Title

Fitness factors impacting survival of a subsurface bacterium in contaminated groundwater

Journal

The ISME Journal: Multidisciplinary Journal of Microbial Ecology, 18(1)

ISSN

1751-7362

Authors

Thorgersen, Michael P

Goff, Jennifer L

Trotter, Valentine V

et al.

Publication Date

2024-01-08

DOI

10.1093/ismejo/wrae176

The ISME Journal

Fitness Factors Impacting Survival of a Subsurface Bacterium in Contaminated Groundwater

Michael P. Thorgersen[1], Jennifer L. Goff[1], Valentine V. Trotter[2], Farris L. Poole II[1], Adam P. Arkin[2,3], Adam M. Deutschbauer[2,4], and Michael W. W. Adams[1, †]

[1]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA

[2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[3]Department of Bioengineering, University of California, Berkeley, Berkeley, CA, USA

[4]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA

[†]Address correspondence to adamsm@uga.edu

Corresponding Author:

Michael W. W. Adams

(ORCID ID: 0000-0002-9796-5014)

Department of Biochemistry & Molecular Biology

University of Georgia, Athens, GA 30602, USA

Email: adamsm@uga.edu. Phone: 706 542 2060

Running title: Fitness Factors Impacting Groundwater Survival

Abstract

Many factors contribute to the ability of a microbial species to persist when encountering complexly contaminated environments including time of exposure, the nature and concentration of contaminants, availability of nutritional resources, and possession of a combination of appropriate molecular mechanisms needed for survival. Herein we sought to identify genes that are most important for survival of Gram-negative *Enterobacteriaceae* in contaminated groundwater environments containing high concentrations of nitrate and metals using the metal-tolerant Oak Ridge Reservation (ORR) isolate, *Pantoe*a sp. MT58 (MT58). Survival fitness experiments in which a randomly barcoded transposon insertion (RB-TnSeq) library of MT58 was exposed directly to contaminated ORR groundwater samples from across a nitrate and mixed metal contamination plume were used to identify genes important for survival with increasing exposure times and concentrations of contaminants, and availability of a carbon source. Genes involved in controlling and using carbon, encoding transcriptional regulators, and related to Gram-negative outer membrane processes were among those found to be important for survival in contaminated ORR groundwater. A comparative genomics analysis of 75 *Pantoea* genus strains allowed us to further separate the survival determinants into core and non-core

genes in the *Pantoea* pangenome, revealing insights into the survival of subsurface microorganisms during contaminant plume intrusion.

Key words: *Pantoea*, pangenome, survival, fitness, outer membrane, NADH dehydrogenase

Introduction

Subsurface microbial communities are drivers of key steps in global elemental cycles (1). Anthropogenic contamination of subsurface environments with nitrate and various metals can disrupt these microbial communities resulting in loss of species and functional diversity disrupting normal biogeochemistry cycles that in turn hinders natural recovery of the contaminated site (2, 3). Investigations into anthropogenically altered microbial communities often focus on metagenomic data to identify key genomic determinants that enable microorganisms to survive in the contaminated environment (4, 5). These methods, however, do not consider phenotypic data from microorganisms exposed to the contaminated environment and provide no direct insights into the molecular mechanisms needed for microbes to persist within the altered environment.

The contamination plume extending from the former S-3 ponds at Oak Ridge Reservation (ORR) in Bear Creek Valley (TN, USA) is an anthropogenically altered site where nitrate and mixed metal contamination affect the subsurface microbial community structure and function (Fig. 1A) (6). From 1951 to 1983, waste, primarily from uranium processing at the Y-12 Plant and consisting of uranium and various other metals dissolved in nitric acid (pH < 2), were deposited in the S-3 ponds (6, 7). Despite cleanup efforts, contamination from the ponds seeped into the surrounding environment forming several overlapping plumes (7, 8). The microbial community structure within these plumes is highly correlated to multiple contamination related geochemical features, including low pH and high concentrations of metals, including Ni,

Co, Mn, Cd, and Pb, as well as U (4, 9). ORR is a model ecosystem for studying how gradients of environmental stressors impact the composition and function of subsurface microbial communities.

The Gram-negative *Enterobacteriaceae* family is important as it is abundant and widely dispersed in natural and host associated environments (10). In particular, the *Pantoea* genus within the *Enterobacteriaceae* consists of versatile and diverse strains with the ability to inhabit wide ranging and dissimilar ecological niches. These include those inhabiting contaminated groundwater and soil, insect symbionts, and important plant pathogens (11). To a large extent, the survival of a microorganism in a particular environment is dependent on the molecular mechanisms it possesses encoded by the genes in its genome. Consequently, comparative genomics have been used to investigate genomic determinants important for enabling various lifestyles of *Pantoea* species, such as genes associated with onion pathogenicity (12) and coding sequences with orthologs restricted to plant, animal and insect hosts (12). In a phylogenomic analysis comparing *Pantoea stewarti* and *Pantoea ananatis* strains, a large core genome of over 3,500 shared protein coding sequences was found that was accompanied by a diverse accessory genome composed of genes located on various mobile genetic elements including plasmids, integrated prophages, and insertion elements (13). The wide array of mobile genetic element types and the resulting accessory genome supported by the core *Pantoea* genome is likely a key factor in the environmental diversity of this genus. Consequently, the goal of this work was to use a groundwater- and sediment-associated *Pantoea* strain to identify genomic determinants that are key for *Enterobacteriaceae* survival in both non-contaminated and contaminated groundwater environments.

MT58 was isolated from a non-contaminated ORR sediment sample (14). Despite the fact that MT58 originated from a non-contaminated region, this strain has a high tolerance for both nitrate and various metals (e.g., Cu, Cd, Co, Ni, Mn, U, and Cr) at contaminated site-

relevant concentrations (14). A 16S rRNA gene (V4 region) amplicon sequence variant (ASV) that matches the MT58 16S rRNA gene sequence was detected in ORR groundwater from across the S-3 ponds contamination plume, including highly contaminated samples with pH < 4 and U concentrations > 50 µM, suggesting that this ORR *Pantoea* lineage or related *Pantoea* species can persist in the presence of these stressors (14). MT58 is genetically tractable (15) and reduces nitrate while growing in the presence of elevated concentrations of various metal mixtures, making this strain a relevant model for understanding microbial persistence during anthropogenic nitrate and metal perturbations.

Herein we used survival fitness experiments in which an RB-TnSeq deletion library (16) of MT58 was directly exposed to environmental groundwater samples taken from across the S-3 ponds contamination plume. The goal was to identify genes that are critical for survival across a range of groundwater challenge conditions, including exposure time, contamination level, and carbon substrate availability. Additionally, we used comparative genomics to investigate whether any of these “survival” genes are unique to strain MT58, are found in a subset of *Pantoea* species or are part of the core makeup of the *Pantoea* pangenome. Together these data allowed us to identify a range of groundwater survival genes from those broadly relevant to Gram-negative *Enterobacteriaceae* to other less widespread accessory genes that are specific to MT58.

Materials and methods

Comparative genomics

Complete *Pantoea* genus genomes were obtained from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (17). The rooted phylogenetic species tree was constructed using the KBase (18) Insert Set of Genomes Into SpeciesTree – v2.2.0 app, which uses FastTree 2.0.0 and a set of 49 core genes to infer maximum-likelihood phylogenies (19). The tree was visualized using Interactive Tree of Life v6.8.1 (20). Pangenomes were constructed using the mOTUpan – v0.3.2 app (21, 22) in KBase (18) with MMseqs2 easy-cluster mode and 80% minimum coverage for orthologs. Pangenome rarefaction and amplification curves were calculated using a previously published R script with R v4.2.1 (23). The pangenome openness parameter (γ) was calculated using the equation $G = cN^{\gamma}$ where G is the pangenome size, c is the core genome size, and N is the number of genomes (24). The estimated full pangenome size (N_1) was calculated using the equation $N_1 = N_{obs} + F_1^2 / 2F_2$ where N_{obs} is the observed pangenome size, F_1 is the number of singletons and F_2 is the number of doubletons in the pangenome (present in only 1 or 2 genomes respectively) (24). MT58 genes and related information were gathered from BV-BRC (17). Genomic island predictions for MT58 were obtained using IslandViewer 4 (25). COG categories and annotations for the MT58 genes were obtained using EggNOG-mapper v2.1.5 (26, 27). Sequence similarity networks for genes of unknown function were constructed using the EFI-Enzyme Similarity Tool and EFI Database v2024_0./100 (28, 29) with the default E-Value of 10^{-5}. The sequence similarity networks were finalized with an alignment score threshold of 35% and by filtering out sequences that deviated from the input sequence by over 20% in length. The sequence similarity networks were visualized using Cytoscape v3.10.2.

Solutions and media

The 1X salts solution contained 4.7 mM NH_4Cl, 1.3 mM KCl, 2 mM $MgSO_4$, 0.2 mM NaCl, and 0.1 mM $CaCl_2$. *Pantoea* minimal medium contained 1X salts solution, 5 mM NaH_2PO_4, 1 mM glucose and a 1X trace elements solution described previously (15). All solutions and media were filter sterilized (0.22 µM) before use.

Survival experiments

The MT58 RB-TnSeq library (16) was grown on Luria-Bertani (LB) broth (30) with 50 µg/mL kanamycin (Kn) (50 mL) aerobically with shaking (150 rpm) at 23°C to an OD_{650} of 0.5. Six reference samples of the LB culture (1 mL) were harvested, and pellets were frozen at -80°C. The remaining LB culture was harvested (10 min, 6,000 x *g*), and the supernatant was decanted. The pellet was washed once with 20 mL of 1X salts and then suspended in 3.5 mL 1X salts. Anoxic (80% N_2, 20% CO_2) survival challenge vials containing 3 mL of either *Pantoea* minimal medium or ORR groundwater (filter sterilized, 0.22 µM) with the indicated additions were prepared in quintuplet and inoculated with 100 µL of washed cell suspension (final OD_{650} of about 0.2). Survival challenge vials were then incubated with gentle rocking at 23°C for the indicated amount of time. After the incubation, 100 µL from each survival challenge vial was used to inoculate 5 mL of LB Kn (50 µg/mL) broth for outgrowth (20 hr) aerobically with shaking (150 rpm) at 23°C). The final OD_{650} of the outgrowth cultures was recorded and 1 mL samples were harvested, and the pellets frozen at -80°C as outgrowth samples.

DNA extraction, sequencing, and fitness analysis

Frozen reference and outgrowth sample pellets were processed for DNA extraction and sequencing using the BarSeq98 method as previously described (16). Briefly, we PCR-amplified the barcodes from the population using a pair of dual indexed primers that contain all of the adapter sequences necessary for sequencing. The primer pairs all have unique dual indexes to

identify instances of index hopping, and the unique indexes allowed us to multiplex hundreds of BarSeq samples on a single lane. After PCR amplification, we mixed these amplicons together in equal volumes, and purified the mixture over a Zymo clean and concentrator 5 column. These PCR amplicons were sequenced using BarSeqV3 primers on the HiSeq2000 platform (Illumina) (31). Strain fitness values, the normalized $\log_2$ ratio of counts between the outgrowth and reference samples, were calculated for each bar-coded strain in the library and gene fitness values, the weighted averages of the strain fitness values for each gene, were calculated as previously described (16). Quality control and normalization of gene fitness values were performed as previously reported (16). For the largest gene fitness values and largest gene fitness value changes observed in the survival fitness experiments, a cutoff of ≥ |1.5| was selected, which was previously seen to highlight significance results for the MT58 RB-TnSeq library (15). In cases where gene fitness changes were monitored over a time, pH or glucose concentration, large fitness changes were additionally filtered by fitting the gene fitness value curves to a second order polynomial and removing genes with poor fits (R^2 value < 0.50).

Viable cell quantification

Viable cells over time were quantified from survival challenge vials prepared as described above. At indicated time points during a challenge, 20 µL of cells were removed from the survival challenge vial and serially diluted 1 to 10 across the columns of a 96 well plate into LB medium. The plate was incubated for 20 hr aerobically at 23ºC with gentle rocking before reading the final OD_{650}. The number of viable cells was then determined from the final OD_{650} of the most serially diluted well with growth by comparing the results to those of similarly treated cultures of MT58 with known viable cell counts. Viable cell counts for the MT58 cultures were determined using traditional dilution plating techniques on LB agar plates grown at 23ºC for 20 hr.

Element and ion quantification

Concentrations of 17 different elements were determined using ICP-MS as described previously (32). Nitrate concentrations were previously determined using a Dionex 2100 system with an AS9 column (U.S. EPA Methods 300.1 and 317.0) (9). Ammonia concentrations were determined using the Amplite Colorimetric Ammonia Quantitation Kit *Blue Color* (AAT Bioquest, Pleasanton, CA, USA).

Results

Survival fitness experimental setup

Five ORR groundwater samples ranging in pH from 3.9 to 7.4 with varying degrees of nitrate and metal contamination from the S-3 ponds contamination source were selected. Sample FW300 (pH 7.4) was from a non-contaminated ORR site located 7,000 m from the S-3 ponds, Whereas the other samples were taken from wells within the contamination plume (20 – 40 m distant) (Fig. 1A). The concentrations of contaminating metals (Fig. S1) and other elements / compounds of interest (Fig. S2) were measured. Although most of the metals tend to decrease in concentration with pH, Mo concentration increases (Fig. S1).

Survival fitness experiments were used to identify molecular mechanisms important for microbial persistence during diverse groundwater challenges (Fig. 1B). These experiments involved the use of a RB-TnSeq library in which transposon mutagenesis with constructs containing random DNA bar codes enable high throughput of fitness profiling experiments compared to traditional TnSeq experiments (16). Each survival fitness experiment was started with MT58 RB-TnSeq library cells harvested from aerobic LB cultures, the growth condition used to construct the RB-TnSeq library (33). Cells were then resuspended directly into the challenge conditions. These included in non-contaminated groundwater, a time course in non-

contaminated groundwater, challenges with groundwater samples containing increasing amounts of contamination, and non-contaminated groundwater samples with increasing concentrations of glucose added as a carbon source (Fig. 1C). All challenge conditions were conducted anoxically at 23°C (original state of the groundwater samples), and all except for the glucose titration experiment contained 1 mM glucose. Additionally, a set of challenges was run using the *Pantoea* minimal medium adjusted to the same pH value as the ORR groundwater samples to evaluate differences in fitness observed with pH changes alone rather than exposure to multiple contaminants. After incubation in the challenge conditions for the indicated time periods, the challenge samples were sub cultured back into the original aerobic LB growth condition for outgrowth and fitness analysis. The resulting fitness changes reflect survival of the library members during the groundwater challenge incubation step. A negative change in gene fitness denotes that the challenge resulted in decreased abundance of library mutants lacking that gene. Survival of the wild-type MT58 strain over time in ORR groundwater with 1 mM glucose added was used to determine limits for the challenge incubation times. Strain MT58 was able to survive without significant decrease in population for 24 hr in non-contaminated FW300 groundwater (pH 7.4). In contrast, a significant decrease in cell viability occurred within 30 min of incubation in the most extremely contaminated FW106 groundwater sample (pH 3.9), with complete loss of viability within 3 hr (Fig. 1D).

Transition to anaerobic carbon metabolism key to groundwater survival

Even under mild survival incubation challenges in FW300 groundwater (pH7.4) for 30 min. (Fig. 1C), 35 genes were important for fitness with fitness values ≥ |1.5| (Table S1). One of the main changes between the survival incubations and the LB growth condition to which they are compared is the switch to anoxic conditions, a necessary consequence of the experimental design. This approach also mimics an important property of groundwater environments which undergo changes in O_2 availability (9, 34). Consistent with the anoxic environment, the *fnr* gene,

which encodes a global transcriptional regulator controlling expression of genes for anaerobic metabolism (35), has a large negative fitness value (-2.3) in the mild challenge. In addition, several genes involved in anaerobic carbon flow that are positively controlled by FNR in *Escherichia coli* also have large negative fitness values (Fig. 1E) (35).

Stringent response, transcriptional regulators, and outer membrane related genes important for fitness with increasing time in non-contaminated groundwater

A 48 hr survival fitness time course experiment was conducted in non-contaminated FW300 groundwater supplemented with 1 mM glucose (Fig. 1C). A total of 57 MT58 genes were identified as having large negative fitness changes between the 0.5 and 48 hr timepoints (Δfitness:$gw_{48h-0.5h}$) (Table S2). Several of the genes identified above as being important for fitness in mild conditions (Table S1), had large negative $\Delta fitness_{48h-0.5h}$ values. Additionally, several genes involved in the stringent response, including *relA* (-3.1), *dksA* (-3.6) and *sspA* (-2.1) (36-38), and several genes encoding transcriptional regulators and transcription factors, had large negative Δfitness:$gw_{48h-0.5h}$ values (Table S2). The latter included *uvrY* (-2.2) which is important for biofilm formation (39), as well as *cspE* (-2.2), part of the RpoS general stress response (40), and *mntR* (-2.2) which regulates Mn homeostasis (Table S2) (41). Nearly half of the genes with large negative $\Delta fitness_{48h-0.5h}$ values (28 of 57) are involved in various outer membrane processes (Fig. 2A, Table S3).

Genes important for fitness in contaminated groundwater include those involved in multiple different outer membrane systems

Survival fitness of the MT58 library was tested in different ORR groundwater samples (supplemented with 1 mM glucose) for 30 min. For comparison, similar survival experiments were also run in the *Pantoea* minimal medium adjusted to the five different pH values of the ORR groundwater samples (Fig. 1C). Only three genes had large negative fitness changes over

the pH 7.4 to pH 3.9 gradient in minimal medium (Δfitness:$mm_{pH3.9\text{-}pH7.4}$) (Table S4). In contrast, there were 45 MT58 genes with large negative fitness changes over the decreasing pH / increasing contamination gradient of the ORR groundwater samples (Δfitness:$gw_{pH3.9\text{-}pH7.4}$, Table S4). This includes *sspA* that encodes a transcriptional regulator critical for acid tolerance in *E. coli* (42) and the Mn efflux gene *mntP* (Fig. S3). As with the time course experiment, most of the genes (32 of 46) with large negative fitness changes with decreasing pH and increasing contamination were related to outer membrane processes (Fig. 2B, Table S3). A total of 20 LPS-related genes had large negative fitness changes with either increased time or increased contamination in groundwater and 13 of these were important for fitness in both conditions (Table S3). The largest negative fitness change trends for non-LPS outer membrane-related genes generally did not overlap between the time course and increasing contamination experiments (Fig. 2, Table S3).

Genes involved in aerobic glucose respiration negatively impact fitness in anoxic groundwater

To investigate the metabolic systems critical under different carbon availability regimes, a survival fitness experiment was conducted using non-contaminated FW300 (pH7.4) groundwater containing increasing amounts of glucose (0 to 5 mM) for a 24 hr incubation period (Fig. 1C). Several outer membrane integrity related genes were important for fitness with increasing glucose (Δfitness:$glu_{5mM\text{-}0mM}$), including *lpp* (-4.0), *tolQ* (-2.4), *ybiS* (-2.7), and *tolB* (-2.0), as well as the genes *mdoH* (-3.3) and *mdoG* (-3.3), which are involved in osmoregulation (Table S5) (43, 44). Also, several genes involved in glucose utilization had large negative Δfitness:$glu_{5mM\text{-}1mM}$ values, including the glycolytic gene *pgi* (-3.5) and anaerobically-expressed glucose fermentation genes *pflAB* (-1.7, -1.6) and *adhE* (-1.8) (Table S5). By contrast, several genes involved in the aerobic respiration of glucose had large positive Δfitness:$glu_{5mM\text{-}0mM}$ values (Table S6). This indicates a competitive advantage for mutant strains in the library that lack

functional copies of these genes. Genes with positive Δfitness:$glu_{5mM-0mM}$ values include those encoding subunits of dehydrogenase enzymes in the TCA cycle (45), as well as genes encoding subunits of complex 1 (encoded by the *nuo* genes) and cytochrome o oxidase of the aerobic respiratory chain (46, 47).

Like the genes encoding aerobic respiration, four genes of unknown function also had large positive Δfitness:$glu_{5mM-0mM}$ values with increasing glucose (Table S5). Sequence similarity networks for these genes (Fig. S4 – S7) revealed related sequences widely distributed in multiple families of the order *Enterobacteriaceae* (*IAI47_RS02400*, *IAI47_RS20625*), multiple orders of the class *Gammaproteobacteria* (*IAI47_RS09550*) and multiple classes of the phylum *Pseudomonadota* (*IAI47_RS05090*). However, none of the sequences in any of the networks were associated with a SwissProt description for a known function.

Pantoea **phylogroup containing MT58 enriched in contaminated ORR environment**

We used comparative genomics to contextualize the fitness results with MT58 to determine if the molecular mechanisms important for its survival were a core aspect of the *Pantoea* genus or limited to a subset of *Pantoea* species. Genomes with complete status are available for 75 *Pantoea* strains in the BV-BRC database (17) (Table S7) and phylogenetically these fall into three main groups (Fig. 3). Phylogroup 1 (30 strains) was mainly composed of *Pantoea* strains of the *Pantoea agglomerans* and *Pantoea vagans* species and includes the ORR isolate MT58. Phylogroup 2 (20 strains) was composed of *Pantoea* strains of the *Pantoea ananatis* and *Pantoea stewartii* species, whereas Phylogroup 3 (25 strains) were more diverse containing *Pantoea dispersa* strains among others. The majority of the strains were isolated from plant-related sources (42) whereas others were isolated from environmental water and sediment samples (10), from fungi (5) and from various animals including humans (6) (Table S7).

The prevalence of various *Pantoea* strains downstream of the S-3 Ponds contamination source was examined using 16S rRNA gene (V4 region) amplicon sequencing data from 32 sediment samples that were previously acquired by cone penetration testing (CPT) (48, 49). Two *Pantoea* specific ASVs were observed in the CPT samples, (ASV1 and ASV2) with ASV1 being present in more samples (19 vs 2) and at higher relative abundance (Fig. 4A). The more abundant ASV1 was exclusive to Phylogroup 1 and was a 100% match to the rRNA gene sequence of MT58 and to all but two of the other Phylogroup 1 strains. The less abundant ASV2 was a 100% match for all of the Phylogroup 2 strains and a majority of the Phylogroup 3 strains (Table S7). Hence, the extreme ORR environment has a preponderance of Phylogroup 1 *Pantoea* strains.

Pantoea **genus and phylogroup pangenomes are open with large accessory genomes**

Pangenomes for the 75 *Pantoea* genus strains as well as the three different *Pantoea* phylogroups were calculated (Fig. 4B) (21, 22). The pangenomes of Phylogroups 1 and 2 had similar percentages of orthologs in the core (present in all group strains), shell (within between 15% and 100% of group strains), and cloud (in ≤ 15% of group strains) of the pangenomes (Fig. 4C). The Phylogroup 3 and *Pantoea* genus pangenomes had larger percentages of genes in the cloud portion of the pangenome with less in the core. The openness parameter can vary from 0 to 1 and indicates that all four pangenomes are open (Fig. 4B) (24), with higher values being associated with a more open pangenome in which the inclusion of additional genomes to the pangenome is more likely to increase the total number of orthologues. Rarefaction curves for the four pangenomes approach the amount of core genome orthologues and ranged from 1491 to 2308 (Fig. S8). Amplification curves further demonstrate the openness of the pangenomes by not approaching an asymptote (Fig. S8). The estimated size for the four pangenomes ranges from 13,000 to 27,000 orthologues (Fig. 4B), revealing a large accessory genome for the *Pantoea* genus and all three phylogroups.

Non-core genes and genes located within mobile genetic elements are important for groundwater survival

While most of the genes identified as important for groundwater fitness were part of the core pangenome (Table S1-S5), next we investigated if there were any metabolic features apparent from the fitness studies that were unique to MT58 or a subset of *Pantoea* species that increase survivability in ORR groundwater. A handful of the genes important for fitness in the time course (6) and increasing contamination (13) groundwater survival experiments (14 total due to overlapping genes) fit this description as they are part of the cloud or low shell (orthologs in < 60 *Pantoea* species) of the *Pantoea* genus pangenome (Table S8). Gene locus *IAI47_RS07480* (a hypothetical gene) was the sole MT58 singleton with no orthologs found the remaining 74 analyzed *Pantoea* species. Three additional genes important for fitness in the increasing contamination experiment were part of the cloud of the *Pantoea* genus pangenome (Table S8). Several of the remaining low shell genes important for fitness can be described as phylogroup specific. These include four with no representative ortholog in Phylogroup 2 and one (*ydcX*) with no representative in Phylogroup 3 (Table S8). Eight of the 14 subset unique genes are predicted by annotation to be involved in LPS biosynthesis, which we showed was important in both the time course and increasing contamination survival experiments (Fig. 2).

Mobile genetic elements can move genetic information from one organism to another contributing to the organisms' accessory genome. They can be an important factor promoting rapid adaptation to changing environmental stresses. There are several different types of mobile genetic elements known including extrachromosomal plasmids as well as transposons, integrons, and phages that can integrate into the chromosome (50). To identify MT58 genes that were potentially acquired from mobile genetic elements, annotation information on each of the MT58 genes was retrieved from the BV-BRC database (Table S9) (17). The MT58 genome contains three plasmids (pMT58-1 (496 genes, 522 kbp), pMT58-2 (110 genes, 131 kbp, and

pMT58-3 (5 genes, 4.3 kbp)) (Table S9). Additionally, we analyzed the MT58 chromosome (3558 genes, 4,018 kbp) for integration of other mobile genetic elements using Island Viewer 4.0. This predicted the presence of 19 genomic islands encompassing approximately 5.7 % of the chromosome and 211 genes (Table S9) (25).

We divided the MT58 genes into four groups based on their genomic location; encoded on the chromosome, within a genomic island, or on pMT58-1 or pMT58-2. The distribution of chromosomally-encoded genes is heavily weighted to the high shell and core of the *Pantoea* pangenome, whereas genomic island and to a lesser extent the plasmid-encoded genes are found within the lower shell and cloud of the *Pantoea* genus pangenome (Fig. 4D). The same comparison using the Phylogroup 1 pangenome (of which MT58 is a member), shows a larger proportion of the pMT58-1 and pMT58-2 genes as part of the high shell and core (Fig. 4D). This indicates that these plasmids (or related plasmids) are observed more often in Phylogroup 1 strains compared to the *Pantoea* strains in the other phylogroups. Visualization of gene fitness values in non-contaminated (FW300) and highly contaminated (FW106) groundwater over gene prevalence in the *Pantoea* genus pangenome highlights accessory genes potentially acquired by mobile genetic elements important for contaminated groundwater survival (Fig. 5). A majority of the genes with large negative fitness values in FW106 groundwater that have orthologs in < 60 of the 75 *Pantoea* genomes are encoded on plasmids or genomic islands (Fig. 5).

Several plasmid and genomic island associated genes were found to be important in the survival fitness experiments. Three pMT58-1 genes had large fitness changes in the survival experiments. Two were hypothetical genes (*IAI47_RS19945* and *IAI47_RS20625*) that had large fitness changes in the glucose availability experiment (Table S5). The other was *ydcX* in the increasing contamination experiment (Fig. 5, Table S4). This gene encodes the orphan toxin OrtT that was shown in *E. coli* to slow cell growth in response to multiple different stress conditions (51). Two pMT58-2 genes (*arnE* and *arnF*) were important for fitness in the time

course survival experiment (Table S2). The *arnE* gene was also important for fitness in the increasing contamination experiment (Table S4). Both *arnE* and *arnF* are involved in LPS biosynthesis (52). Six genes of a predicted genomic island on the MT58 chromosome (*IAI47_RS07005* to *IAI47_RS07030*) were important for survival in contaminated groundwater (Fig. 5, Table S4). Within the genomic island, the *wbbO* (-1.7) gene encodes a glycosyl transferase important for O-antigen construction (53). The O-antigen construction process for *wbbO* is also dependent on an ATP-binding cassette transporter that likely includes *wzm* (-3.7) and *wzt* (-3.3), located within the same genomic island (53). Two additional hypothetical genes on different genomic islands (*IAI47_RS 07480* and *IAI47_RS 16520*) also had large negative Δfitness:$gw_{pH3.9\text{-}pH7.4}$ values (Table S4).

Discussion

Our *Pantoea* species tree partitioned into three major phylogroups. Phylogroup 1, which contains MT58, corresponded to the main *Pantoea* ASV found in sediment impacted by the S-3 ponds contamination plume (Fig. 4A). Comparison of the frequency of pMT58-1 and pMT58-2 gene orthologs in the Phylogroup 1 and *Pantoea* genus pangenomes shows that both of these plasmids are enriched in Phylogroup 1 and may be a discriminating factor characteristic of this phylogroup (Fig. 4D). The importance of these plasmids for survival of MT58 in increasingly contaminated groundwater was substantiated with survival fitness experiments that identified one pMT58-1 gene (*ydcX*, -1.7) and one pMT58-2 gene (*arnE*, -2.0) with large negative Δfitness:$gw_{pH3.9\text{-}pH7.4}$. Several other Phylogroup 1 specific genes were also found to be important in the time course and increasing contamination fitness experiments (Table S8), supporting the idea that Phylogroup 1 organisms are more prevalent than other *Pantoea* strains in the S-3 ponds contaminated environment due to their specific genomic content. Future efforts to obtain and characterize other *Pantoea* isolates from the ORR contaminated environment could help further define the importance of these fitness relevant genes and potentially identify others.

The *mntP* and *sspA* genes, important for Mn homeostasis and acid resistance, respectively (Fig. S3), along with 43 other genes had large negative fitness changes in increasingly contaminated ORR groundwater but negligible changes in minimal medium with decreasing pH (Table S4). Clearly, survival in contaminated ORR groundwater is due to multiple factors in addition to pH including exposure to toxic nitrate and metals (Fig. S1, Fig. S2). An aspect of bacterial survival in the contaminated ORR environment is synergistic effects from multiple challenges. The toxicity of metals can vary greatly with pH due to differences in solubility and oxidation states (54). In addition, the various ORR metals themselves can have synergistic negative impacts on microorganisms, as was previously observed for a contaminated site relevant ORR *Bacillus cereus* strain (55).

Genes involved in controlling and using carbon and energy sources were among the most important for groundwater survival. Under mild survival conditions, genes involved in the anaerobic catabolism of pyruvate were shown to be important for fitness (Fig. 1E). By contrast, large positive fitness changes with increasing glucose concentrations were observed for genes encoding several enzymes involved in aerobic respiration indicating the function of these genes inhibits survival in anoxic groundwater (Table S6). Although the exact underlying mechanism for this observation is unknown, several possibilities exist. Immediately after transfer from aerobic to anaerobic growth conditions, the TCA cycle and electron transport chain enzymes are still present and active while the terminal electron acceptor O_2 is absent (56), which could have adverse effects on the redox balance of cellular electron carriers. Active aerobic respiration machinery could also expose MT58 to mechanisms of toxicity in the survival challenges. In any case, the positive fitness phenotype highlights the challenge of a facultative lifestyle in groundwater environments and the importance of the molecular mechanisms that are used to react to changes in groundwater O_2 concentrations.

One benefit of library-based fitness experiments is gaining phenotypic information on poorly annotated genes (57). Several genes of unknown function had positive fitness changes with increasing glucose availability, as was seen for genes involving aerobic respiration (Table S5). Like the genes encoding redox enzymes of the TCA cycle and aerobic respiratory chain, these genes of unknown function are part of the core in the *Pantoea* pangenome (Table S5) and are widely distributed phylogenetically (Fig. S4-S7). It is intriguing to consider what functions these conserved hypothetical proteins carry out that is possibly related to the well-studied aerobic respiration systems.

In addition to individual genes, fitness experiments can also be useful in uncovering and confirming metabolic connections between sets of genes. A large number of outer membrane related genes were among the most important for survival in both the time course and

increasing contamination experiments (Table S3). Based on annotations (Table S9) and the reported survival fitness trends, several groupings of outer membrane-related genes emerged (Fig. 2, Table S3). Genes involved in maintaining outer membrane integrity were increasingly important for survival in non-contaminated groundwater with increasing time but not increasing contamination (Fig. 2). These genes with high negative $\Delta fitness_{48h-0.5h}$ values encode several physiologically-related proteins involved in outer membrane integrity and peptidoglycan processing. For example, with corresponding $\Delta fitness_{48h-0.5h}$ values in parentheses, TolB (-6.5) is a periplasmic protein that interacts with the transmembrane (cytoplasmic) proteins TolQ (-4.8), TolR, and TolA, while also interacting with outer membrane peptidoglycan associated proteins Lpp (-6.3), OmpA (-2.6), and Pal (58). Both NlpI (-5.6) and YbiS (-2.7) are involved in attachment of Lpp to peptidoglycan (59, 60) whereas Skp (-2.3) is a periplasmic chaperone critical in the biogenesis of OmpA (61). NagA (-2.2), LdcA (-2.9), and Prc (-5.5) are all involved in the synthesis and/or recycling of peptidoglycan (Table S3) (62-64). Previously, it has been proposed that the role of the Tol-Pal proteins in Gram-negative organisms is to link peptidoglycan to the outer membrane (58, 65). The similar fitness trends that we observed for multiple peptidoglycan related genes and the *tolBQ* genes supports this idea of an integrated outer membrane integrity system and clearly this is important for survival in the ORR environment.

Genes involved in LPS synthesis were critical for survival both with increased exposure time in non-contaminated groundwater and with decreasing pH / increasing contamination (Fig. 2, Table S3). LPS is the outermost component of the Gram-negative cell and is critical for outer membrane integrity as well as serving as a permeability barrier (66). Over 100 genes can be involved in the formation of the LPS layer, which is composed of three parts; the hydrophobic lipid A core, the hydrophilic core polysaccharide and the O-antigen hydrophilic oligosaccharide (67). The LPS genes we observed that were critical for groundwater survival are involved in

diverse aspects of LPS formation, including lipid A core synthesis (*yrbG*, *rfaG*, *rfaB*), O-antigen synthesis (*wbnF*, *wecA*), and transport of LPS components across the cytoplasmic membrane (*arnE*, *arnF*, *wzm*, *wzxE*) (52, 68-71). LPS-related mutants can be lethal or sensitive to multiple stress conditions related to groundwater survival, including osmotic shock, temperature shock, and exposure to antibiotics and metals (67, 72, 73).

The LPS genes that were important for groundwater survival were of disparate origins within the MT58 genome and *Pantoea* pangenome even though LPS is a universal molecular mechanism within the *Pantoea* genus and the larger *Enterobacteriaceae* family. Although several of the genes, including *rfaG*, *wzxE*, and *yrbG,* are chromosomally located and part of the core *Pantoea* pangenome, others, like *arnE* and *arnF,* are located on pMT58-2, and still others are part of the pangenome cloud and located within genomic islands (Fig. 3). Some of the chromosomally located LPS genes of interest, such as *IAI47_RS00670* and *IAI47_RS00670*, are also phylogroup specific (Table S3). Taken together, this shows that LPS synthesis genes critical for MT58 survival in groundwater have been acquired throughout the diversification of the *Pantoea* genus via acquisition of multiple different mobile genetic elements (Fig. 3). LPS is often identified as an important factor in microbial pathogenicity in which modifications can result in host immune system evasion (74). For example, horizontally transferred LPS genes impact interactions between pathogenic *Xanthomonas* strains and various plant hosts (75). The data herein suggests that LPS genes acquired from mobile genetic elements are also critical for adaptation in contaminated groundwater environments.

Conclusion

Microbial communities in anthropogenically-altered environments, like the ORR S-3 ponds, often have low species diversity but are enriched in functional genes essential to survival in the contaminated environment (4, 76). Metagenomic analyses, however, do not provide direct phenotypic insight and typically overlook genes of interest while providing little insight into mechanisms. Survival fitness experiments in which a MT58 RB-TnSeq library was directly exposed to environmental groundwater successfully uncovered genes important for groundwater survival including outer membrane related genes pertinent to Gram-negative *Enterobacteriaceae* and carbon catabolic genes pertinent to facultative organisms in general. Contextualization of these results with comparative genomics further revealed a tableau in which a microorganism from a non-contaminated environment relies on a combination of core and accessory metabolic mechanisms for survival when encountering an anthropogenically-altered environment.

Acknowledgements

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare no competing financial interests.

References

1. Thullner M, Regnier P. Microbial controls on the biogeochemical dynamics in the subsurface. Rev Mineral Geochem. 2019;85(1):265-302.

2. Liu S, Zheng T, Li Y, Zheng X. A critical review of the central role of microbial regulation in the nitrogen biogeochemical process: New insights for controlling groundwater nitrogen contamination. J Environ Manage. 2023;328:116959.

3. Riedel T, Kübeck C, Quirin M. Legacy nitrate and trace metal (Mn, Ni, As, Cd, U) pollution in anaerobic groundwater: Quantifying potential health risk from “the other nitrate problem”. Appl Geochem. 2022;139:105254.

4. Ning D, Wang Y, Fan Y, Wang J, Van Nostrand JD, Wu L, et al. Environmental stress mediates groundwater microbial community assembly. Nat Microbiol. 2024:490-501.

5. Techtmann SM, Hazen TC. Metagenomic applications in environmental monitoring and bioremediation. J Ind Microbiol Biot. 2016;43(10):1345-54.

6. Brooks SC. Waste characteristics of the former S-3 ponds and outline of uranium chemistry relevant to NABIR Field Research Center studies. Oak Ridge National Laboratory (US). 2001.

7. Shevenell L, Moore G, Dreier R. Contaminant spread and flushing in fractured rocks near Oak Ridge, Tennessee. Groundwater Monit Remediat. 1994;14(2):120-9.

8. Revil A, Skold M, Karaoulis M, Schmutz M, Hubbard SS, Mehlhorn TL, et al. Hydrogeophysical investigations of the former S-3 ponds contaminant plumes, Oak Ridge Integrated Field Research Challenge site, Tennessee. Geophysics. 2013;78(4):EN29-EN41.

9. Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu L, et al. Natural bacterial communities serve as quantitative geochemical biosensors. mBio. 2015;6(3).

10. Janda JM, Abbott SL. The changing face of the family *Enterobacteriaceae* (Order:"*Enterobacterales*"): New members, taxonomic issues, geographic expansion, and new diseases and disease syndromes. Clin Microbiol Rev. 2021;34(2):10.1128/cmr. 00174-20.

11. Walterson AM, Stavrinides J. *Pantoea*: insights into a highly versatile and diverse genus within the Enterobacteriaceae. FEMS Microbiol Rev. 2015;39(6):968-84.

12. Agarwal G, Gitaitis RD, Dutta B. Pan-genome of novel *Pantoea stewartii* subsp. indologenes reveals genes involved in onion pathogenicity and evidence of lateral gene transfer. Microorganisms. 2021;9(8):1761.

13. De Maayer P, Aliyu H, Vikram S, Blom J, Duffy B, Cowan DA, et al. Phylogenomic, pan-genomic, pathogenomic and evolutionary genomic insights into the agronomically relevant enterobacteria *Pantoea ananatis* and *Pantoea stewartii*. Front Microbiol. 2017;8:1755.

14. Thorgersen MP, Ge X, Poole FL, Price MN, Arkin AP, Adams MW. Nitrate-utilizing microorganisms resistant to multiple metals from the heavily contaminated Oak Ridge Reservation. Appl Environ Microb. 2019;85(17):e00896-19.

15. Thorgersen MP, Xue J, Majumder EL, Trotter VV, Ge X, Poole FL, et al. Deciphering microbial metal toxicity responses via random bar code transposon site sequencing and activity-based metabolomics. Appl Environ Microb. 2021;87(21):e01037-21.

16. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio. 2015;6(3).

17. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, et al. Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res. 2023;51(D1):D678-D89.

18. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: the United States department of energy systems biology knowledgebase. Nat Biotechnol. 2018;36(7):566-9.

19. Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS One. 2010;5(3):e9490.

20. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293-W6.

21. Buck M, Mehrshad M, Bertilsson S. mOTUpan: a robust Bayesian approach to leverage metagenome-assembled genomes for core-genome estimation. NAR Genom Bioinform. 2022;4(3):lqac060.

22. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026-8.

23. Méric G, Yahara K, Mageiros L, Pascoe B, Maiden MC, Jolley KA, et al. A reference pan-genome approach to comparative bacterial genomics: identification of novel epidemiological markers in pathogenic *Campylobacter*. PloS One. 2014;9(3):e92798.

24. Liao J, Guo X, Weller DL, Pollak S, Buckley DH, Wiedmann M, et al. Nationwide genomic atlas of soil-dwelling *Listeria* reveals effects of selection and population ecology on pangenome evolution. Nat Microbiol. 2021;6(8):1021-30.

25. Bertelli C, Laird MR, Williams KP, Group SFURC, Lau BY, Hoad G, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 2017;45(W1):W30-W5.

26. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38(12):5825-9.

27. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309-D14.

28. Zallot R, Oberg N, Gerlt JA. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. Biochem. 2019;58(41):4169-82.

29. Oberg N, Zallot R, Gerlt JA. EFI-EST, EFI-GNT, and EFI-CGFP: enzyme function initiative (EFI) web resource for genomic enzymology tools. J Mol Biol. 2023;435(14):168018.

30. Bertani G. Studies on lysogenesis I: the mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol. 1951;62(3):293-300.

31. Huang YY, Price MN, Hung A, Gal-Oz O, Ho D, Carion H, et al. Barcoded overexpression screens in gut *Bacteroidales* identify genes with new roles in carbon utilization and stress resistance. bioRxiv. 2022:2022.10. 10.511384.

32. Ge X, Vaccaro BJ, Thorgersen MP, Poole FL, Majumder EL, Zane GM, et al. Iron-and aluminium-induced depletion of molybdenum in acidic environments impedes the nitrogen cycle. Environ Microbiol. 2019;21(1):152-63.

33. Thorgersen MP, Lancaster WA, Ge X, Zane GM, Wetmore KM, Vaccaro BJ, et al. Mechanisms of chromium and uranium toxicity in *Pseudomonas stutzeri* RCH2 grown under anaerobic nitrate-reducing conditions. Front Microbiol. 2017;8:1529.

34. McGuire JT, Long DT, Klug MJ, Haack SK, Hyndman DW. Evaluating behavior of oxygen, nitrate, and sulfate during recharge and quantifying reduction rates in a contaminated aquifer. Environ Sci Technol. 2002;36(12):2693-700.

35. Gunsalus R, Park S-J. Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the ArcAB and Fnr regulons. Res Microbiol. 1994;145(5-6):437-50.

36. Loveland AB, Bah E, Madireddy R, Zhang Y, Brilot AF, Grigorieff N, et al. Ribosome• RelA structures reveal the mechanism of stringent response activation. Elife. 2016;5:e17029.

609 37. Pal RR, Bag S, Dasgupta S, Das B, Bhadra RK. Functional characterization of the stringent
610 response regulatory gene *dksA* of *Vibrio cholerae* and its role in modulation of virulence phenotypes. J
611 Bacteriol. 2012;194(20):5638-48.
612 38. Williams MD, Ouyang TX, Flickinger MC. Starvation-induced expression of SspA and SspB: the
613 effects of a null mutation in *sspA* on *Escherichia coli* protein synthesis and survival during growth and
614 prolonged starvation. Mol Microbiol. 1994;11(6):1029-43.
615 39. Suzuki K, Wang X, Weilbacher T, Pernestig A-K, Melefors O, Georgellis D, et al. Regulatory
616 circuitry of the CsrA/CsrB and BarA/UvrY systems of *Escherichia coli*. J Bacteriol. 2002;184(18):5130-40.
617 40. Phadtare S, Inouye M. Role of CspC and CspE in regulation of expression of RpoS and UspA, the
618 stress response proteins in *Escherichia coli*. J Bacteriol. 2001;183(4):1205-14.
619 41. Waters LS, Sandoval M, Storz G. The *Escherichia coli* MntR miniregulon includes genes encoding
620 a small protein and an efflux pump required for manganese homeostasis. J Bacteriol.
621 2011;193(21):5887-97.
622 42. Hansen AM, Qiu Y, Yeh N, Blattner FR, Durfee T, Jin DJ. SspA is required for acid resistance in
623 stationary phase by downregulation of H-NS in *Escherichia coli*. Mol Microbiol. 2005;56(3):719-34.
624 43. Lequette Y, Odberg-Ferragut C, Bohin J-P, Lacroix J-M. Identification of *mdoD*, an *mdoG* paralog
625 which encodes a twin-arginine-dependent periplasmic protein that controls osmoregulated periplasmic
626 glucan backbone structures. J Bacteriol. 2004;186(12):3695-702.
627 44. Bontemps-Gallo S, Bohin J-P, Lacroix J-M. Osmoregulated periplasmic glucans. EcoSal Plus.
628 2017;7(2):10.1128/ecosalplus. ESP-0001-2017.
629 45. Huynen MA, Dandekar T, Bork P. Variation and evolution of the citric-acid cycle: a genomic
630 perspective. Trends Microbiol. 1999;7(7):281-91.

631 46. Unden G, Steinmetz PA, Degreif-Dünnwald P. The aerobic and anaerobic respiratory chain of
632 *Escherichia coli* and *Salmonella enterica*: enzymes and energetics. EcoSal Plus.
633 2014;6(1):10.1128/ecosalplus. ESP-0005-2013.
634 47. Cotter P, Chepuri V, Gennis R, Gunsalus R. Cytochrome o (*cyoABCDE*) and d (*cydAB*) oxidase
635 gene expression in *Escherichia coli* is regulated by oxygen, pH, and the *fnr* gene product. J Bacteriol.
636 1990;172(11):6333-8.
637 48. Goff JL, Szink EG, Thorgersen MP, Putt AD, Fan Y, Lui LM, et al. Ecophysiological and genomic
638 analyses of a representative isolate of highly abundant *Bacillus cereus* strains in contaminated
639 subsurface sediments. Environ Micriobiol. 2022;24(11):5546-60.
640 49. Lunne T, Powell JJ, Robertson PK. Cone penetration testing in geotechnical practice: CRC press;
641 2002.
642 50. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. Nat Rev
643 Genet. 2015;16(8):472-82.
644 51. Islam S, Benedik MJ, Wood TK. Orphan toxin OrtT (YdcX) of *Escherichia coli* reduces growth
645 during the stringent response. Toxins. 2015;7(2):299-321.
646 52. Wang X, Quinn PJ. Lipopolysaccharide: Biosynthetic pathway and structure modification. Prog
647 Lipid Res. 2010;49(2):97-107.
648 53. Kos V, Whitfield C. A membrane-located glycosyltransferase complex required for biosynthesis
649 of the D-galactan I lipopolysaccharide O antigen in *Klebsiella pneumoniae*. J Biol Chem.
650 2010;285(25):19668-78.
651 54. Wang Z, Meador JP, Leung KM. Metal toxicity to freshwater organisms as a function of pH: A
652 meta-analysis. Chemosphere. 2016;144:1544-52.
653 55. Goff JL, Chen Y, Thorgersen MP, Hoang LT, Poole FL, Szink EG, et al. Mixed heavy metal stress
654 induces global iron starvation response. ISME J. 2023;17(3):382-92.

56. Tolla DA, Savageau MA. Regulation of aerobic-to-anaerobic transitions by the FNR cycle in *Escherichia coli*. J Mol Biol. 2010;397(4):893-905.

57. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, et al. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature. 2018;557(7706):503-9.

58. Clavel T, Germon P, Vianney A, Portalier R, Lazzaroni JC. TolB protein of *Escherichia coli* K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp and OmpA. Mol Microbiol. 1998;29(1):359-67.

59. Schwechheimer C, Rodriguez DL, Kuehn MJ. NlpI-mediated modulation of outer membrane vesicle production through peptidoglycan dynamics in *Escherichia coli*. Microbiologyopen. 2015;4(3):375-89.

60. Magnet S, Bellais S, Dubost L, Fourgeaud M, Mainardi J-L, Petit-Frère S, et al. Identification of the L, D-transpeptidases responsible for attachment of the Braun lipoprotein to *Escherichia coli* peptidoglycan. J Bacteriol. 2007;189(10):3927-31.

61. Schäfer U, Beck K, Müller M. Skp, a molecular chaperone of Gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. J Biol Chem. 1999;274(35):24567-74.

62. Arai Y, Sparks DL. ATR–FTIR spectroscopic investigation on phosphate adsorption mechanisms at the ferrihydrite–water interface. J Colloid Interf Sci. 2001;241(2):317-26.

63. Debanu Das MH, Marc-André Elsliger, Rameshwar U. Kadam, Joanna C. Grant, Hsiu-Ju Chiu, Mark W. Knuth, , Heath E. Klock MDM, Adam Godzik, Scott A. Lesley, Ashley M. Deacon, Dominique Mengin-Lecreulx, , Wilson IA. Structure and Function of a Novel LD-Carboxypeptidase A Involved in Peptidoglycan Recycling. J Bacteriol. 2013;195(24):5555.

64. Hara H, Yamamoto Y, Higashitani A, Suzuki H, Nishimura Y. Cloning, mapping, and characterization of the *Escherichia coli prc* gene, which is involved in C-terminal processing of penicillin-binding protein 3. J Bacteriol. 1991;173(15):4799-813.

65. Lloubès R, Cascales E, Walburger A, Bouveret E, Lazdunski C, Bernadac A, et al. The Tol-Pal proteins of the *Escherichia coli* cell envelope: an energized system required for outer membrane integrity? Res Microbiol. 2001;152(6):523-9.

66. Sperandeo P, Martorana AM, Polissi A. Lipopolysaccharide biosynthesis and transport to the outer membrane of Gram-negative bacteria2019. 9-37 p.

67. Klein G, Raina S. Regulated assembly of LPS, its structural alterations and cellular response to LPS defects. Int J Mol Sci. 2019;20(2):356.

68. Sperandeo P, Pozzi C, Dehò G, Polissi A. Non-essential KDO biosynthesis and new essential cell envelope biogenesis genes in the *Escherichia coli yrbG–yhbG* locus. Res Microbiol. 2006;157(6):547-58.

69. Parker C, Kloser A, Schnaitman C, Stein M, Gottesman S, Gibson B. Role of the *rfaG* and *rfaP* genes in determining the lipopolysaccharide core structure and cell surface properties of *Escherichia coli* K-12. J Bacteriol. 1992;174(8):2525-38.

70. Lehrer J, Vigeant KA, Tatar LD, Valvano MA. Functional characterization and membrane topology of Escherichia coli WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide. J Bacteriol. 2007;189(7):2618-28.

71. Wang K-C, Huang C-H, Chang P-R, Huang M-T, Fang S-B. Role of *wzxE* in *Salmonella Typhimurium* lipopolysaccharide biosynthesis and interleukin-8 secretion regulation in human intestinal epithelial cells. Microbiol Res. 2020;238:126502.

72. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res. 2018;46(W1):W486-W94.

700 73. Schut GJ, Thorgersen MP, Poole FL, Haja DK, Putumbaka S, Adams MW. Tungsten enzymes play
701 a role in detoxifying food and antimicrobial aldehydes in the human gut microbiome. Proc Natl Acad Sci
702 USA. 2021;118(43):e2109008118.
703 74. Beutler B. LPS in microbial pathogenesis: promise and fulfilment. Journal of Endotoxin Research.
704 2002;8(5):329-35.
705 75. Patil PB, Bogdanove AJ, Sonti RV. The role of horizontal transfer in the evolution of a highly
706 variable lipopolysaccharide biosynthesis locus in *xanthomonads* that infect rice, citrus and crucifers.
707 BMC Evol Biol. 2007;7(1):1-13.
708 76. Waldron PJ, Wu L, Nostrand JDV, Schadt CW, He Z, Watson DB, et al. Functional gene array-
709 based analysis of microbial community structure in groundwaters with a gradient of contaminant levels.
710 Environ Sci Technol. 2009;43(10):3529-34.
711 77. Crain AV, Broderick JB. Pyruvate formate-lyase and its activation by pyruvate formate-lyase
712 activating enzyme. J Biol Chem. 2014;289(9):5723-9.
713

Figure legends

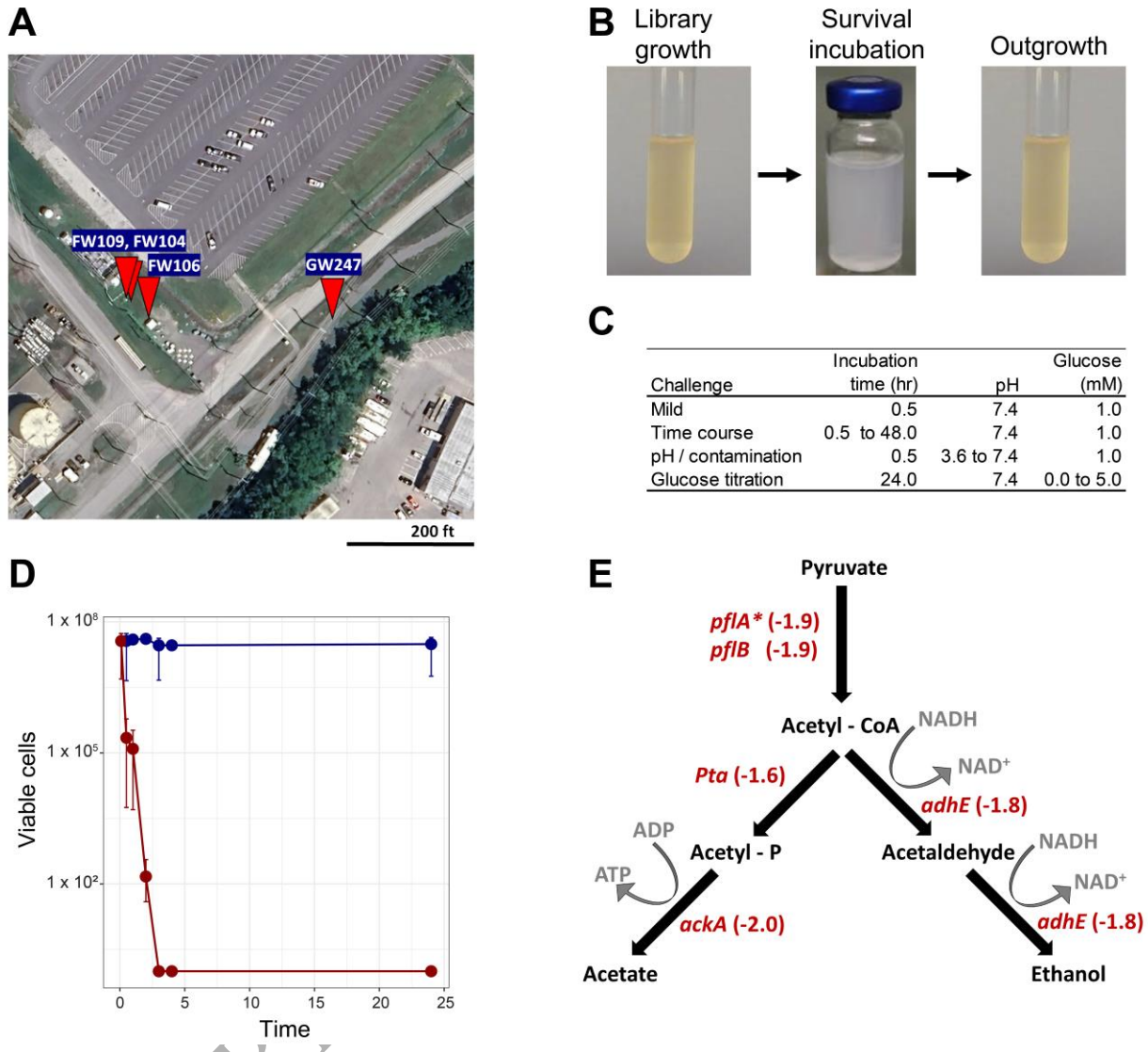
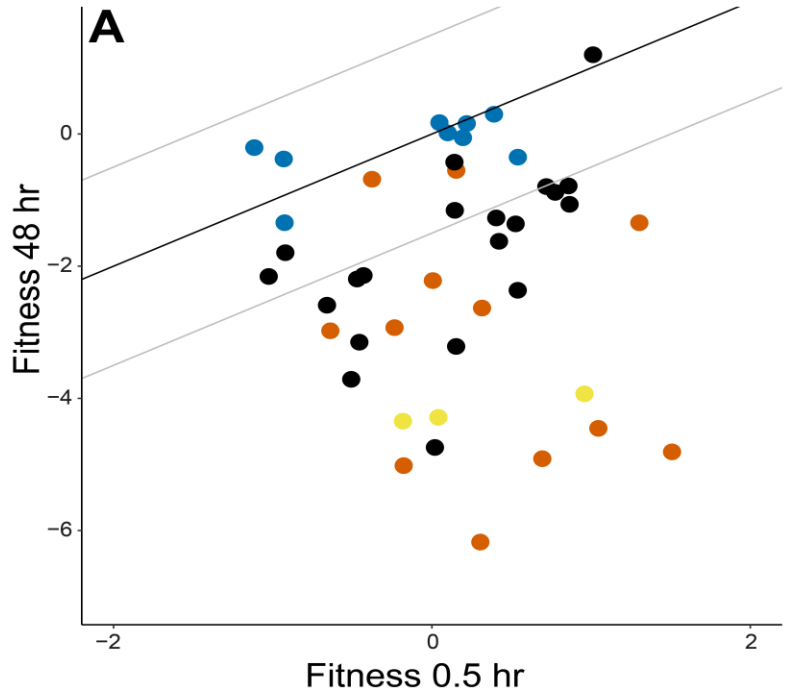
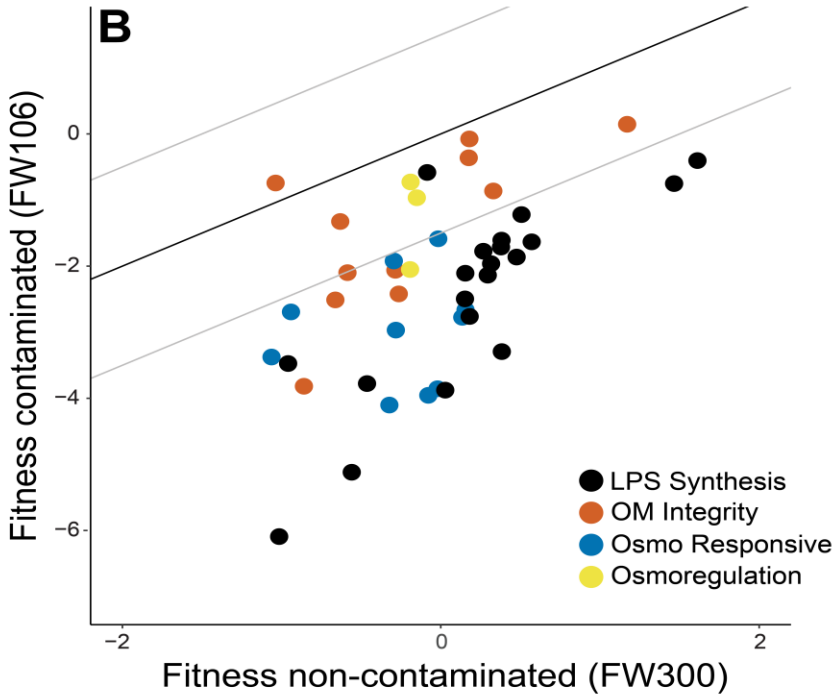


Figure 1: Survival fitness experiments of MT58 in ORR groundwater. A) Map of ORR S-3 ponds contaminated site in Bear Creek Valley (TN, USA). Red markers denote the location of wells within the S-3 ponds contamination plume that were used in this study. Wells used in this study included non-contaminated FW300 (pH7.4) (not depicted), as well as contaminated wells GW247 (pH 5.4), FW104 (pH 5.0), FW109 (pH 4.4), and FW106 (pH 3.9). Imagery: ©2024 Airbus, Maxer Technologies. **B)** Setup of survival fitness experiments with groundwater incubation challenges. **C)** Parameters of survival fitness experiments. **D)** Survival of wild-type

MT58 in non-contaminated (FW300, pH7.4, blue) and highly contaminated (FW106, pH 3.9, red) groundwater. **E)** Anaerobic pyruvate catabolism in MT58. Red values indicate negative fitness values observed under mild groundwater challenge conditions (non-contaminated FW300 groundwater for 30 min). * *pflA* is an accessory gene required for activation of PflB (77).





728 **Figure 2: Groupings of outer membrane related genes with similar fitness patterns. A)**
729 Fitness change over time scatter plot after survival incubation challenge in non-contaminated
730 FW300 groundwater for 48 hr vs 0.5 hr. **B)** Fitness change with contamination scatter plot after
731 survival incubation challenge for 30 min in contaminated FW106 groundwater versus non-
732 contaminated FW300 groundwater. The black line is the 1:1 line, and the grey lines show the
733 cutoff for large fitness changes at ≥ |1.5|. Genes belonging to different categories of outer
734 membrane related genes are detailed in Table S3.

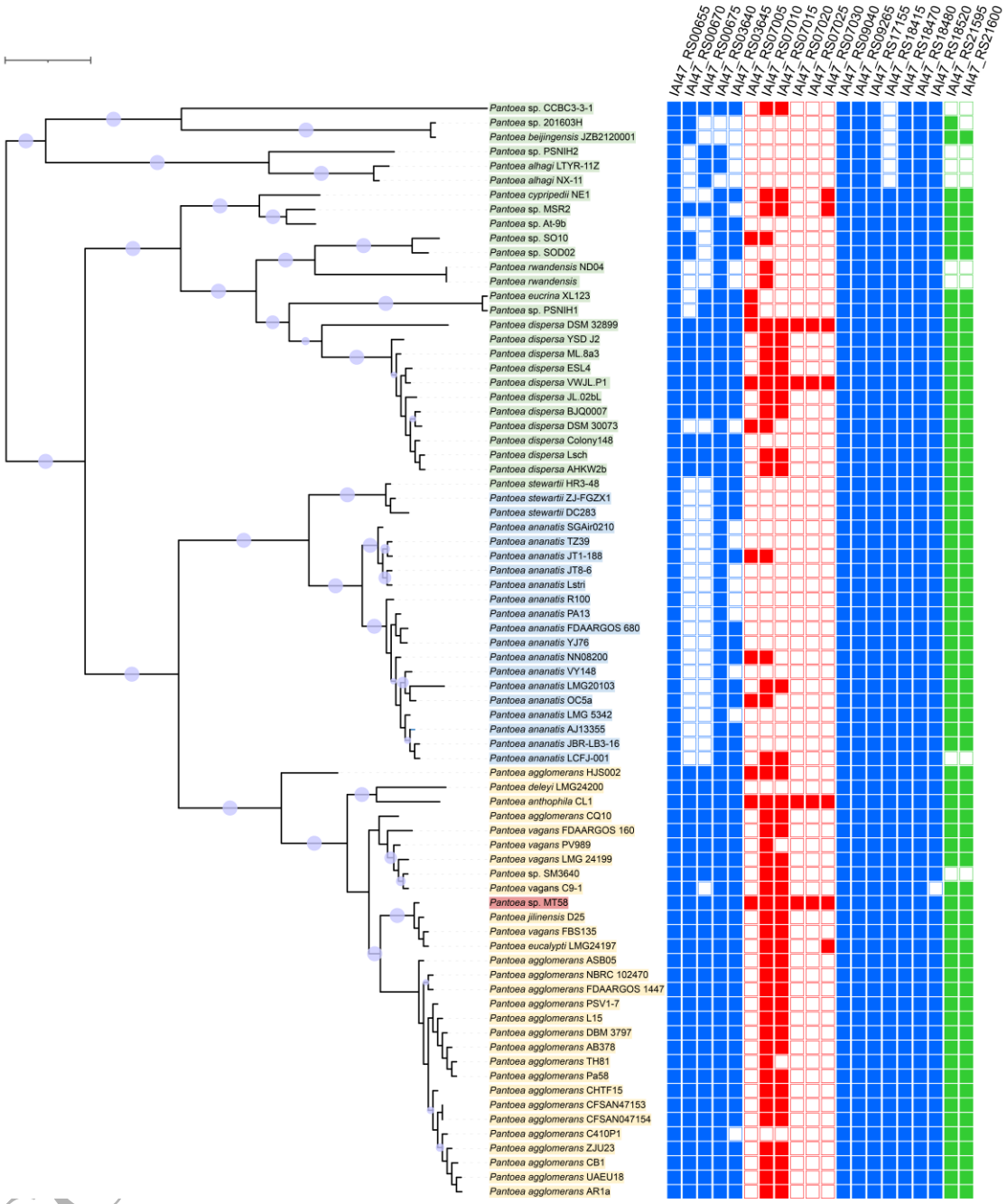
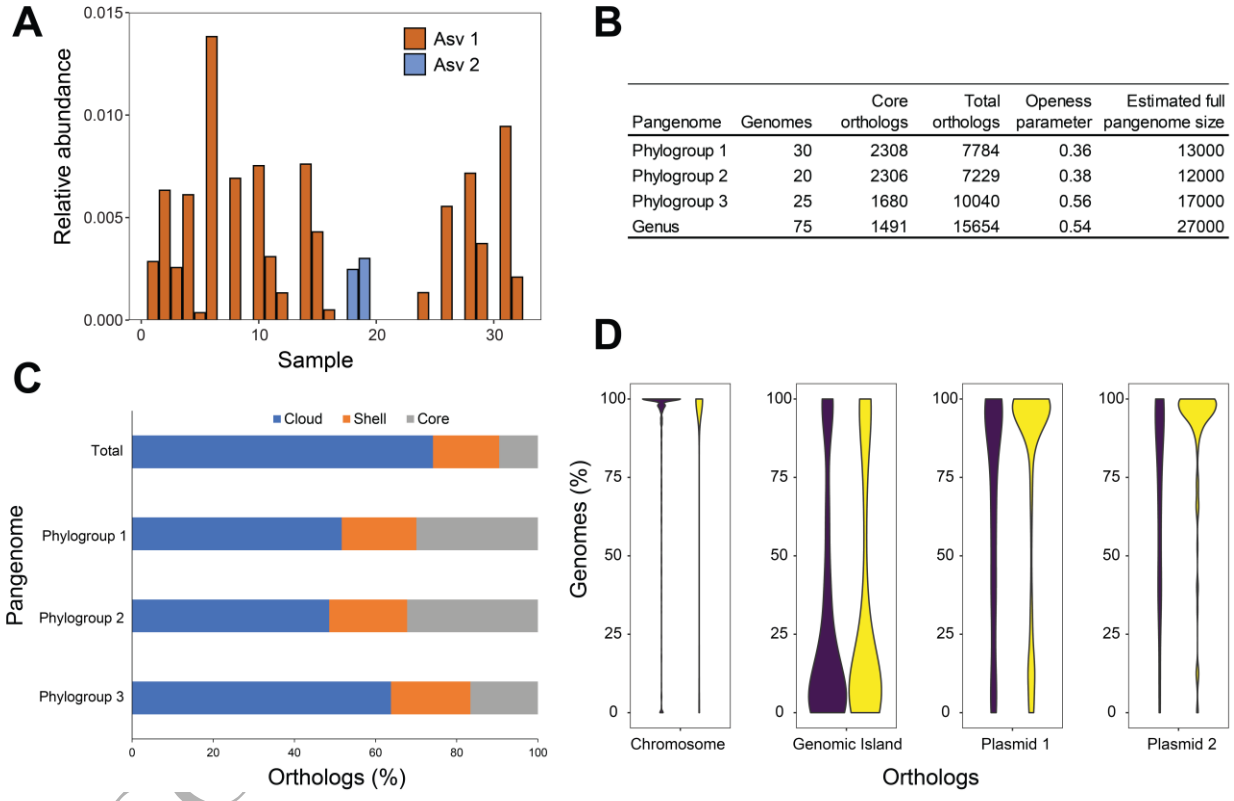


Figure 3: *Pantoea* genus phylogenetic tree. Phylogenetic tree of 75 *Pantoea* genus strains used in this study broken into phylogroups 1 (orange), 2 (blue), and 3 (green). The ORR isolate MT58 is highlighted in red. The tree was constructed using multiple sequence alignments of COGs for 49 core universal genes (19). Tree scale of 0.01 is indicated in top left corner.

Bootstrap values from 0.9 to 1.0 are indicated as purple circles with larger values having larger circles. The array to the right of the tree consists of genes involved in LPS synthesis that were found to be important for survival in at least one fitness experiment with filled boxes indicating which of the *Pantoea* strains contain a homologue of the MT58 gene. Data was obtained from the *Pantoea* genus pangenome data. The genome location of the gene in MT58 is indicated by color; chromosomal (blue), genomic island (red), and plasmid (green).



Pangenome	Genomes	Core orthologs	Total orthologs	Openess parameter	Estimated full pangenome size
Phylogroup 1	30	2308	7784	0.36	13000
Phylogroup 2	20	2306	7229	0.38	12000
Phylogroup 3	25	1680	10040	0.56	17000
Genus	75	1491	15654	0.54	27000

Figure 4: *Pantoea* genus distribution across Oak Ridge Reservation (ORR) contaminated site. **A)** Relative abundance of *Pantoea* ASV sequences detected in contaminated ORR sediment samples (48). Values reported are the highest of two for each sample tested. **B)** Summary data for *Pantoea* genus and phylogroup pangenomes. **C)** Orthologs located in the core (present in all group strains), shell (within between 15% and 100% of group strains), and

cloud (in ≤ 15% of group strains) of the Total and Phylogroup based pangenomes. **D)** MT58 ortholog distribution in the *Pantoea* genus (purple) and Phylogroup 1 (yellow) pangenomes by genome location.

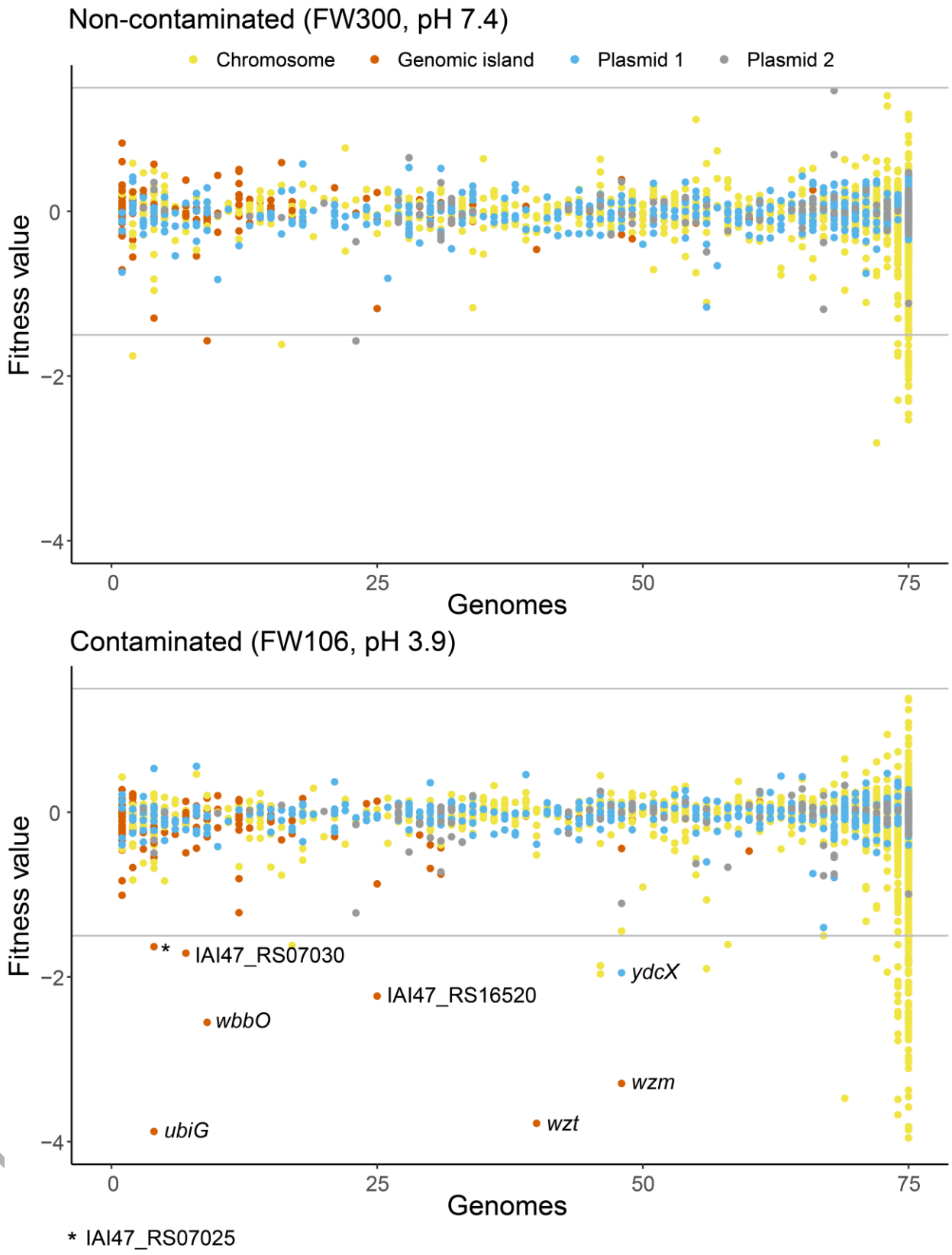


Figure 5: Comparison of gene fitness values after incubation challenge in non-contaminated (FW300) and contaminated (FW106) groundwater. Gene fitness values are from 30-minute incubation challenges in the indicated ORR groundwater sample. Gene fitness values are colored based on genome location of the gene and are arrayed on the x-axis based on the number of genomes containing an ortholog of the gene in the *Pantoea* genus pangenome. Genes of interest in the low shell or cloud of the Total pangenome (Table S8) are labeled to the right of their corresponding dot.