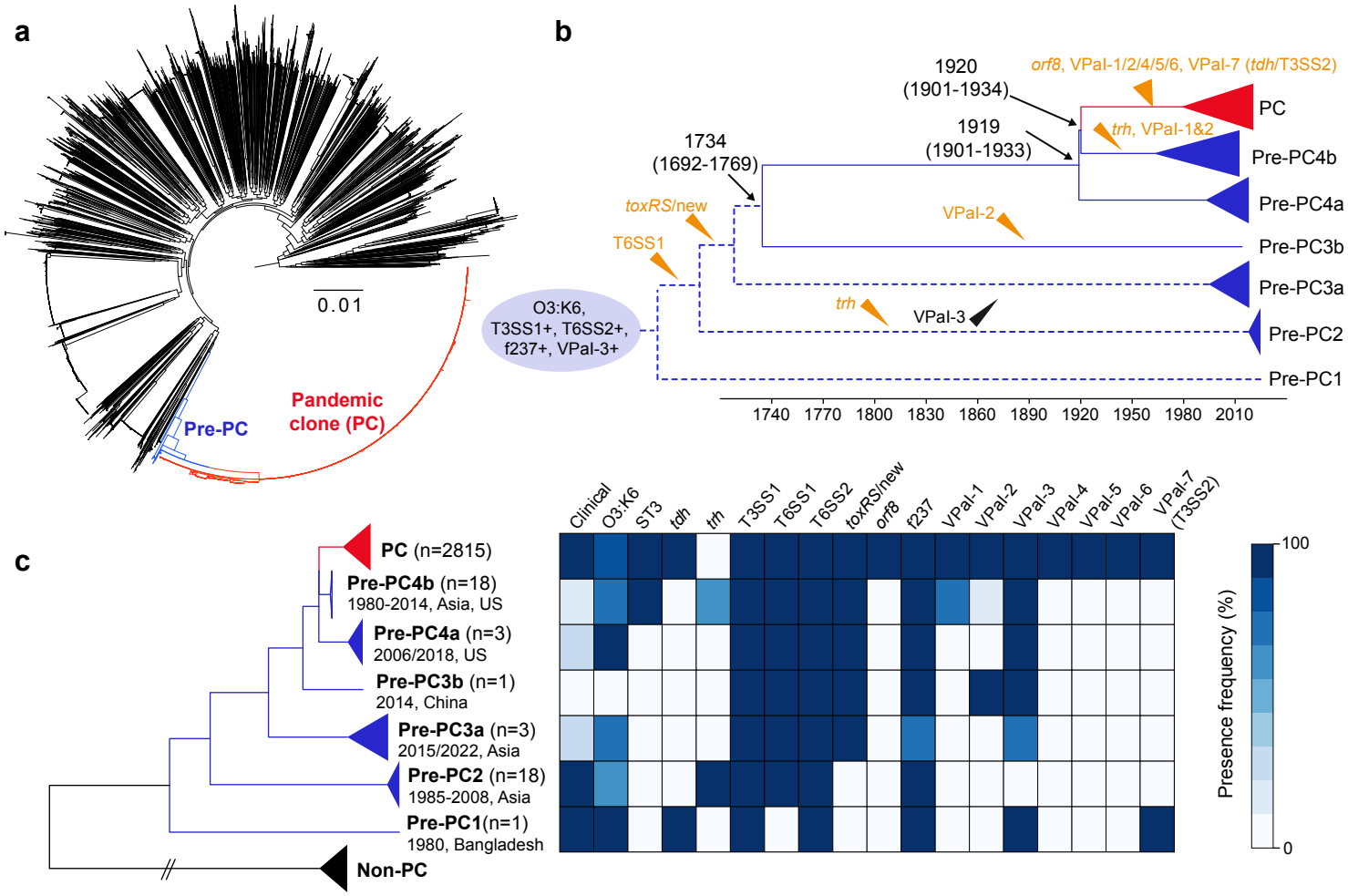
978

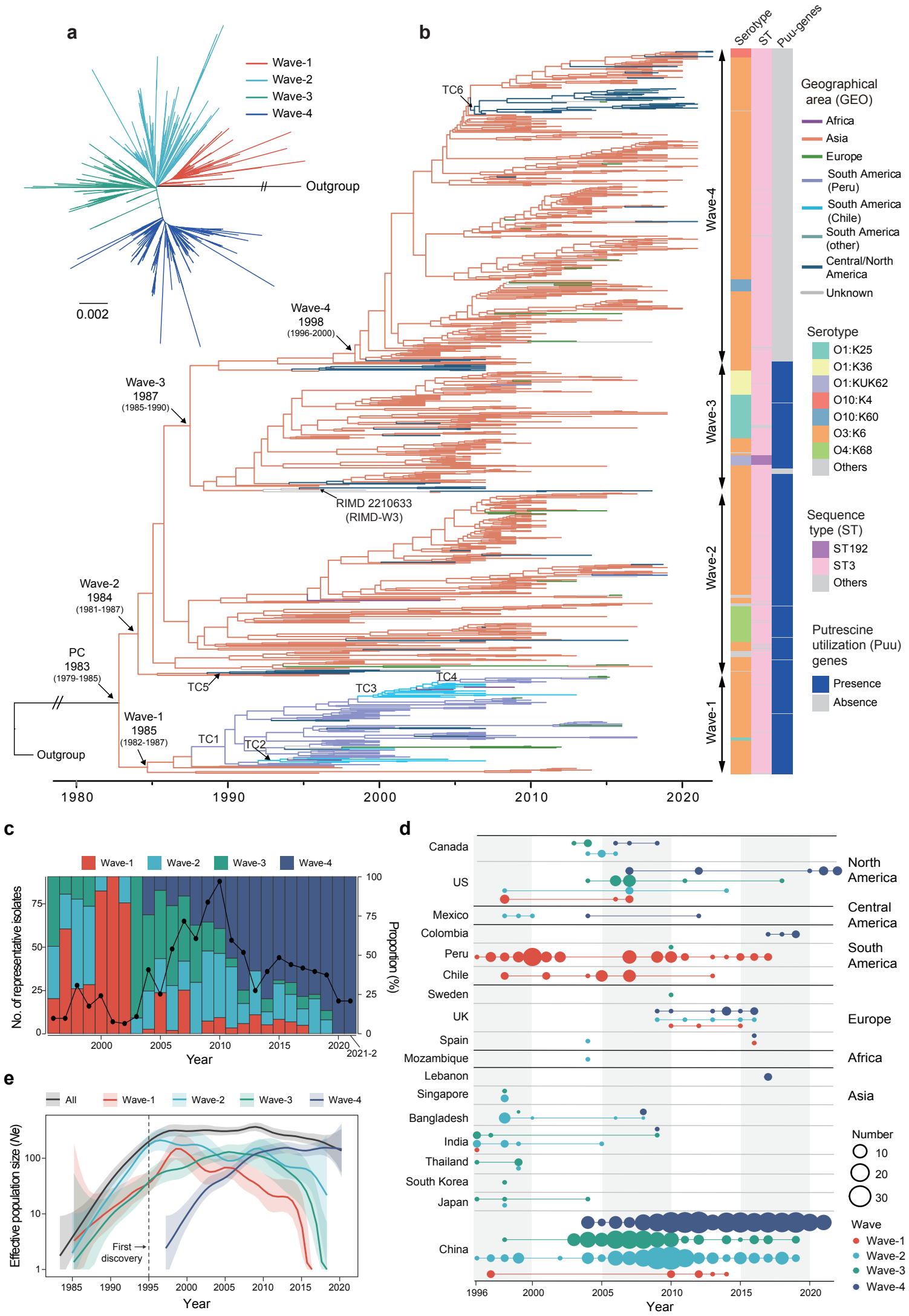
979 **Figure 3. Effect of gene-loss in putrescine utilization (Puu) pathway.** a) Positions of wave-
980 4 specific variations on the reference genome (top) and gene structure and coverage of Puu-
981 genes (colored in blue) in different waves (bottom). b) Summary of Puu-genes and metabolites
982 based on KEGG annotations. The *puuC* gene, which appears in three sub-pathway, was
983 highlighted in bold. c) Growth curves of wildtype (RIMD-W3), mutants (Δ*puuC*) and
984 complement strains (C-Δ*puuC*) when putrescine (Put) was used as sole carbon or nitrogen
985 source. d) Biofilm levels in M9 minimal media with moderate salt concentrations measured by
986 optical density at 595 nm (OD_{595}) with (+Put) or without exogenous putrescine. e)
987 Quantification of intra- and extra-cellular metabolites during logarithmic and stationary phases.
988 N.D.: not detected, indicating levels below the lower limit of detection. f) Transcriptomic
989 volcano plots showing differentially expressed genes (DEGs). The colors represent genes of
990 specific category as in the legend. Red and blue backgrounds represent up- and down-regulated
991 genes in Δ*puuC* compared to wildtype strain. g) c-di-GMP dynamics of strains with (+Put) or
992 without exogenous putrescine. h) The adhesion ability of strains to Caco-2 cells relative to
993 wildtype (RIMD-W3). i) *In vivo* bioluminescent imaging of mice at 5 hours post-infection by
994 gavage. j) Survival curves of mice by intraperitoneal infection. The measurements in panel e
995 and h were converted into relative level to wildtype. Mann-Whitney U test was used for
996 pairwise comparisons except for panel j, where log-rank (Mantel-Cox) tests were applied. *:
997 $p<0.05$, **: $p<0.01$.

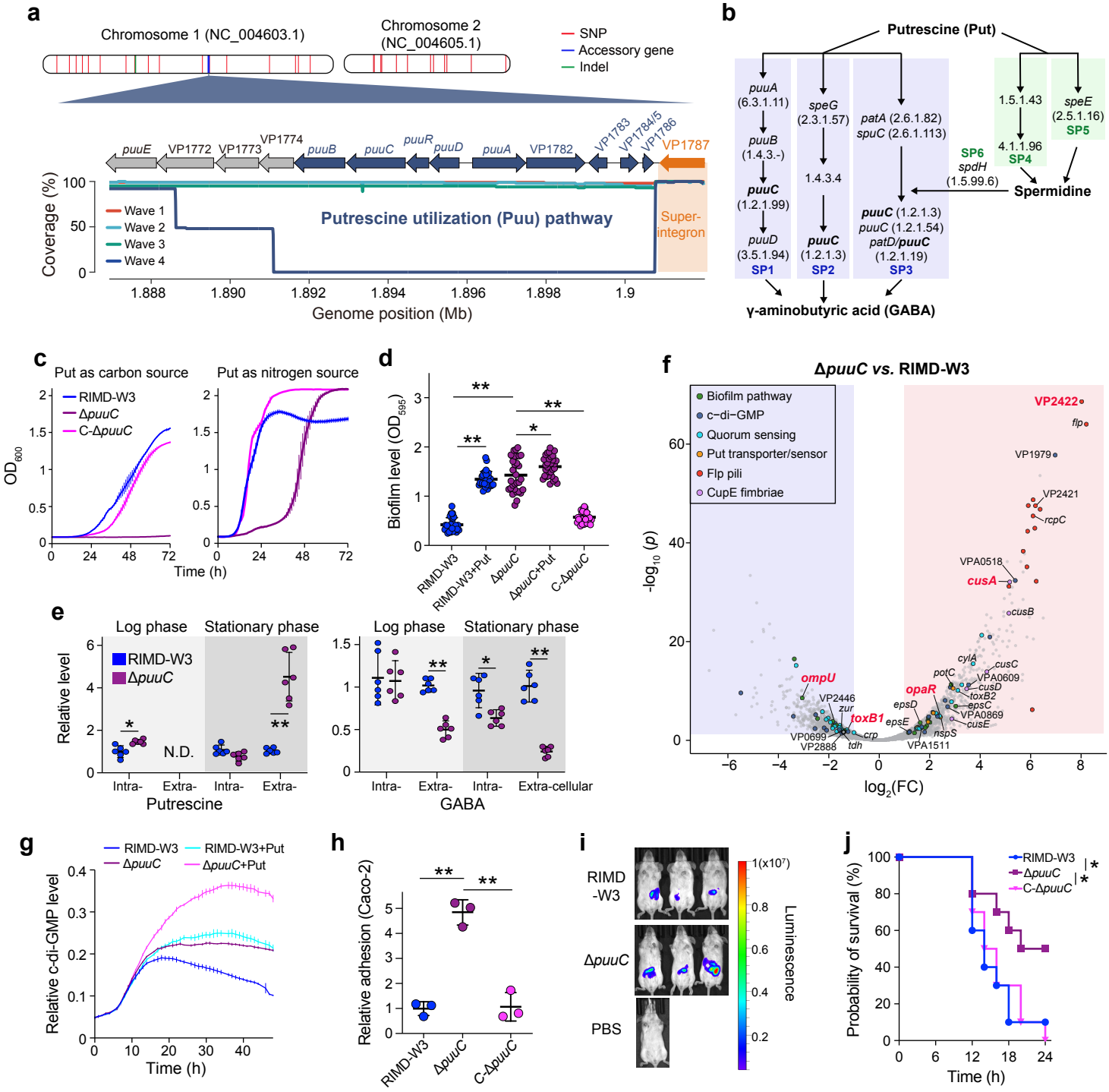
998

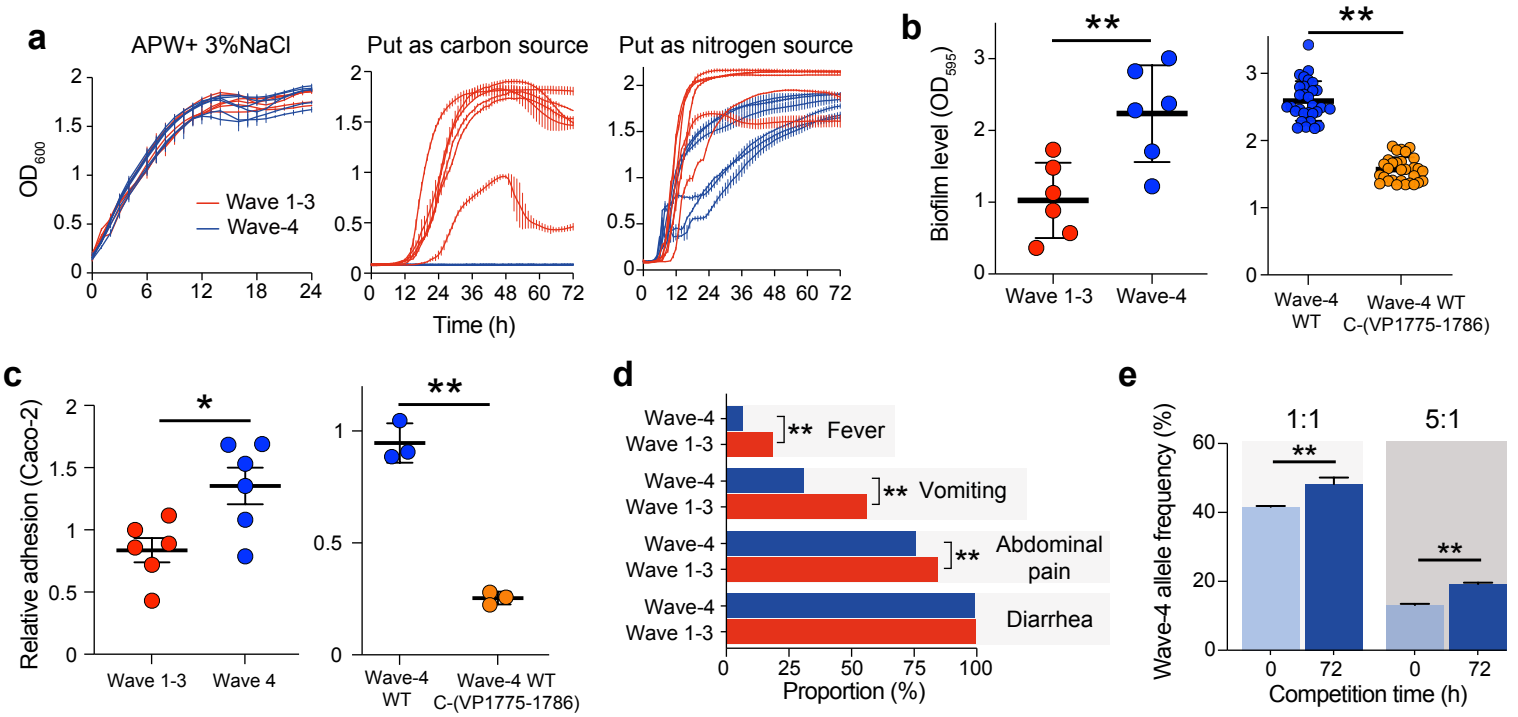
999 **Figure 4. Phenotypes and clinical symptoms of wave1-3 and wave-4 strains.** a) Growth
1000 curves under high-nutrition and high-salt (APW+ 3%NaCl) conditions and when putrescine
1001 was used as the sole carbon or nitrogen source. b) Biofilm level between wave1-3 and wave-4

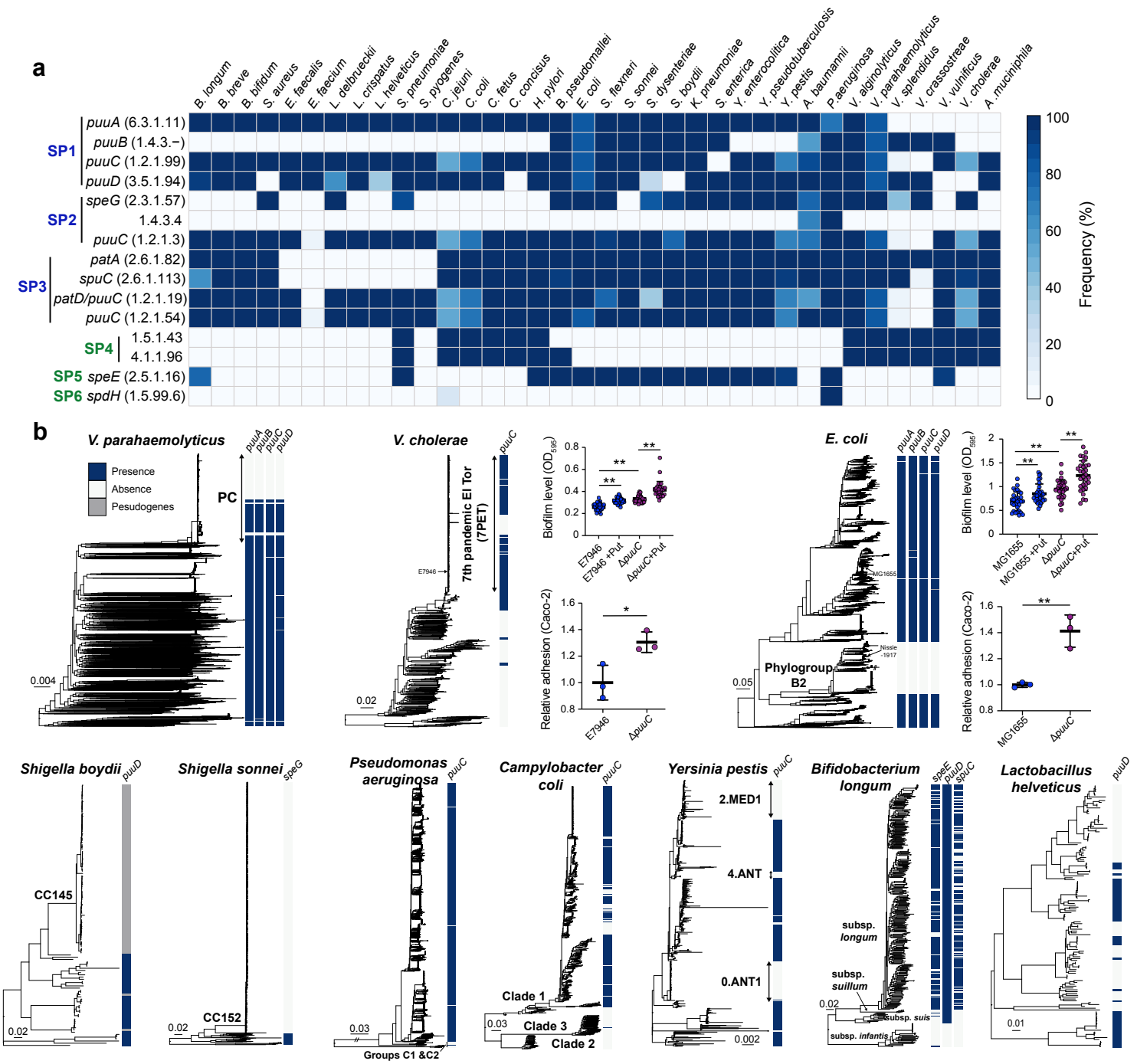
1002 strains, and between the wave-4 wildtype strain (VPCZ58, wave-4 WT) and its complemented
1003 strain (wave-4 WT C-(VP1775-1786)) carrying the nine puu-related genes (VP1775-1786). c)
1004 The adhesion ability of strains to Caco-2 cells between wave1-3 and wave-4 strains, and
1005 between the wave-4 WT and wave-4 WT C-(VP1775-1786). d) Clinical symptoms of wave-4
1006 and other wave-associated patients. e) Wave-4 specific allele frequency after 72 hours of
1007 competition under seawater with chitin conditions. The Mann-Whitney U test was used for
1008 pairwise comparisons, except for panel d, where Chi-squared tests were applied. *: $p<0.05$, **:
1009 $p<0.01$.
1010
1011 **Figure 5. Puu-genes in 36 bacterial species.** a) Prevalence of Puu-genes in 36 species. Genes
1012 and sub-pathways (SP) on the left are as in Fig. 3b. The colors (right) indicate the frequency of
1013 genes. b) Phylogenetic distribution of Puu-genes in 10 species and the biofilm and adhesion of
1014 wildtype (E7946 and MG1655) and Δ*puuC* mutant strains in *V. cholerae* and *E. coli* with (+Put)
1015 or without exogenous putrescine. Mann-Whitney U test was used for pairwise comparisons. *:
1016 $p<0.05$, **: $p<0.01$.
1017
1018 **Figure 6. Proposed model of pandemic *V. parahaemolyticus* wave succession driven by**
1019 **gene-loss in the putrescine utilization pathway.** The dashed boxes and dashed arrows
1020 represent the hypothesized relationships inferred from the transcriptomic and metabolic results.
1021 Cartoon schematics were created using the software from BioRender.com.

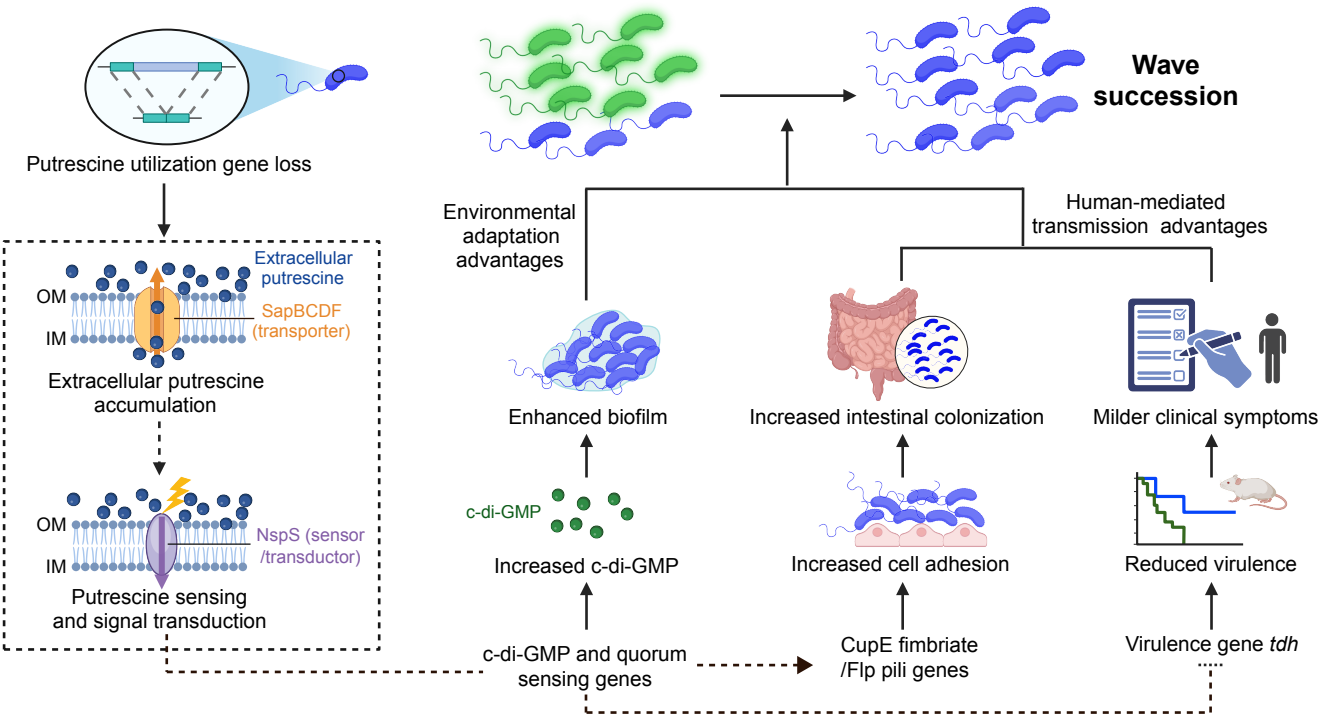

a
0.01
Pre-PC
Pandemic clone (PC)
b
1920 (1901-1934)
orf8, VPal-1/2/4/5/6, VPal-7 (tdh/T3SS2)
PC
1919 (1901-1933)
trh, VPal-1&2
Pre-PC4b
1734 (1692-1769)
Pre-PC4a
toxRS/new
VPal-2
T6SS1
Pre-PC3b
Pre-PC3a
O3:K6, T3SS1+, T6SS2+, f237+, VPal-3+
trh
VPal-3
Pre-PC2
Pre-PC1
1740 1770 1800 1830 1860 1890 1920 1950 1980 2010
c
PC (n=2815)
Pre-PC4b (n=18)
1980-2014, Asia, US
Pre-PC4a (n=3)
2006/2018, US
Pre-PC3b (n=1)
2014, China
Pre-PC3a (n=3)
2015/2022, Asia
Pre-PC2 (n=18)
1985-2008, Asia
Pre-PC1(n=1)
1980, Bangladesh
Non-PC
Clinical
O3:K6
ST3
tdh
trh
T3SS1
T6SS1
T6SS2
toxRS/new
orf8
f237
VPal-1
VPal-2
VPal-3
VPal-4
VPal-5
VPal-6
VPal-7 (T3SS2)
Presence frequency (%)
100
0


a
Wave-1
Wave-2
Wave-3
Wave-4
Outgroup
0.002
b
Serotype
ST
Puu-genes
TC6
Wave-4
1998
(1996-2000)
Wave-3
1987
(1985-1990)
RIMD 2210633
(RIMD-W3)
Wave-2
1984
(1981-1987)
PC
1983
(1979-1985)
TC5
TC4
TC3
Wave-1
1985
(1982-1987)
TC1
TC2
Outgroup
1980
1990
2000
2010
2020
Geographical area (GEO)
Africa
Asia
Europe
South America (Peru)
South America (Chile)
South America (other)
Central/North America
Unknown
Serotype
O1:K25
O1:K36
O1:KUK62
O10:K4
O10:K60
O3:K6
O4:K68
Others
Sequence type (ST)
ST192
ST3
Others
Putrescine utilization (Puu) genes
Presence
Absence
c
Wave-1
Wave-2
Wave-3
Wave-4
No. of representative isolates
Proportion (%)
Year
2021-2
d
Canada
US
Mexico
Colombia
Peru
Chile
Sweden
UK
Spain
Mozambique
Lebanon
Singapore
Bangladesh
India
Thailand
South Korea
Japan
China
North America
Central America
South America
Europe
Africa
Asia
Number
10
20
30
Wave
Wave-1
Wave-2
Wave-3
Wave-4
Year
e
All
Wave-1
Wave-2
Wave-3
Wave-4
Effective population size (Ne)
First discovery
Year

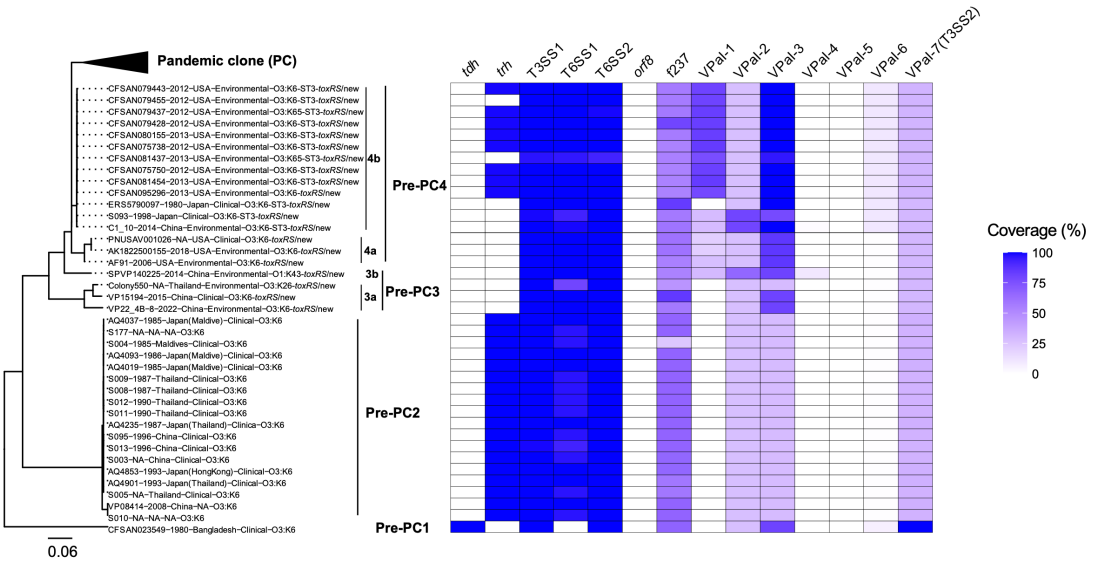

a
Chromosome 1 (NC_004603.1)
Chromosome 2 (NC_004605.1)
SNP
Accessory gene
Indel
puuE
VP1772
VP1773
VP1774
puuB
puuC
puuR
puuD
puuA
VP1782
VP1783
VP1784/5
VP1786
VP1787
Coverage (%)
Wave 1
Wave 2
Wave 3
Wave 4
Putrescine utilization (Puu) pathway
Super-integron
Genome position (Mb)
b
Putrescine (Put)
puuA (6.3.1.11)
puuB (1.4.3.-)
puuC (1.2.1.99)
puuD (3.5.1.94)
SP1
speG (2.3.1.57)
1.4.3.4
puuC (1.2.1.3)
SP2
patA (2.6.1.82)
spuC (2.6.1.113)
puuC (1.2.1.3)
puuC (1.2.1.54)
patD/puuC (1.2.1.19)
SP3
1.5.1.43
4.1.1.96
SP4
speE (2.5.1.16)
SP5
SP6
spdH (1.5.99.6)
Spermidine
γ-aminobutyric acid (GABA)
c
Put as carbon source
Put as nitrogen source
RIMD-W3
ΔpuuC
C-ΔpuuC
OD600
Time (h)
d
Biofilm level (OD595)
RIMD-W3
RIMD-W3+Put
ΔpuuC
ΔpuuC+Put
C-ΔpuuC
e
Log phase
Stationary phase
RIMD-W3
ΔpuuC
Relative level
N.D.
Intra-
Extra-
Putrescine
GABA
Extra-cellular
f
ΔpuuC vs. RIMD-W3
Biofilm pathway
c-di-GMP
Quorum sensing
Put transporter/sensor
Flp pili
CupE fimbriae
VP2422
flp
VP1979
VP2421
rcpC
VPA0518
cusA
cusB
cylA
cusC
potC
VPA0609
cusD
toxB2
epsC
VPA0869
cusE
opaR
epsD
epsE
toxB1
nspS
VPA1511
ompU
VP2446
zur
VP0699
VP2888
tdh
crp
-log10 (p)
log2(FC)
g
RIMD-W3
RIMD-W3+Put
ΔpuuC
ΔpuuC+Put
Relative c-di-GMP level
Time (h)
h
Relative adhesion (Caco-2)
RIMD-W3
ΔpuuC
C-ΔpuuC
i
RIMD-W3
ΔpuuC
PBS
1(x10⁷)
Luminescence
j
RIMD-W3
ΔpuuC
C-ΔpuuC
Probability of survival (%)
Time (h)


a
APW+ 3%NaCl
Put as carbon source
Put as nitrogen source
OD600
Wave 1-3
Wave-4
Time (h)
b
Biofilm level (OD595)
**
Wave 1-3
Wave-4
Wave-4 WT
Wave-4 WT C-(VP1775-1786)
c
Relative adhesion (Caco-2)
*
**
Wave 1-3
Wave 4
Wave-4 WT
Wave-4 WT C-(VP1775-1786)
d
Wave-4
Wave 1-3
** Fever
** Vomiting
** Abdominal pain
Diarrhea
Proportion (%)
e
1:1
5:1
Wave-4 allele frequency (%)
**
Competition time (h)

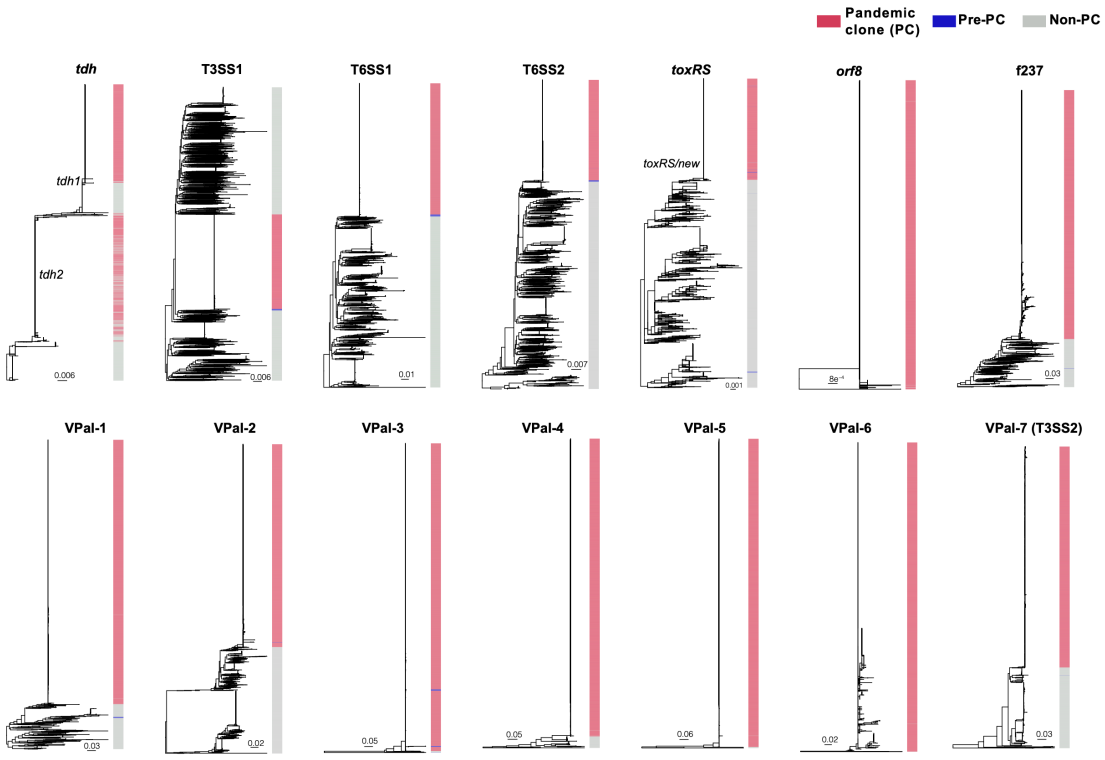

a
B. longum
B. breve
B. bifidum
S. aureus
E. faecalis
E. faecium
L. delbrueckii
L. crispatus
L. helveticus
S. pneumoniae
S. pyogenes
C. jejuni
C. coli
C. fetus
C. concisus
H. pylori
B. pseudomallei
E. coli
S. flexneri
S. sonnei
S. dysenteriae
S. boydii
K. pneumoniae
S. enterica
Y. enterocolitica
Y. pseudotuberculosis
Y. pestis
A. baumannii
P. aeruginosa
V. alginolyticus
V. parahaemolyticus
V. splendidus
V. crassostreae
V. vulnificus
V. cholerae
A. muciniphila
SP1
puuA (6.3.1.11)
puuB (1.4.3.−)
puuC (1.2.1.99)
puuD (3.5.1.94)
SP2
speG (2.3.1.57)
1.4.3.4
puuC (1.2.1.3)
SP3
patA (2.6.1.82)
spuC (2.6.1.113)
patD/puuC (1.2.1.19)
puuC (1.2.1.54)
SP4
1.5.1.43
4.1.1.96
SP5 speE (2.5.1.16)
SP6 spdH (1.5.99.6)
Frequency (%)
100
80
60
40
20
0
b
V. parahaemolyticus
Presence
Absence
Pesudogenes
PC
puuA
puuB
puuC
puuD
0.004
V. cholerae
puuC
7th pandemic El Tor (7PET)
E7946
0.02
Biofilm level (OD595)
E7946
E7946 +Put
ΔpuuC
ΔpuuC+Put
Relative adhesion (Caco-2)
E. coli
MG1655
Nissle 1917
Phylogroup B2
0.05
MG1655 +Put
Shigella boydii
CC145
0.02
Shigella sonnei
CC152
speG
Pseudomonas aeruginosa
0.03
Groups C1 &C2
Campylobacter coli
Clade 1
Clade 3
Clade 2
Yersinia pestis
2.MED1
4.ANT
0.ANT1
0.002
Bifidobacterium longum
speE
subsp. longum
subsp. suillum
subsp. suis
subsp. infantis
Lactobacillus helveticus
0.01


Putrescine utilization gene loss
Extracellular putrescine
OM
IM
SapBCDF (transporter)
Extracellular putrescine accumulation
OM
IM
NspS (sensor /transductor)
Putrescine sensing and signal transduction
Wave succession
Environmental adaptation advantages
Human-mediated transmission advantages
Enhanced biofilm
Increased intestinal colonization
Milder clinical symptoms
c-di-GMP
Increased c-di-GMP
Increased cell adhesion
Reduced virulence
c-di-GMP and quorum sensing genes
CupE fimbriate /Flp pili genes
Virulence gene tdh

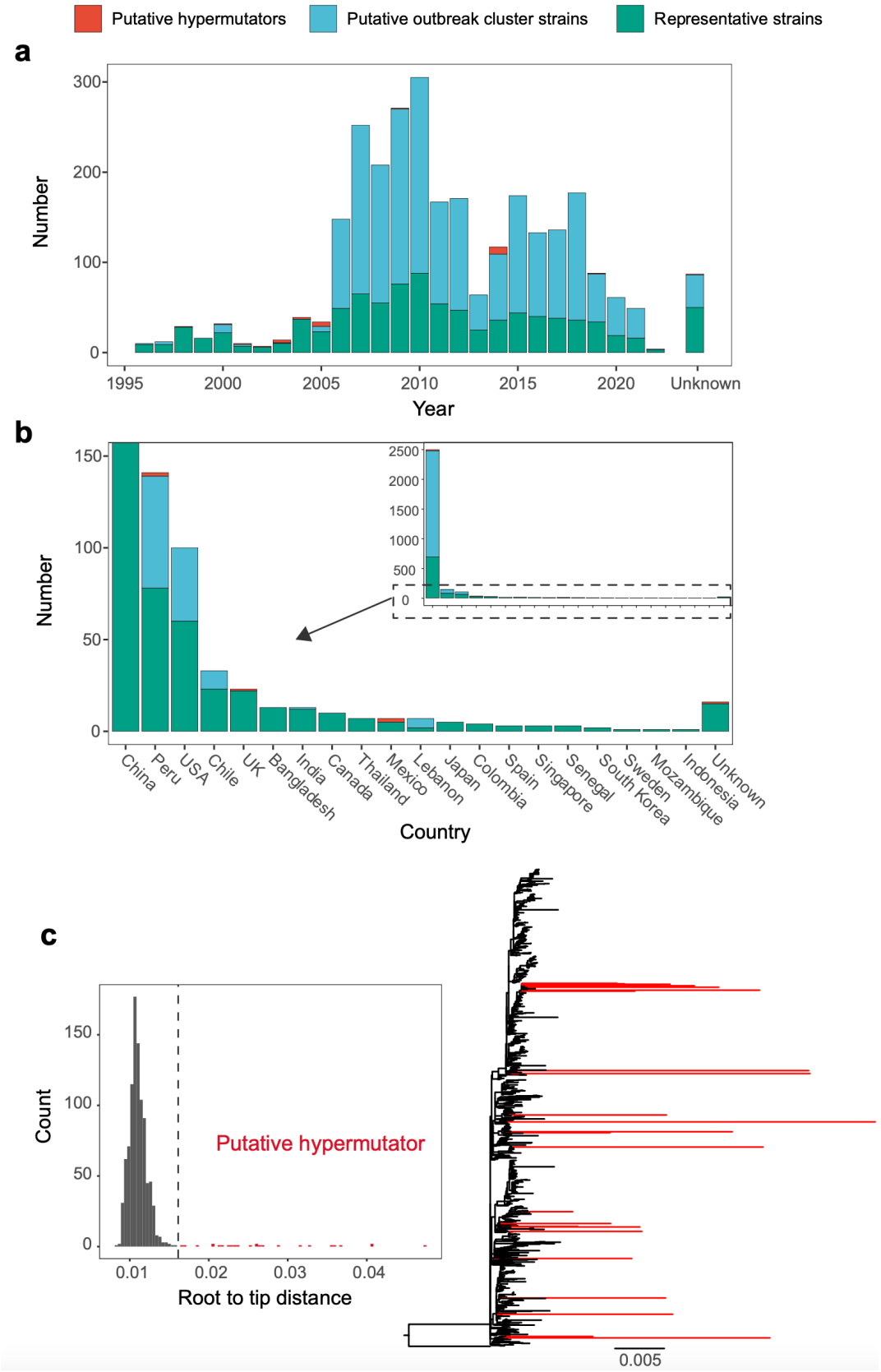
Extended Data Figure 1. Phylogenetic tree of pre-pandemic (pre-PC) strains and sequence coverage of known pandemic clone (PC) marker genes/islands. The colors on the right indicate BLASTN coverages of genes/islands.



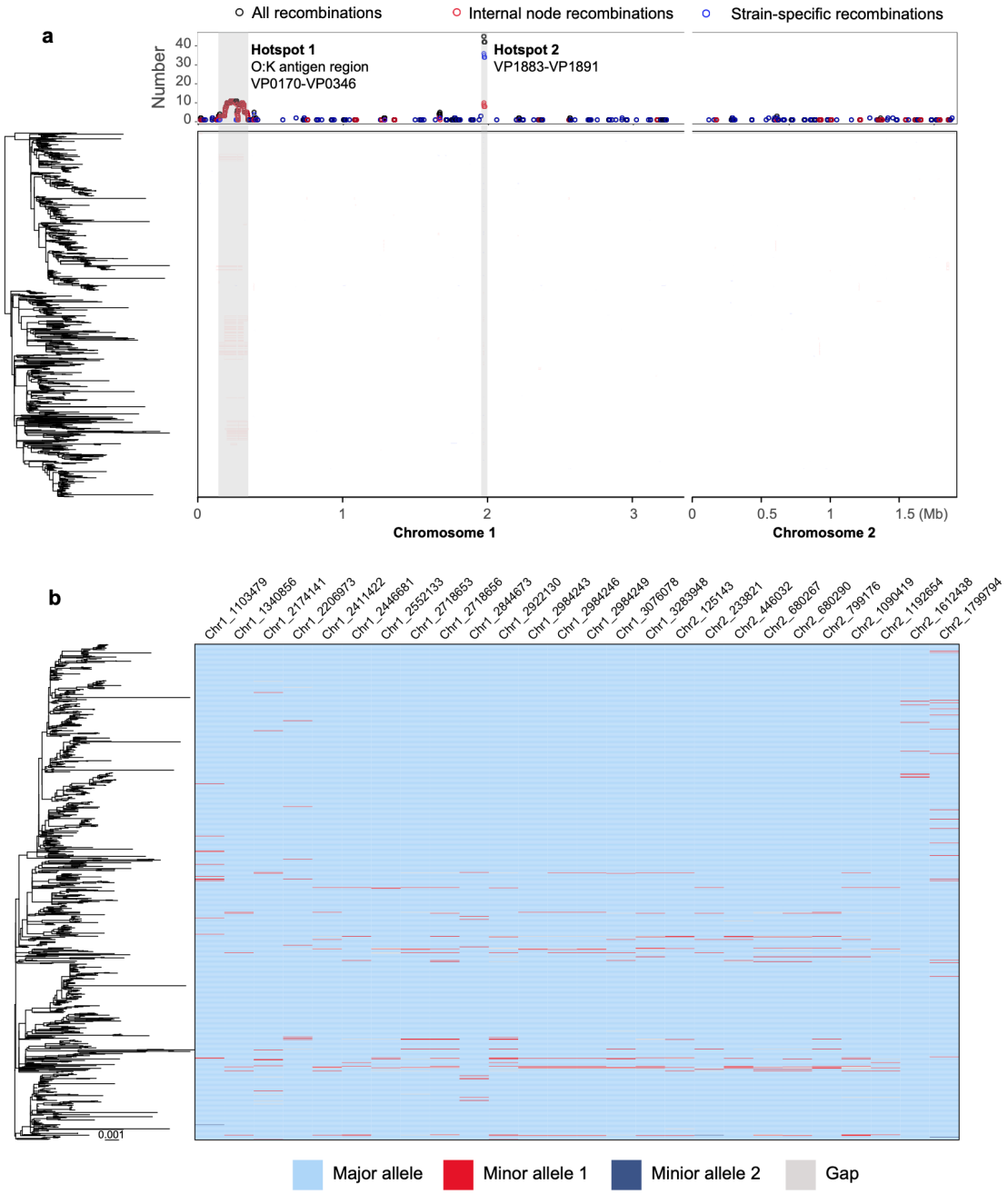
Extended Data Figure 2. Phylogenetic distribution of known pandemic clone (PC) marker genes/islands in PC, pre-PC and non-PC strains. The colors indicate the strains classification.



Extended Data Figure 3. The temporal and geographical distribution of pandemic clone (PC) strains (a,b) and root-to-tip distance and phylogenetic branch lengths of putative hypermutators (c).

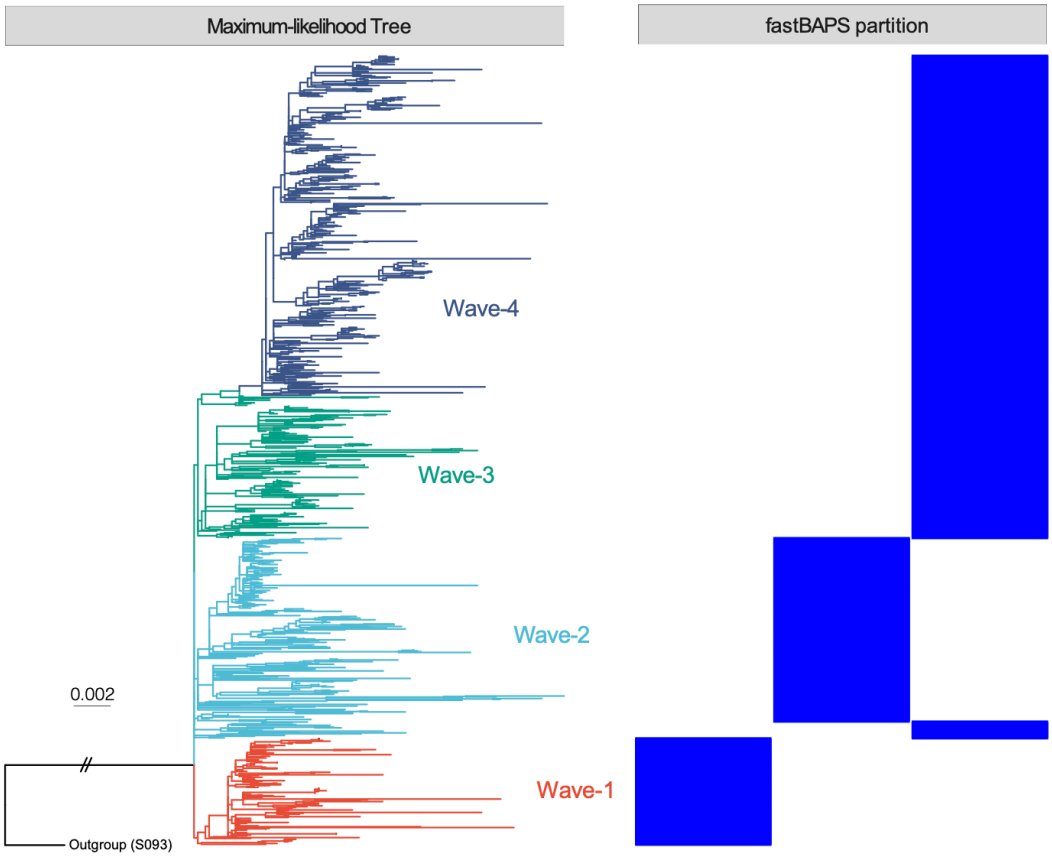


13 **Extended Data Figure 4. Distribution of homologous recombination events (a,b) and**
14 **homoplastic SNP in pandemic clone strains (c).**



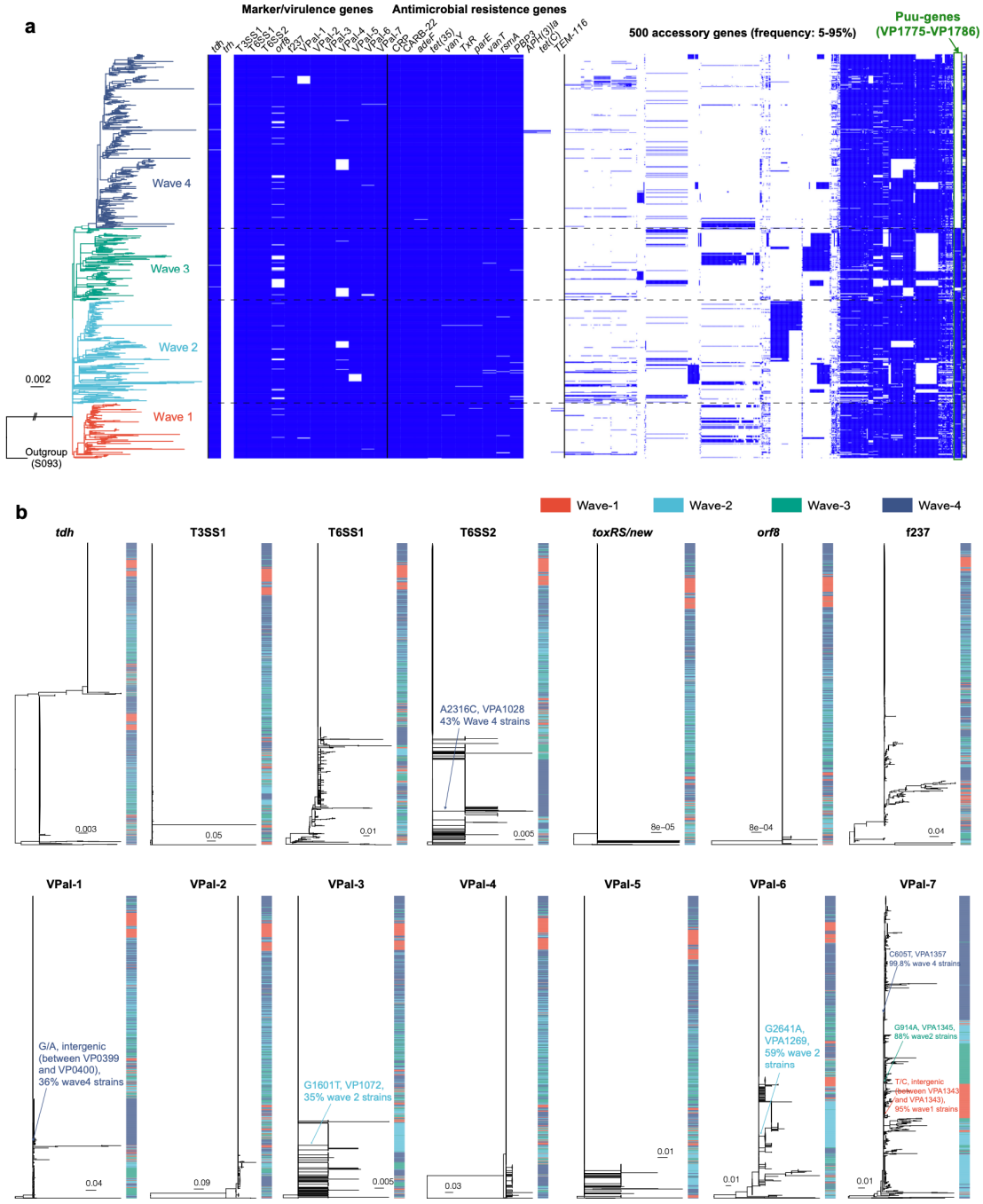
15
16

17 **Extended Data Figure 5. Maximum likelihood tree (left) and fastBAPS hierarchical**
18 **clusters (right) of pandemic clone representative strains.**



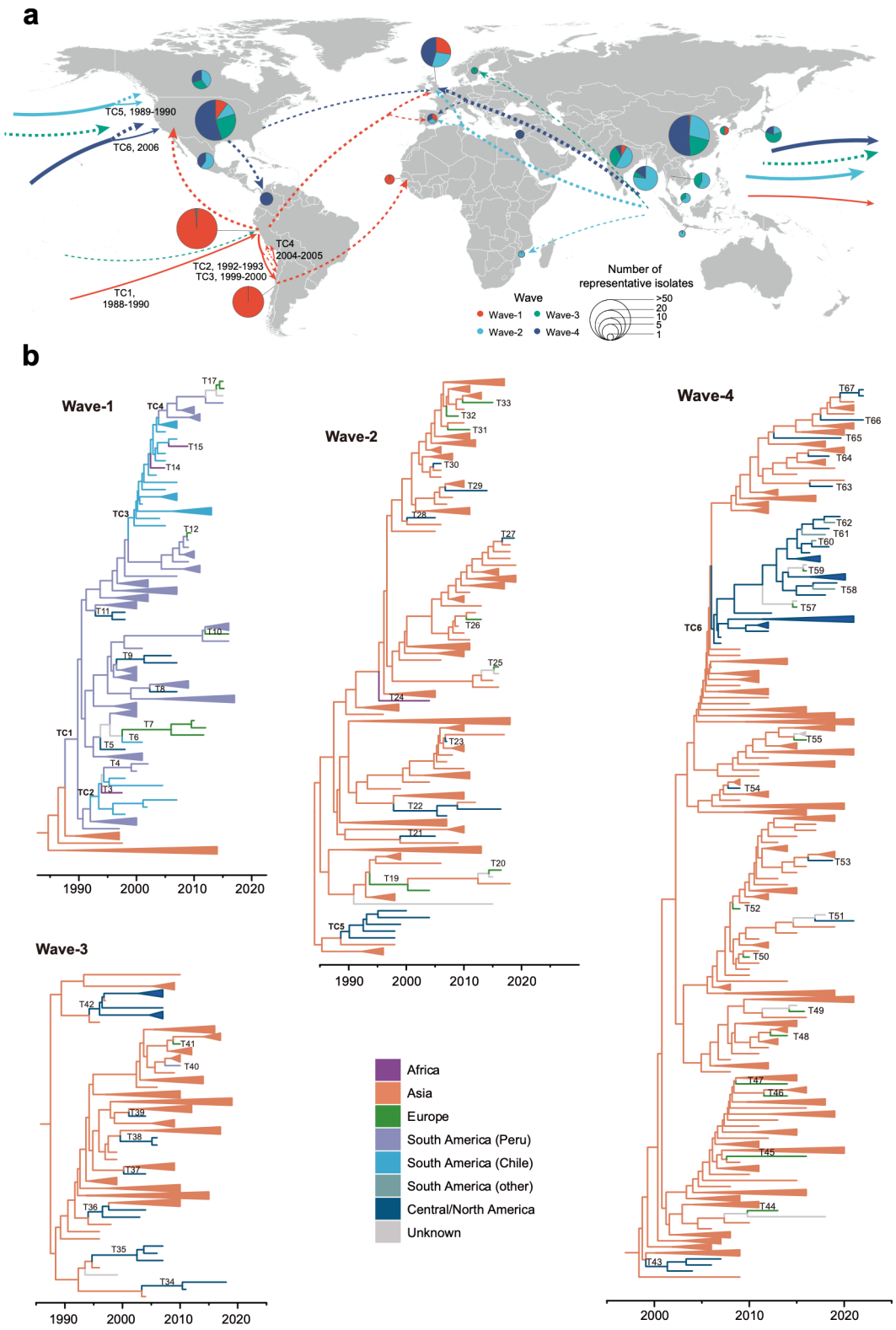
19
20

21 **Extended Data Figure 6. Distribution of known pandemic clone (PC) marker/virulence**
22 **gene, antimicrobial resipiscence genes and accessory genes (a) and phylogenetic clustering**
23 **of PC marker/virulence genes (b).**

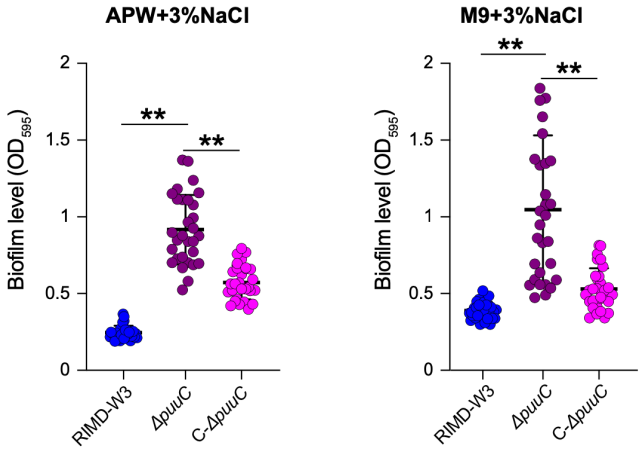


24
25

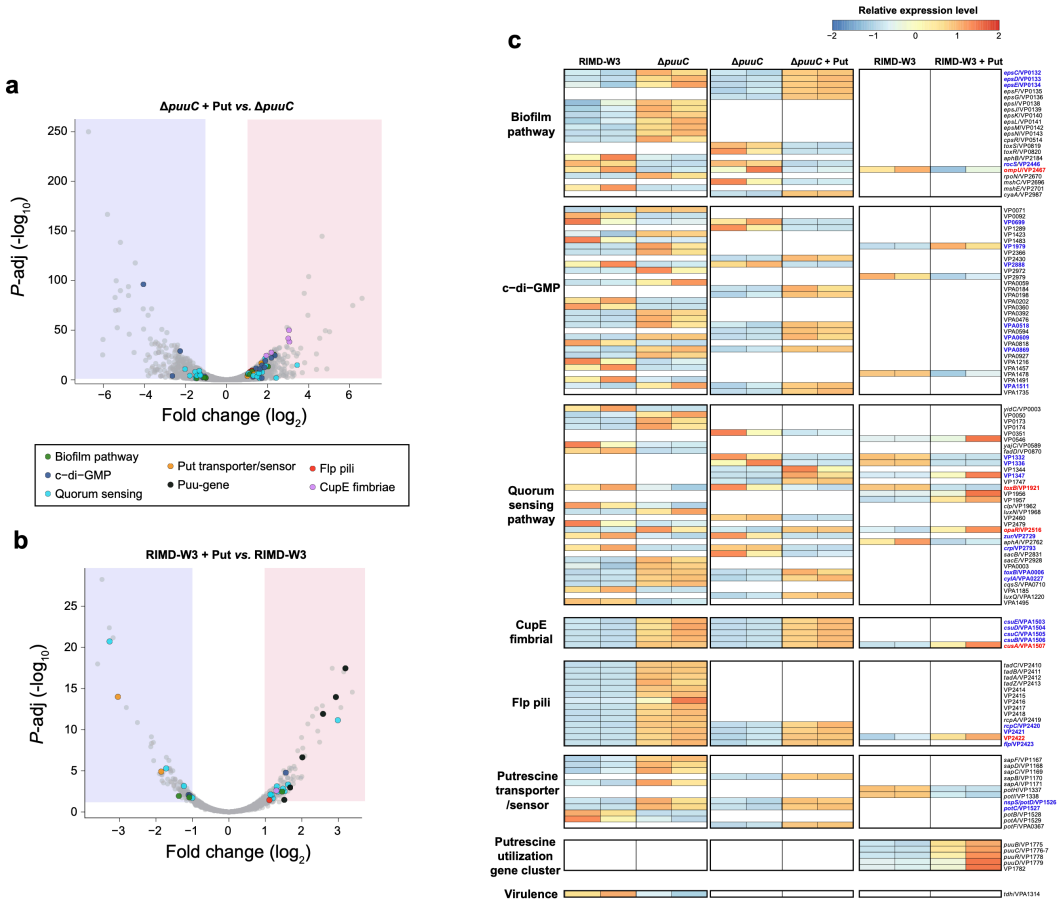
Extended Data Figure 7. Global transmission events of pandemic clone. a) Summary of global transmission events. Dashed lines represent transient transmission not leading to local colonization. Solid lines represent transmissions resulting in local colonization (TC). The colors indicate different waves. Pie charts indicate the composition of waves of countries. The size of circles scale with the number of representative strains. b) Transmission events of different waves. The branch colors indicate the geographical regions of strains.



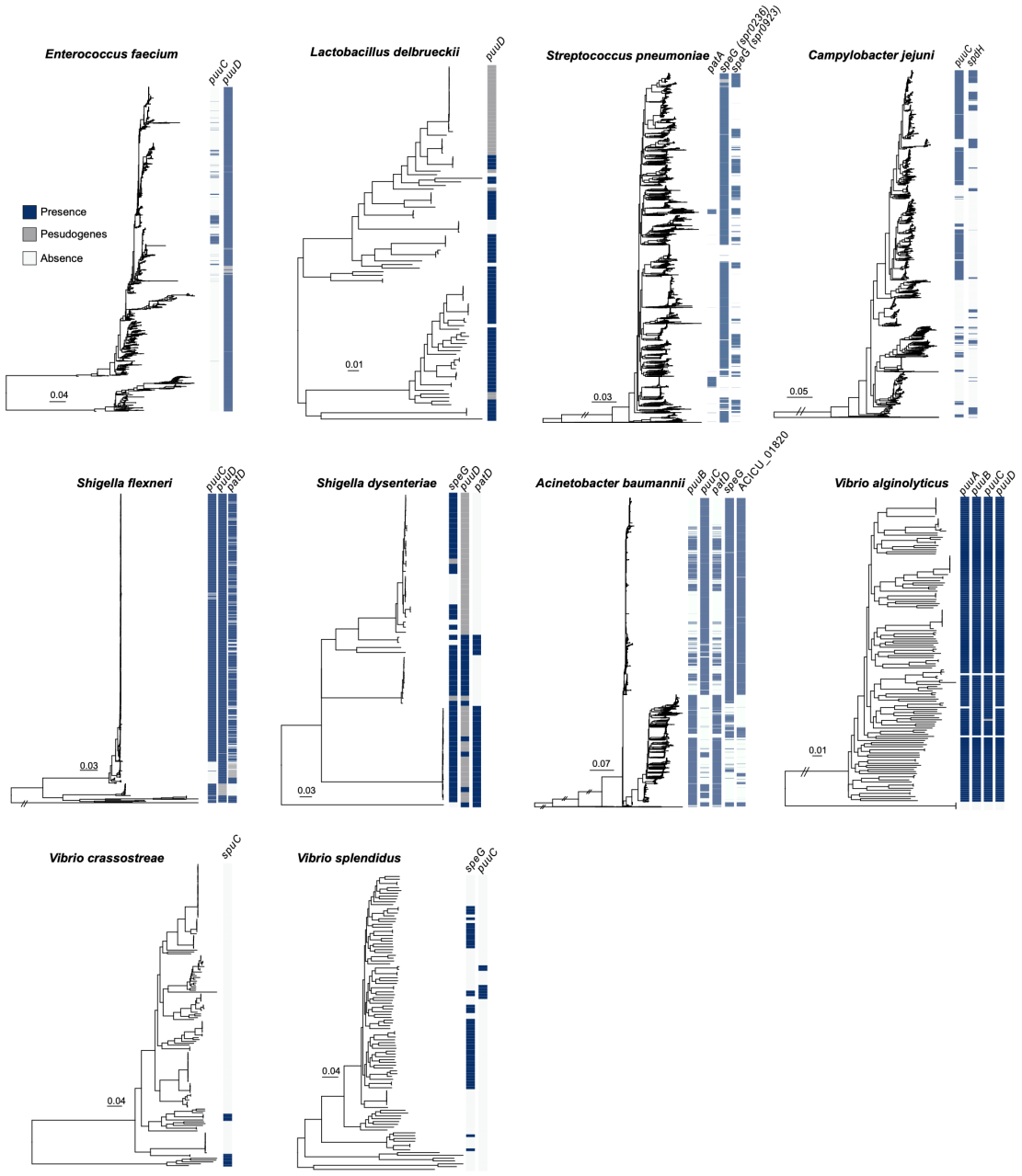
Extended Data Figure 8. Biofilm level of wildtype (RIMD-W3), mutants (Δ*puuA* and Δ*puuC*) and complements (C-Δ*puuA* and C-Δ*puuC*) strains under rich- (alkaline peptone water, APW) or low-nutrition (M9) media with high-salt (3%) concentration.



Extended Data Figure 9. Volcano plot (a,b) and heatmap (c) showing differentially expressed genes (DEGs) between wildtype strain or mutants with (Put+) and without (Put-) exogenous putrescine. Genes colored in red and blue indicate DEGs shared by two or three comparisons, respectively.



Extended Data Fig. 10. Phylogenetic trees and distribution of Puu-genes in 10 bacterial species.



Extended Data Tables

Extended Data Table 1. Wave-4-specific genomic variations.

a) Wave-4-specific SNPs.

Allele composition				Position	SNP alleles	Type	Related Gene	Function of related gene
Wave1	Wave2	Wave3	Wave4					
T(100%);	T(100%);	C(1%);T(99%);	C(93%);T(7%);	156938	T/C	Nonsynonymous	VP0147	amidophosphoribosyltransferase
G(100%);	G(100%);	G(99%);T(1%);	G(7%);T(93%);	300131	G/T	Nonsynonymous	VP0305	DUF1107 domain-containing protein
G(100%);	-(1%);G(99%);	G(99%);T(1%);	-(0%);T(100%);	377670	G/T	Nonsynonymous	VP0376	EAL domain-containing protein
G(100%);	G(100%);	A(1%);G(99%);	A(100%);	449471	G/A	Synonymous	VP0449	penicillin-binding protein activator
G(100%);	G(100%);	G(100%);	T(100%);	517530	G/T	Intergenic	VP0502_intergenic_VP0503	DUF3545 family protein(VP0502); sodium:alanine symporter family protein(VP0503)
C(100%);	C(100%);	C(99%);T(1%);	-(0%);T(100%);	653250	C/T	Nonsynonymous	VP0623	D-amino acid dehydrogenase
T(100%);	T(100%);	T(100%);	C(96%);T(4%);	949025	T/C	Nonsynonymous	VP0914	sigma-54 dependent transcriptional regulator
C(100%);	C(100%);	C(98%);T(2%);	T(100%);	1050003	C/T	Synonymous	VP1002	ImpA family metalloprotease
C(100%);	C(100%);	C(98%);T(2%);	T(100%);	1204871	C/T	Synonymous	VP1147	EAL domain-containing protein
G(100%);	G(100%);	G(98%);T(2%);	T(100%);	1331671	G/T	Nonsynonymous	VP1259	DEAD/DEAH box helicase
-(1%);G(99%);	G(100%);	G(100%);	-(2%);A(98%);	1824857	G/A	Synonymous	VP1712	sensor histidine kinase
-(2%);C(98%);	-(5%);C(95%);	-(2%);C(96%);T(2%);	T(100%);	1905951	C/T	Nonsynonymous	VP1794	GNAT family N-acetyltransferase
T(100%);	T(100%);	G(1%);T(99%);	G(100%);	2103884	T/G	Nonsynonymous	VP2009	response regulator
G(100%);	G(100%);	A(1%);G(99%);	A(100%);	2593552	G/A	Intergenic	VP2469_intergenic_VP2470	magnesium transporter(VP2469);tyrosine--tRNA ligase(VP2470)
C(100%);	C(100%);	C(98%);T(2%);	T(100%);	2891192	C/T	Nonsynonymous	VP2722	NADPH flavoprotein subunit
C(100%);	C(100%);	C(100%);	T(100%);	2924872	C/T	Synonymous	VP2759	N-acetyl-gamma-glutamyl-phosphate reductase
C(100%);	C(100%);	C(100%);	T(100%);	3060071	C/T	Synonymous	VP2888	sensor domain-containing diguanylate cyclase
C(100%);	C(100%);	C(99%);G(1%);	C(0%);G(100%);	3615671	C/G	Nonsynonymous	VPA0333	MFS transporter
G(100%);	G(100%);	A(7%);G(93%);	A(100%);	3623698	G/A	Synonymous	VPA0344	efflux RND transporter permease subunit
G(100%);	-(0%);G(100%);	A(1%);G(99%);	A(100%);	3697691	G/A	Nonsynonymous	VPA0414	hypothetical protein
A(100%);	A(100%);	A(100%);	G(100%);	3708431	A/G	Nonsynonymous	VPA0427	heme anaerobic degradation radical SAM methyltransferase ChuW/HutW
T(100%);	T(100%);	T(100%);	G(100%);	3873928	T/G	Synonymous	VPA0587	pyroglutamyl-peptidase I
G(100%);	G(100%);	G(99%);T(1%);	T(100%);	3998728	G/T	Synonymous	VPA0684	anaerobic sulfatase maturase
T(100%);	-(0%);T(100%);	-(1%);A(1%);T(98%);	-(0%);A(100%);	4259092	T/A	Intergenic	VPA0938_intergenic_VPA0939	AraC family transcriptional regulator(VPA0938); MATE family efflux transporter(VPA0939)
-(1%);A(99%);	-(0%);A(100%);	A(98%);C(2%);	-(0%);A(1%);C(99%);	4292802	A/C	Nonsynonymous	VPA0961	LysR family transcriptional regulator
-(1%);G(99%);	G(100%);	C(2%);G(98%);	C(100%);	4417739	G/C	Nonsynonymous	VPA1069	diguanylate cyclase
-(1%);T(99%);	T(100%);	A(1%);T(99%);	A(100%);	4443534	T/A	Synonymous	VPA1095	hypothetical protein
G(100%);	G(100%);	A(1%);G(99%);	-(0%);A(100%);	4719056	G/A	Nonsynonymous	VPA1357	hypothetical protein
C(100%);	C(100%);	C(99%);T(1%);	C(5%);T(95%);	5108626	C/T	Nonsynonymous	VPA1701	FCD domain-containing protein

b) Wave-4-specific accessory genes.

Gene presence frequency (%)				GeneLocusTag	GeneLocusTag old	Gene	Gene function
Wave1	Wave2	Wave3	Wave4				
100	99.2	95.3	0	VP_RS08535	VP1775	*puuB*	FAD-binding oxidoreductase
100	99.2	95.3	0	VP_RS08540	VP1776-7	*puuC*	aldehyde dehydrogenase
99.2	99.2	95.3	0	VP_RS08545	VP1778	*puuR*	HTH-type transcriptional regulator PuuR
99.2	99.2	95.3	0	VP_RS08550	VP1779	*puuD*	gamma-glutamyl-gamma-aminobutyrate hydrolase family protein

99.2	98.7	95.3	0	VP_RS08555	VP1781	*puuA*	glutamine synthetase family protein
98.4	98.7	94.7	0	VP_RS08560	VP1782		sodium:proton antiporter
98.4	98.3	94.7	0	VP_RS08565	VP1783		hypothetical protein
98.4	97.1	94.1	0	VP_RS08570	VP1784/5		DUF3291 domain-containing protein
98.4	95.4	93.5	0	VP_RS08575	VP1786		hypothetical protein

c) Wave-4-specific unitigs.

Unitig presence frequency (%)				Unitig sequences	Reference genome accession	Mapped position start	Mapped position end	Variation type	Related genes/intergenic	Function of related genes
Wave -1	Wave -2	Wave -3	Wave -4							
100	98.3	85.3	7.1	GGTTTATTGGCTTTGGCTTTGGCTTTGGCTTTGGCT	NC_004603.1	1057835	1057800	Indels (6bp)	VP1010	pseudouridine_synthase
100	98.3	85.3	7.1	TTGGCTTTGGCTTTGGCTTTGGCTTTGGCTGTGCGA	NC_004603.1	1057841	1057806	Indels (6bp)	VP1010	pseudouridine_synthase
0	0	0.6	93.3	AAGATTTTCCGAGAAAGCACATAACGCCCAGTTAAGGGGTGAAGCACG	CP031781.1	1323943	1323990	Indel (6bp)	D0853_06625	DUF3265 domain-containing protein
92.9	95.8	89.4	0	CATCACTGAGCAATTGTTCTTTGCCAGTTGGTTTTGTGCTTTTGACTTTTTTGTCTGTCAG	CP031781.1	2588397	2588457	SNP:C2588427T, nonsynonymous	D0853_13065	DUF4102 domain-containing protein
0	0	1.8	95.6	CTGACAGACAAAAAAGTCAAAAGCACAAAATCAACTGGCAAAGAACAATTGCTCAGTGATG	CP031781.1	2588397	2588457	SNP:T2588427C, nonsynonymous	D0853_13065	DUF4102 domain-containing protein
92.9	95.4	90	0	TGCAACTCTAGCGGTTCTTCGCTGTGTAGAAGTAAGTAAAGTTCT	CP031781.1	2589628	2589672	SNP:G2589658A, intergenic	D0853_13065 -Intergenic- D0853_13070	DUF4102 domain-containing protein (D0853_13065); EAL domain-containing protein (D0853_13070)
0	0.4	1.2	95.6	AGAACTTTACTTACTTCTACACAGCGAAGAGCCGCTAGAGTTGCAGATCAGCCTTGGGCGA	CP031781.1	2589628	2589688	SNP:G2589658A, intergenic	D0853_13065 -Intergenic- D0853_13070	DUF4102 domain-containing protein (D0853_13065); EAL domain-containing protein (D0853_13070)
92.9	90.4	90	0	TCGCCCAAGGCTGATCTGCAACTCTAGCGGTTCTTCGCTGTGTAGA	CP031781.1	2589643	2589688	SNP:G2589658A, intergenic	D0853_13065 -Intergenic- D0853_13070	DUF4102 domain-containing protein (D0853_13065); EAL domain-containing protein (D0853_13070)
92.9	95.4	89.4	0	GCTTGATAATGATCTCTATTTGTTGATGGTATCACATTTTTTAGCCAAATTGACGAGTATC	CP031781.1	2589863	2589923	SNP:C2589893T, intergenic	D0853_13065 -Intergenic- D0853_13070	DUF4102 domain-containing protein (D0853_13065); EAL domain-containing protein (D0853_13070)
0	0	1.8	94.8	GATACTCGTCAATTTGGCTAAAAAATGTGACACCATCAACAAATAGAGATCATTATCAAGC	CP031781.1	2589863	2589923	SNP:C2589893T, intergenic	D0853_13065 -Intergenic- D0853_13070	DUF4102 domain-containing protein (D0853_13065); EAL domain-containing protein (D0853_13070)
92.9	94.1	89.4	0	AGATACAAAGCTTCCGGTATCTTTTGTTTTTAAATTTACTTTTATCTTT	CP031781.1	2590927	2590975	SNP:G2590957T, nonsynonymous	D0853_13075	AraC family transcriptional regulator
0	0	1.2	94.8	AAAGATAAAAGTAAATTTAAAAACAAAAGAGACCGGAAGCTTTGTATCTAAGTCAAAACAT	CP031781.1	2590927	2590987	SNP:G2590957T, nonsynonymous	D0853_13075	AraC family transcriptional regulator
92.9	95	89.4	0	ATGTTTTGACTTAGATACAAAGCTTCCGGTATCTTT	CP031781.1	2590952	2590987	SNP:G2590957T, nonsynonymous	D0853_13075	AraC family transcriptional regulator

92.9	95.8	89.4	0	TGTGTTCTAGTCAAAAGGTGTCCATTTTAAATT	CP031781.1	2591664	2591693	SNP:T2591694A, intergenic	D0853_13075 -Intergenic- D0853_13080	AraC family transcriptional regulator (D0853_13075); hypothetical protein (D0853_13080)
0	0	1.8	94.8	AATTTAAAATGGACACCTTTTGACTAGAACTCAATTAATAGTCAACTAGTGTTGTGTTGGGTTT	CP031781.1	2591664	2591727	SNP:T2591694A, intergenic & C2591706T, intergenic	D0853_13075 -Intergenic- D0853_13080	AraC family transcriptional regulator (D0853_13075); hypothetical protein (D0853_13080)
92.9	95.4	89.4	0	GTTTTAAGAAAACCCAACACAACACTAGTTAACTATTAATTGTGTTCTAGTCAAAAGGTGTCCATTTTAA	CP031781.1	2591667	2591736	SNP:T2591694A, intergenic & C2591706T, intergenic	D0853_13075 -Intergenic- D0853_13080	AraC family transcriptional regulator (D0853_13075); hypothetical protein (D0853_13080)
0	0.4	1.8	94.8	TTAATAGTCAACTAGTGTTGTGTTGGGTTTTCTTAAAAC	CP031781.1	2591698	2591736	SNP:C2591706T, intergenci	D0853_13075 -Intergenic- D0853_13080	AraC family transcriptional regulator (D0853_13075); hypothetical protein (D0853_13080)
0	0.4	1.2	95.3	CCAAAAAGGATAATTCGCAAGAAACCTCATGTAAGACGAGAATGTGAGGTGTTTCGAGATG	CP031781.1	2593590	2593650	SNP:G2593620A, nonsynonymous	D0853_13090	hypothetical protein
91.3	95.4	90	0	CATCTCGAAACACCTCACATTCTCGTCTTATATGAGGTTTCTTGCGAATTATCCTTTTTGG	CP031781.1	2593590	2593650	SNP:G2593620A, nonsynonymous	D0853_13090	hypothetical protein

Extended Data Table 2. Homolog genes of Puu-genes in 36 bacterial species.

Sub-pathway	**SP1**			**SP2**		**SP2/3**	**SP3**			**SP4**		**SP5**		**SP6**	
EC/Gene	**EC:6.3.1.11**	**EC:1.4.3.-**	**EC:1.2.1.19**	**EC:3.5.1.94**	**EC:2.3.1.57**	**EC:1.4.3.4**	**EC:1.2.1.3**	**EC:2.6.1.82**	**EC:1.2.1.19**	**EC:2.6.1.113**	**EC:1.2.1.54**	**EC:1.5.1.43**	**EC:4.1.1.96**	**EC:2.5.1.16**	**EC:1.5.99.6**
Gene	***puuA***	***puuB***	***puuC***	***puuD***	***speG***		***puuC***	***patA***	***patD/puuC***	***spuC***	***puuC***			***speE***	***spdH***
Escherichia coli	b1297	b1301	b1300	b1298	b1584		b2661	b3073	b1444	b3073	b1300			b0121	
Shigella boydii	SBO_1765	SBO_1761	SBO_1762	CUC36_02800	CUC36_01690		SBO_4411	SBO_2932	SBO_3588	SBO_2932	SBO_1762			SBO_0110	
Shigella dysenteriae	SDY_1947	SDY_1943	GH770_23095	C5S64_RS19865	C5S64_RS18920		GH770_23095	GH770_14085	SDY_1730	GH770_14085	GH770_23095			SDY_0028	
Shigella flexneri	SFyv_1927	SFyv_1933	SFyv_1932	SFyv_1930	SFyv_2338		SFyv_3896	SFyv_4521	SFyv_2596	SFyv_4521	SFyv_1932			SFyv_0158	
Shigella sonnei	SSON_1843	SSON_1839	SSON_1840	SSON_1842	HUX90_20125		SSON_2805	SSON_3211	SSON_1693	SSON_3211	SSON_1840			SSON_0129	
Vibrio parahaemolyticus	VP1781	VP1775	VP1776-7	VP1779	VPA0474		VP1776-7	VP2797	VP1776-7	VP0630	VP1776-7	VP1941	VP1940		
Vibrio alginolyticus	N646_0402	N646_0406	N646_0405	N646_0403	N646_4685		N646_3146	N646_1889	N646_3146	N646_3797	N646_3146	N646_1023	N646_1022		
Vibrio vulnificus		VVMO6_03438	VVMO6_04237				VVMO6_04237	VVMO6_00258	VVMO6_04237	VVMO6_03430	VVMO6_04237	VVMO6_01981	VVMO6_01980	FOB71_05510	
Vibrio cholerae			VC0819		VCA0947		VC0819	VC2618	VC0819	VCA0605	VC0819	VC1624	VC1623		
Vibrio splendidus		LTQ54_23680	LTQ54_17875	LTQ54_11215	LTQ54_17570		LTQ54_17875	LTQ54_10125	LTQ54_17875	LTQ54_03365	LTQ54_17875	LTQ54_14800	LTQ54_14805		
Vibrio crassostreae	EDB58_103383	IS519_18960	IS519_02870	IS519_13010	IS519_03160		IS519_02870	IS519_11840	IS519_02870	IS519_17355	IS519_02870	IS519_00195	IS519_00200		

Pseudomonas aeruginosa	PA0296	PA1565/PA2317/PA2776/PA5309	PA0219/PA4189	PA0297	PA4114	PA0421	PA4899	PA0895	PA5373	PA0221/PA0299	PA5312			PA1687/PA4774	PA3713
Enterococcus faecium	EFAU004_00692		EFAU004_00239	EFAU004_02102	WQ1_02180		FBF65_00190				FBF65_00190				
Enterococcus faecalis	EF62_2460		EF62_0622	EF62_1589	EGM183_03505		EF62_0622	I5Q89_03735	EF62_0622	I5Q89_03735	EF62_0622				
Klebsiella pneumoniae	KP1_2006	KP1_1462/KP1_2001	KP1_2002	KP1_2005	KP1_2606		KP1_1995	KP1_4798	KP1_2989	KP1_1518	KP1_2002			KP1_0952	
Staphylococcus aureus	SA1150		SA0162		SA0330		SA0162/SA1924	SA0179	SA2406	SA2214	SA0162				
Streptococcus pneumoniae	spr0444		spr1028	spr0997	spr0236/spr0923		spr1028	DV131_RS08170	spr1028	DV131_RS08170	spr1028	spr0820	spr0821	spr0819	
Streptococcus pyogenes	spyM18_1942		spyM18_1383	spyM18_1080			spyM18_1383	E9497_RS07210	spyM18_1041	E9497_RS07210	spyM18_1383				
Salmonella enterica	t3614	t3050	t3839	t2854	t1423		t2686	t3138	t1506	t3138	t3839			t0171	
Burkholderia pseudomallei	BPSL2336	BPSS0461/BPSS1470	BPSS0473	BPSS0470	BPSL0096		BPSS0280/BPSL1020/BPSL1550	BPSS0869	BPSS0307	BPSL2335/BPSS0468	BPSS0473			BPSS1453	BPSL2954/BPSS1151
Yersinia enterocolitica	YE0027		YE3793	YE0746	YE1124		YE3793	YE3957	YE3793	YE2908	YE3793			YE0710	
Yersinia pestis	YPO0024		YPO1166	YPO3377	YPO2830		YPO1166	YPO2844	YPO1166	YPO1150	YPO1166			YPO3411	
Yersinia pseudotuberculosis	YPTB0024		YPTB1196	YPTB0754	YPTB2797		YPTB1196	YPTB3731	YPTB1196	YPTB1181	YPTB1196			YPTB0720	
Helicobacter pylori	HP0512		HP0056	HP0349			HP0056	HP0976	HP0056	HP0976	HP0056	HP1507	HP0020	HP0832	
Campylobacter coli	AV440_RS02380		G157_06195	G157_00130	HXP31_06725		G157_06195	G157_07570	G157_06195	G157_07130	G157_06195	G157_07815	G157_01310		
Campylobacter concisus	CVT08_04025		CCON33237_0809				CCON33237_0809	CCON33237_0597	CCON33237_0809	CCON33237_0597	CCON33237_0809	CCON33237_0315	CCON33237_0326		
Campylobacter fetus	CFF8240_1054		CFF8240_0070	CFF8240_1754			CFF8240_0070	CFF8240_0925	CFF8240_0070	CFF8240_0513	CFF8240_0070	CFF8240_0285	CFF8240_0338		
Campylobacter jejuni	CJSA_0663		CJSA_0461	CJSA_0027			CJSA_0461	CJSA_0204	CJSA_0461	CJSA_0281	CJSA_0461	CJSA_0162	CJSA_1436		CJSA_1098
Acinetobacter baumannii	ACICU_02336	ACICU_01128	ACICU_01364	ACICU_02009	ACICU_01628	ACICU_01820	ACICU_02609	ACICU_02660	ACICU_01133	ACICU_02439	ACICU_02298				
Akkermansia muciniphila			Amuc_2066	Amuc_0170			Amuc_2066	Amuc_0493	Amuc_2066	Amuc_0493	Amuc_2066	Amuc_0148	Amuc_1657		
Bifidobacterium bifidum	BBIF_0556		BBIF_1321	BBIF_0842			BBIF_1321	BBIF_1100	BBIF_1321	BBIF_1100	BBIF_1321				
Bifidobacterium breve	HMPREF9228_0590	HMPREF9228_1307	HMPREF9228_0974			HMPREF9228_1307	HMPREF9228_1180	HMPREF9228_1307	HMPREF9228_1180	HMPREF9228_1307					
Bifidobacterium longum	BLD_0159		BLD_0881	BLD_1017			BLD_0881	BLD_0813	BLD_0881	BLD_1417	BLD_0881			BLD_0231	
Lactobacillus crispatus	EC605_RS07450		EC605_RS01585	EC605_RS03085			EC605_RS01585		EC605_RS01585		EC605_RS01585				
Lactobacillus delbrueckii	Ldb1472		Ldb1179	Ldb2120	Ldb1883		Ldb1179		Ldb1179		Ldb1179				

Lactobacillus helveticus	lhv_1558	lhv_1739	GFB61_10560	lhv_1739	lhv_1739	lhv_1739