AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems

| Molecular and Cellular Biology | Research Article

Complex interplay between gene deletions and the environment uncovers cellular roles for genes of unknown function in *Escherichia coli*

Kaat Sondervorst,[1] Kristina Nesporova,[1] Matthew Herdman,[1] Bart Steemans,[1] Joëlle Rosseels,[1] Sander K. Govers[1]

AUTHOR AFFILIATION See affiliation list on p. 18.

ABSTRACT Phenotypic outcomes can be heavily affected by environmental factors. In this study, we exploited the previously observed nutrient dependency of cell biological phenotypic features captured by a cross-condition image-based profiling assay of *Escherichia coli* deletion strains to examine this in more detail. We identified several general principles, including the existence of a spectrum of deviating phenotypes across nutrient conditions (i.e., from nutrient- or feature-specific to pleiotropic phenotypic deviations), limited conservation of phenotypic deviations across nutrient conditions (i.e., limited phenotypic robustness), and a subset of nutrient-independent phenotypic deviations (indicative of consistent genetic determinants of specific phenotypic features). In a subsequent step, we used this cross-condition data set to identify five genes of unknown function (*yibN*, *yaaY*, *yfaQ*, *ybiJ*, and *yijD*), of which the deletion displayed either nutrient-independent phenotypic deviations or phenotypic similarities to genes of known function. These genes showed different levels of phylogenetic conservation, ranging from conserved across the tree of life (*yibN*) to only present in some genera of the Enterobacterales (*yaaY*). Analysis of the structural properties of the proteins encoded by these y-genes, identification of structural similarities to other proteins, and the examination of their subcellular localization yielded new insights into their contribution to *E. coli* cell morphogenesis, cell cycle progression, and cell growth. Together, our approach showcases how bacterial image-based profiling assays and data sets can serve as a gateway to reveal the function of uncharacterized proteins.

IMPORTANCE Despite unprecedented access to genomic information, predicting phenotypes based on genotypes remains notoriously difficult. One major confounding factor is the environment and its ability to modulate phenotypic outcomes. Another fact is that a large fraction of protein-coding genes in bacterial genomes remain uncharacterized and have no known function. In this work, we use a large-scale cross-condition image-based profiling dataset to characterize nutrient-dependent phenotypic variability of *Escherichia coli* deletion strains and exploit it to provide insight into the cellular role of genes of unknown function. Through our analysis, we identified five genes of unknown function that we subsequently further characterized by examining their phylogenetic conservation, predicted structural properties and similarities, and their intracellular localization. Combined, this approach highlights the potential of cross-condition image-based profiling, which extracts many cell biological phenotypic readouts across multiple conditions, to better understand nutrient-dependent phenotypic variability and uncover protein function.

KEYWORDS cell biology, cell cycle, *Escherichia coli*, y-genes, fluorescent image analysis, phylogenetic analysis, protein structure predictions

Editor Julia Willett, University of Minnesota Twin Cities, Minneapolis, Minnesota, USA

Address correspondence to Sander K. Govers, sander.govers@kuleuven.be.

The authors declare no conflict of interest.

See the funding table on p. 18.

Received 10 February 2025
Accepted 14 May 2025
Published 10 June 2025

Copyright © 2025 Sondervorst et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

A cellular phenotype is determined by the interaction of a cell's genotype with stochastic, epigenetic, and environmental factors (1–4). Uncovering the intricate relationships between these aspects is crucial in understanding the emergence of different cellular phenotypes and establishing clear genotype-phenotype associations. A phenotype itself is also complex and multifactorial, consisting of many, often interdependent traits, ranging from cellular fitness and growth capacity to cell morphology, intracellular organization, and cell cycle aspects. In model microbes such as yeast and *Escherichia coli,* systematic genome-wide screens, made possible through the construction of arrayed knockout collections, transposon mutant libraries, or CRISPR-based approaches (5–11), have enabled the identification of genetic factors that contribute to specific phenotypic traits. For example, such screens have identified genes required for swarming motility (12), for growth across different environments (13), and for T7 bacteriophage growth (14) in *E. coli*.

While these initial screens typically focused on specific phenotypic traits, the advent of high-content screening methods has enabled the simultaneous readout of many phenotypic parameters. One prime example is image-based profiling, in which a large number of quantitative cellular and subcellular features are extracted from microscopy images of individual cells (15–18). This approach captures high-content cell biological information and provides more detailed insights into cellular physiology and the many aspects that constitute a given phenotype. Image-based profiling approaches across the genome-wide deletion collection of *E. coli* have revealed that a large fraction of the non-essential genome (~20%) affects different aspects of cell morphogenesis, cell cycle progression, and growth (19–21). However, these screens were performed in a single environment, and we have recently shown that a large fraction of phenotypic alterations in these deletion mutants are, in fact, nutrient-dependent (22). More specifically, we found that the majority of deviating phenotypes (~75%), for all extracted phenotypic features, exhibited nutrient dependency and did not deviate consistently across nutrient conditions. In line with this, a screen of the *E. coli* single-gene deletion library across 30 different nutrient conditions highlighted variable growth phenotypes across conditions for many deletion strains (23). These findings underscore the important role of the environment (in this case, the nutrients available in the growth medium) in determining phenotypic outcomes.

The goal of this study was to provide more detailed insights into the nature of this nutrient-dependency of phenotypic landscapes, while also exploiting both the extent (i.e., the many phenotypic features that were extracted) and the nutrient-dependency of these landscapes to uncover potential cellular roles for genes of unknown function. Even in a well-studied model system as *E. coli*, about one-third of genes lack experimental evidence of function (24, 25). These unannotated genes are typically referred to as "y-genes," stemming from their primary names starting with the letter "y" (26). As metagenomic approaches have provided us with unprecedented access to bacterial genome sequences and the genes that lie within them, such sequence-to-function gaps are becoming increasingly more common in microbiology (27–29). Addressing this gap is challenging given the relatively slow process of experimental characterization, yet it is crucial in expanding our understanding of genetic diversity and cellular functioning, and further optimizing computational gene function predictions. One way to assign a function to a gene relies on the investigation of the consequences of gene inactivation (e.g., through deletion or transposon insertion). In earlier genome-wide screens, this was achieved by examining the growth or fitness of specific strains in which the genes of interest have been deleted or inactivated, across different nutrient conditions (13, 28–30). Here, we extended this approach by examining cross-condition image-based profiles of *E. coli* strains in which such y-genes were deleted (22). We exploited the high-content phenotypic characterization across conditions to identify specific y-gene deletion strains with nutrient-independent deviations of phenotypic features or phenotypic similarities to genes of known function. This yielded a set of five candidate genes that were subjected to bioinformatic analyses (to assess their level of conservation), structural

predictions (to identify their structural properties and homologies with other proteins), and fluorescent labeling (to determine their intracellular localization). Together, our experiments reveal crucial insights into the cellular role of these uncharacterized proteins and their contribution to the cellular functioning of *E. coli*.

RESULTS

Re-analysis of cross-condition image-based profiling data reveals specific and pleiotropic deviating phenotypes

To examine the apparent nutrient dependency of phenotypic landscapes in more detail, we started from an existing data set that we generated in a recent study (22). This data set contains population-level cellular features, extracted from phase contrast and fluorescence microscopy images, for 806 *E. coli* deletion strains. These strains expressed fluorescent protein fusions to a cell division protein (FtsZ) and a DNA replication marker (SeqA) and were sampled across four nutrient conditions (Fig. 1A). These four nutrient conditions varied substantially in carbon source and nutrient quality and were chosen from a panel of 41 nutrient conditions to represent a wide range of different average growth rates, cell sizes, and DNA replication regimes (discrete vs overlapping DNA replication cycles) (22). All four nutrient conditions were based on M9 buffer, supplemented with either L-alanine (M9Lala), glycerol (M9gly), glucose (M9glu), or L-arabinose with casamino acids and thiamine (M9LaraCAAT). In this case, M9Lala was the poorest nutrient condition, giving rise to the slowest growth rates, and richness gradually increases in M9gly, M9glu, and M9LaraCAAT. While the original data set consisted of 77 average population-level features (related to cell morphology, nucleoid morphology, divisome formation, DNA replication, and growth) for each deletion strain per nutrient condition (22), we expanded this data set for the current study to 96 features (Fig. S1A). The additional features consist of coefficients of variation (CVs) for specific cellular properties, which provide a measure for the variability of that property and thus also the level of control that is exercised over that property in a given nutrient condition. For all features in a given nutrient condition, we also calculated normalized scores (*s*), which represent the extent to which a mutant strain deviates from its respective wild type for that feature and nutrient condition (see Materials and Methods). Typically, a threshold score $|s| \geq 3$ for a specific feature is used to designate a deletion strain as one with a significant deviation or defect for that feature. A general overview of the distribution of deviating scores per feature and nutrient condition is provided in Fig. S1A. The phenoprint of a deletion strain is given by the collection of all its feature values and scores across conditions. In total, each phenoprint of our current data set consists of 384 features (96 per nutrient condition for each strain). Both the raw measurements and normalized scores for all strains across four nutrient conditions can be found in Data set S1.

In a first step, we investigated the specificity of deviating phenotypes for a given strain by capturing the extent to which a deviation is restricted to a given feature or nutrient condition. For this, we quantified both the number of nutrient conditions in which a strain displays a deviation (i.e., number of nutrient conditions with $|s| \geq 3$ for at least one feature), as well as the number of features for which a strain displays a deviation (i.e., number of features with $|s| \geq 3$ in at least one nutrient condition) (Fig. 1B). This analysis revealed the existence of a broad range of deviating phenotypes, with strains displaying very specific phenotypes (e.g., deviating in only a single feature in a single nutrient condition), more pleiotropic phenotypes (e.g., deviating in >90 features across all four nutrient conditions), and everything in between (Fig. 1B). Together, this spectrum of deviating phenotypes illustrates the heterogeneity in phenotypic alterations across nutrient conditions.

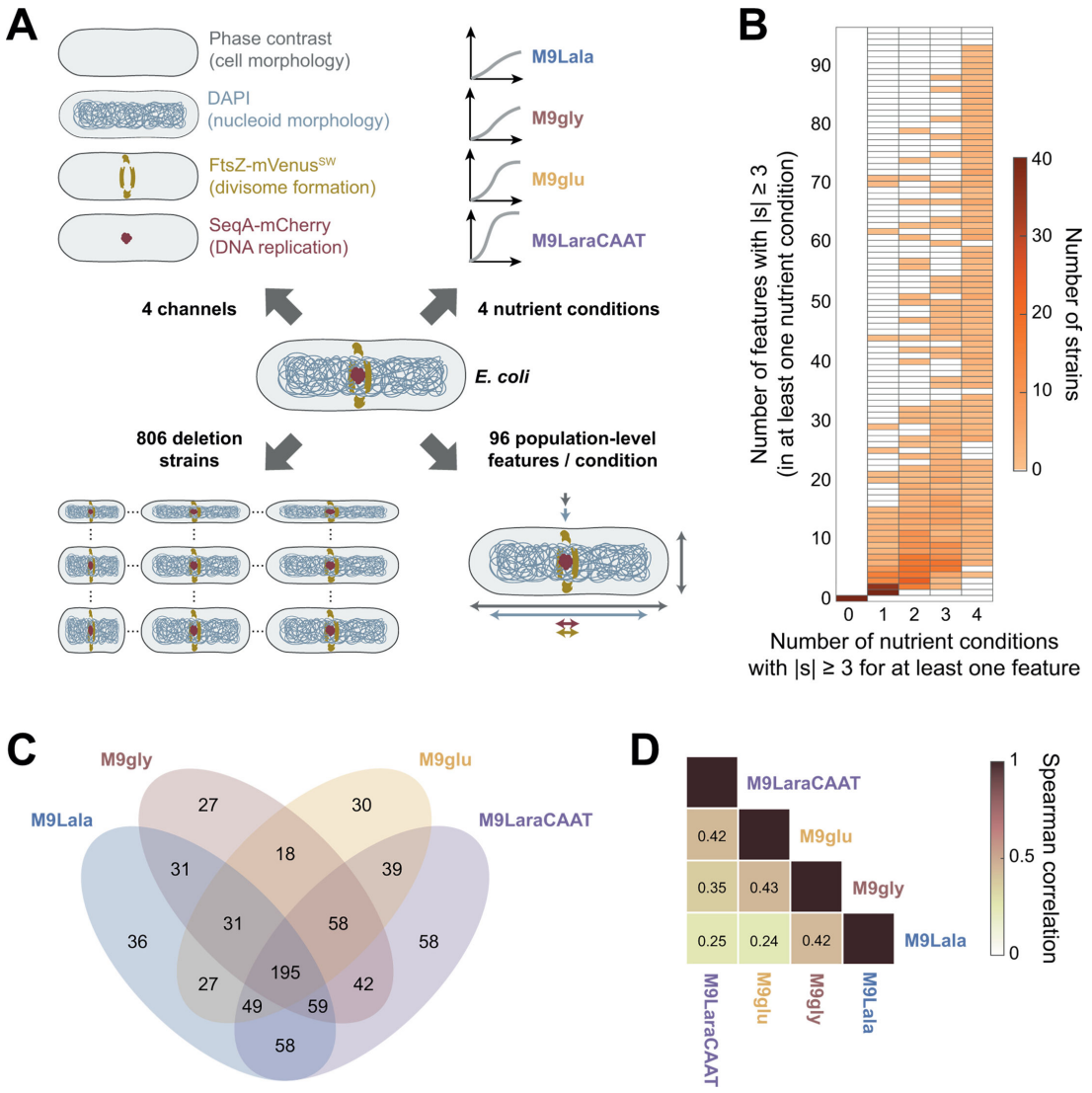


FIG 1 Diversity of deviating phenotypes and limited phenotypic robustness of deviations across nutrient conditions. (A) Schematic showing the complexity of the previously acquired (22) cross-condition image-based profiling data set of a library of 806 *E. coli* deletion strains. These strains were stained with DAPI (4′,6-diamidino-2-phenylindole) to visualize their nucleoids and expressed two chromosomally encoded fluorescent protein fusions to visualize divisome formation (FtsZ-mVenus^{SW}) and DNA replication (SeqA-mCherry). (B) Heatmap showing the distribution of deletion strains upon quantification of the number of features that deviate from the wild type ($|s| \geq 3$) in at least one nutrient condition versus the number of nutrient conditions in which at least one feature deviates from the wild type. (C) Four-way Venn diagram showing the number of strains that display a deviating phenotype ($|s| \geq 3$ for at least one feature) across different combinations of nutrient conditions. (D) Heatmap showing average correlations of deviating phenotypes across nutrient conditions. For each feature, only deletion strains displaying a deviating phenotype ($|s| \geq 3$) in at least one nutrient condition were considered upon calculation of the correlation coefficients, which were then averaged across features.

Limited robustness of phenotypic deviations across nutrient conditions

To probe how this heterogeneity affected the conservation of phenotypes of specific mutants across different nutrient conditions, we first examined in which nutrient condition strains displayed a deviating phenotype ($|s| \geq 3$ for at least one feature). We found that 195 strains (24.2% of all deletion strains) displayed a deviation in all four nutrient conditions, whereas 197 (24.4%), 215 (26.7%), and 151 (18.7%) strains did so in three, two, and one out of four nutrient conditions, respectively (Fig. 1C). In addition, we detected an overlap in deletion strains with deviating phenotypes between all possible combinations of nutrient conditions, indicating extensive nutrient-dependent

phenotypic variability. This extensive phenotypic variability resulted in limited correlations of phenotypic scores of strains across conditions (Fig. 1D). For these correlations, only deletion strains with a deviation ($|s| \geq 3$) in at least a single nutrient condition were considered per feature, and correlations were then averaged across features. While comparisons between more similar nutrient conditions (in terms of richness, as measured by the growth rates and cell sizes they support) gave rise to higher correlation values than more dissimilar nutrient conditions (Fig. S1B), the overall correlation values remained low, indicating limited robustness of deviating phenotypes across nutrient conditions in general.

Nutrient-independent phenotypes enable the identification of genetic determinants of cell morphogenesis and cell cycle progression

While the preservation of phenotypes was generally limited across nutrient conditions, we were able to identify a number of nutrient-independent outliers, defined as having $s > 3$ (or $s < -3$) for at least one nutrient condition and $s > 1.5$ (or $s < -1.5$) in all other conditions, for almost all extracted features (Fig. 2A). The majority of strains displaying a nutrient-independent deviation did so for a limited number of features (Fig. 2B), although we also found some pleiotropic phenotypes (with >40 altered features per strain) that were nutrient-independent (Fig. 2B). The nutrient-independent nature of the deviations enabled us to more confidently identify genetic determinants of specific cellular features (Fig. 2C). An overview of all nutrient-independent outliers for all features is included in Data set S1. Similar to our previous study that focused on cell cycle laws (i.e., general principles that connect different cell cycle events to specific aspects of cell size [22]), many of the genes deleted in the strains that display nutrient-independent deviations encode proteins that localize to either the cell envelope or the nucleoid (Fig. 2C). Most noteworthy was the observation that almost all deletion strains displaying a consistently increased cell width were deletions of cell envelope components (17/22 or 77.3%; Fig. 2C), indicating an important link between envelope integrity and the determination of cell width.

From this analysis, two y-gene deletion strains with interesting nutrient-independent phenotypes also emerged (Fig. 2C). The first one was the Δ*yibN* strain, which consistently gave rise to wider cells across conditions (Fig. 2C). Similar to other deletion strains that gave rise to wider cells, YibN contains a predicted transmembrane region and is expected to localize to the inner membrane. In addition, this strain also displayed longer cells in richer nutrient conditions, giving rise to markedly larger cells (in terms of cell volume and surface area) in these conditions (Fig. S2A). Another remarkable y-gene deletion strain was the Δ*yaaY* strain, which displayed more narrow cells and a later relative timing of FtsZ ring formation across nutrient conditions (Fig. 2C). In the richest nutrient condition, this strain also displayed a more pleiotropic phenotype, with cells being smaller on average (in terms of both length and width) with a large number of additional phenotypic alterations (Fig. S2A). This observation could be a consequence of significant growth defects, illustrated by our inability to obtain reliable growth measurements for this strain across all nutrient conditions (Fig. S2A). Together, these observations implicate both YibN and YaaY in important aspects of cell morphogenesis, cell cycle progression, and growth.

Cross-condition phenotypic profiling can inform the cellular roles of genes of unknown function

An additional advantage of phenotypic profiling across multiple nutrient conditions is that it allows the identification of deletion strains with similar phenoprints in a more reliable and non-nutrient-specific way (i.e., even if certain phenotypic alterations are not preserved across nutrient conditions). We achieved this by examining correlations between the obtained multi-condition phenoprints of strains. High phenoprint correlations between two deletion strains, both positive and negative, are predictive of functional connections between the deleted genes and can inform on gene function

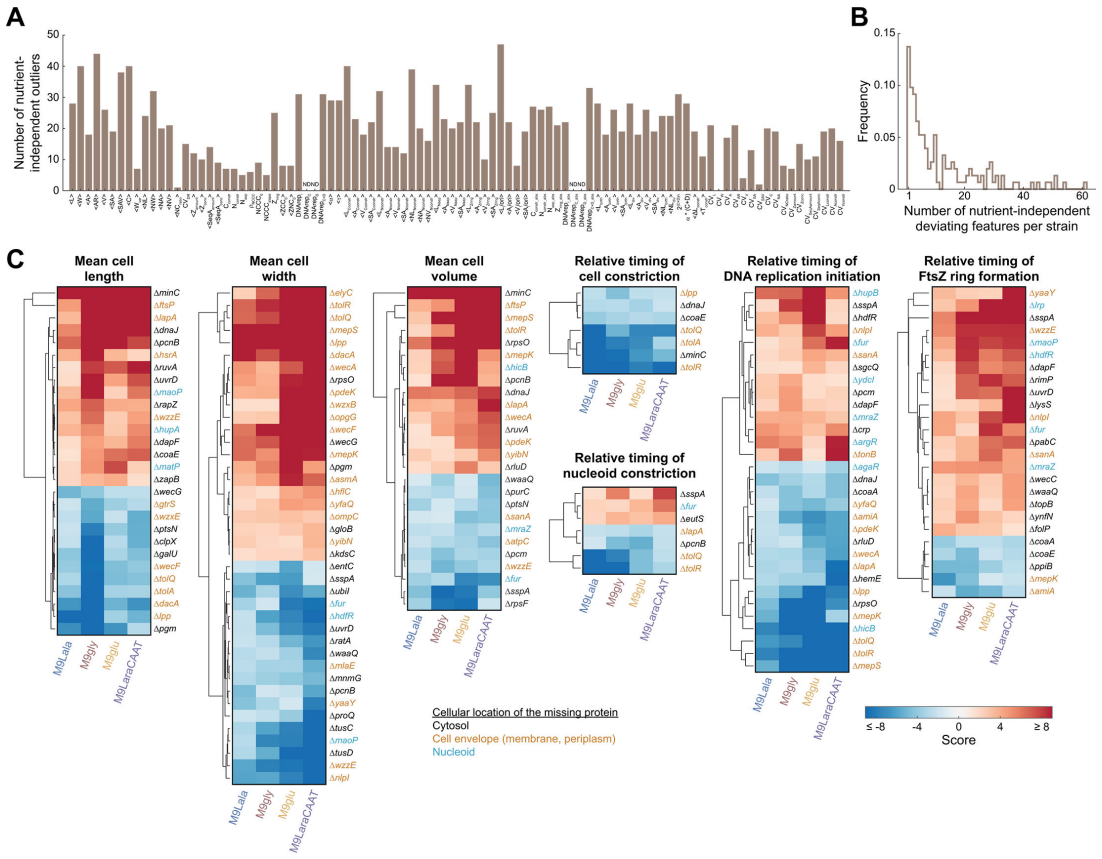


FIG 2 Identification of the genetic determinants of cell morphogenesis and cell cycle progression using nutrient-independent deviations of deletion strains. (A) Bar graph showing the number of nutrient-independent outliers for each feature. Nutrient-independent outliers were defined as having $s > 3$ (or $s < -3$) for at least one nutrient condition and $s > 1.5$ (or $s < -1.5$) in all other conditions. Feature abbreviations are defined in the legend of Data set S1. ND, not determined. The relative and absolute timing of chromosome replication ($DNArep_C$ and $DNArep_{C_abs}$) and chromosome replication termination ($DNArep_D$ and $DNArep_{D_abs}$) could not be determined in M9LaraCAAT due to cells displaying overlapping rounds of chromosome replication. (B) Frequency distribution of the number of nutrient-independent deviating features per strain. (C) Clustergrams showing the normalized scores of nutrient-independent outliers for the indicated features.

(13, 31). To identify deletion strains with related phenoprints, we first calculated pairwise correlation coefficients for all strains that were sampled across all four nutrient conditions (Fig. 3A). This yielded a broad distribution of correlation coefficients (Fig. 3B), of which some of the highest absolute values were obtained between strains in which genes were deleted that encode proteins known to be involved in similar cellular processes (Fig. 3C), serving as a proof-of-concept.

For example, TolR and TolQ are both inner membrane components of the Tol-Pal system that spans the entire cell envelope and is crucial for maintaining its integrity (32, 33). IhfA and IhfB are the two subunits of the nucleoid-associated protein IHF that specifically binds DNA, introduces sharp DNA bends, and affects gene expression (34–36). WecG and WecF both catalyze cytoplasmic steps in the biosynthesis of enterobacterial common antigen (37). For all these pairs, their positive phenoprint correlation exemplifies their similar roles in the same cellular process or pathway (Fig. 3C, top row). In contrast, strong negative correlations often indicated opposing roles (Fig. 3C, bottom row). For example, BamE is a lipoprotein that is part of the β-barrel assembly machinery, the Bam complex, which plays a crucial role in the biogenesis of integral outer membrane proteins in Gram-negative bacteria (38–40). Deletion of *bamE* leads to jamming of the Bam complex by RcsF, a Bam complex substrate and outer membrane protein-dependent surface-exposed lipoprotein involved in stress sensing (41–43). Surface exposure of RcsF normally occurs via the assembly of complexes with

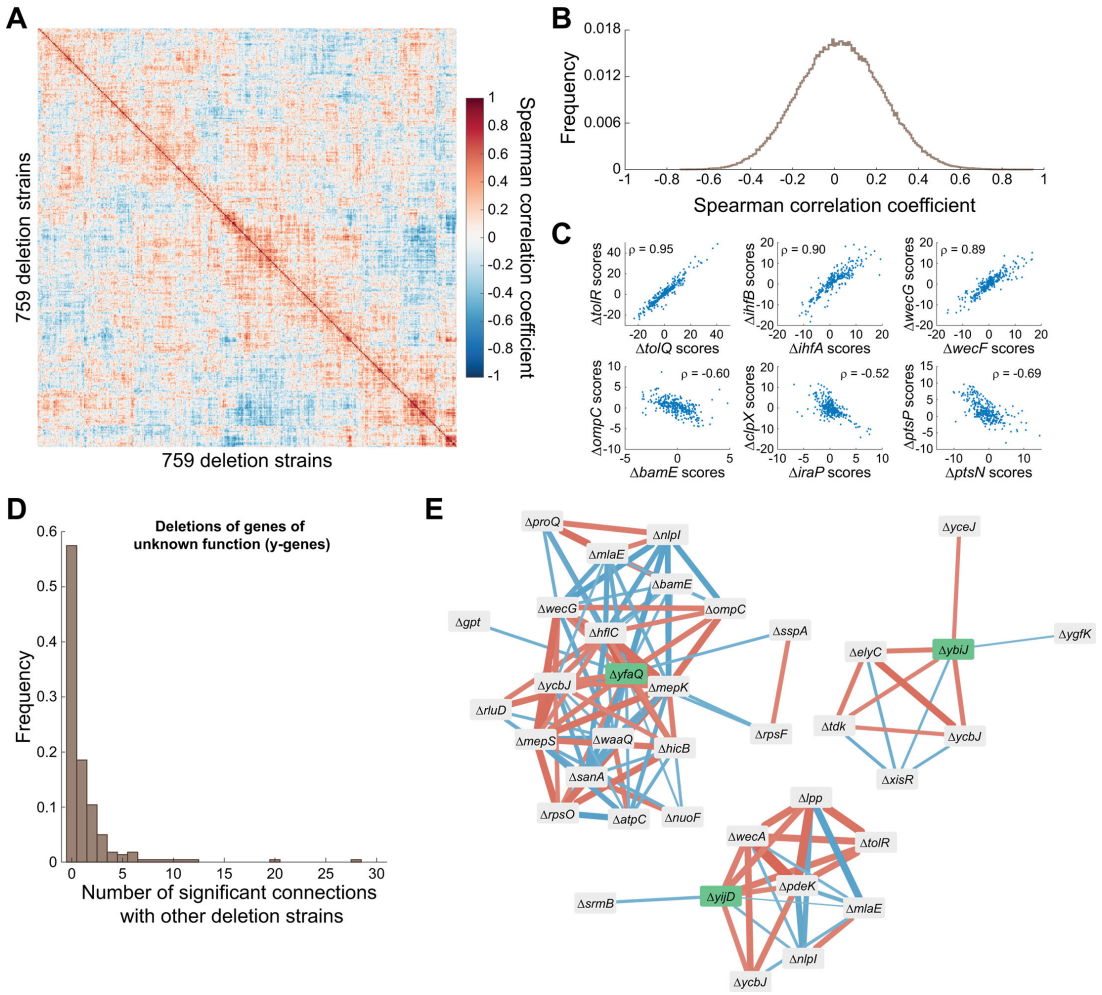


FIG 3 Phenoprint correlation networks of deletion strains can inform the cellular roles of deleted genes. (A) Heatmap showing the pairwise correlations between the cross-condition phenoprints of 759 deletion strains. (B) Frequency distribution of all pairwise phenoprint correlations. (C) Examples of strong positive (top row) and negative (bottom row) correlations between phenoprint scores of strains in which the deleted gene products are known to be involved in similar or opposing cellular functions. (D) Frequency distribution of the number of significant phenoprint connections between strains with deletions of y-genes and other deletion strains. (E) Network representation of significant phenoprint correlations between deletions of the indicated y-genes (green) and other deletion strains. The color of the edges indicates the sign of the correlation (blue: negative, red: positive), and their width indicates the relative strength of the correlation.

various outer membrane proteins, including OmpC (44, 45). Both BamE and OmpC are thus functionally connected to RcsF, but their deletion leads to different RcsF-related outcomes (i.e., jammed in the Bam complex vs not in complex with OmpC) that likely underlie the contrasting phenotypes of their respective deletion strains. In other cases, the opposite roles are more clear, as IraP is an anti-adaptor protein of the ClpXP protease (46, 47), and PtsN and PtsP are competing parts of the nitrogen-related phosphotransferase system of *E. coli* (48, 49).

In a subsequent step, we exploited this approach to provide more insight into the cellular role of specific y-genes by examining potential connections of their deletion phenoprint with that of other genes of known function. For each y-gene deletion strain, we identified significant connections by using empirical thresholds to exclude the middle 99.7% of all correlation coefficients (corresponding to all correlation coefficients that fall within three standard deviations of the mean). This analysis showed that close to 50% of y-gene deletion phenoprints displayed significant connections to other deletion phenoprints, ranging from just one to many significant connections (Fig. 3D). This distribution of significant connections was similar to that obtained when examining the

phenoprint connections of all deletion strains (Fig. S2B). An example of a y-gene deletion strain with many connections was the Δ*yfaQ* strain, which displayed connections to several deletion strains of genes involved in peptidoglycan or membrane remodeling and integrity (Δ*mepK*, Δ*mepS*, Δ*wecG*, Δ*ompC*, Δ*bamE*, Δ*mlaE*, Δ*nlpI*, Δ*waaQ*, Δ*sanA,* Δ*hflC*; Fig. 3E). Together with its predicted signal sequence and phenoprint (wider cells across nutrient conditions; Fig. S2A), this connection network suggested a potential role for the YfaQ protein in cell envelope assembly and/or structuring. Other y-gene deletions that displayed interesting connections and phenoprints included the Δ*ybiJ* strain, which had positive connections with the Δ*elyC* and Δ*ycbJ* strains (Fig. 3E) while displaying smaller cells in nutrient-poor and larger cells in nutrient-rich conditions (Fig. S2A). ElyC is an envelope biogenesis factor (50), whereas the gene encoding YcbJ overlaps with that of ElyC (opposite orientation), leading to the inactivation of *elyC* in a *ycbJ* deletion mutant. YbiJ contains a signal sequence, suggesting it localizes to the cell envelope and could also be involved in maintaining its integrity. Another notable y-gene deletion strain was Δ*yijD*, which displayed gradually wider cells in increasingly richer nutrient conditions, a slower growth rate in nutrient-poor conditions, early cell and nucleoid constriction across all conditions, and positive connections to Δ*tolR* and Δ*lpp* (Fig. 3E; Fig. S2A). As both these proteins are involved in linking different parts of the cell envelope or are parts of complexes that do so (33, 51), YijD, shown to be an inner membrane protein (52), could play a similar role or be part of a similar complex.

Differences in phylogenetic conservation of identified genes of unknown function

To investigate the level of evolutionary conservation of the five identified y-genes (*yibN*, *yaaY*, *yfaQ*, *ybiJ*, and *yijD*), we first evaluated the occurrence of orthologs across the tree of life (Fig. 4A; Fig. S3). Of our five genes, *yaaY* displayed the most limited conservation (Fig. 4A and B) and was only found in a subset of genera of the Enterobacterales (i.e., *Escherichia*, *Salmonella*, and *Citrobacter*). While *yijD* was mostly conserved across the Enterobacterales, it was also present in most Vibrionales (Fig. 4A and B), including in notorious pathogens such as *Vibrio cholerae*. This contrasts with *ybiJ*, which was found across Enterobacterales, but not in other Orders (Fig. 4B; Fig. S4). While *yfaQ* exhibited an irregular and limited conservation pattern across the bacterial domain, *yibN* orthologs were found across the entire tree of life (Fig. 4A and B), even in archaea and eukaryotes.

In a second stage, we examined the conservation of these five y-genes across a large-scale *E. coli* genome collection, which we recently assembled (consisting of 131,610 *E. coli* genomes [53]). This analysis revealed that four genes (*yaaY*, *ybiJ*, *yibN,* and *yijD*) were strongly conserved across *E. coli* strains, as they could be found in (almost) all genomes (Fig. 4C). The only exception was *yfaQ*, which was still present in approximately 98% of the genomes but was notably depleted or even completely absent in strains of specific sequence types (Fig. 4D; Table S1). Together, this analysis revealed clear differences in the broadness of the phylogenetic spectrum in which the y-genes operate.

Cell envelope-associated localization patterns of YibN and YbiJ

To provide further insight into the cellular role of these y-genes, we determined their homology (Table S2), performed structural predictions, and examined their intracellular localization. For this last step, we attempted to construct C-terminal msfGFP fusions to each of them at their native chromosomal locus. For all selected y-genes but *yaaY*, we successfully generated such a fluorescent protein fusion and subjected these strains to microscopy across the four different nutrient conditions of our initial screen (i.e., M9Lala, M9gly, M9glu, and M9LaraCAAT).

YfaQ homology determination and structural predictions suggest a bipartite arrangement of the protein (Fig. S5; Table S2), where the N-terminal domain shares features of transpeptidases, while the C-terminal domain is structurally similar to SpoIID transglycosylase (Fig. S5). YfaQ also harbors an N-terminal Sec signal peptide and SPI cleavage site, suggesting export to the periplasm (Fig. S5). While construction was

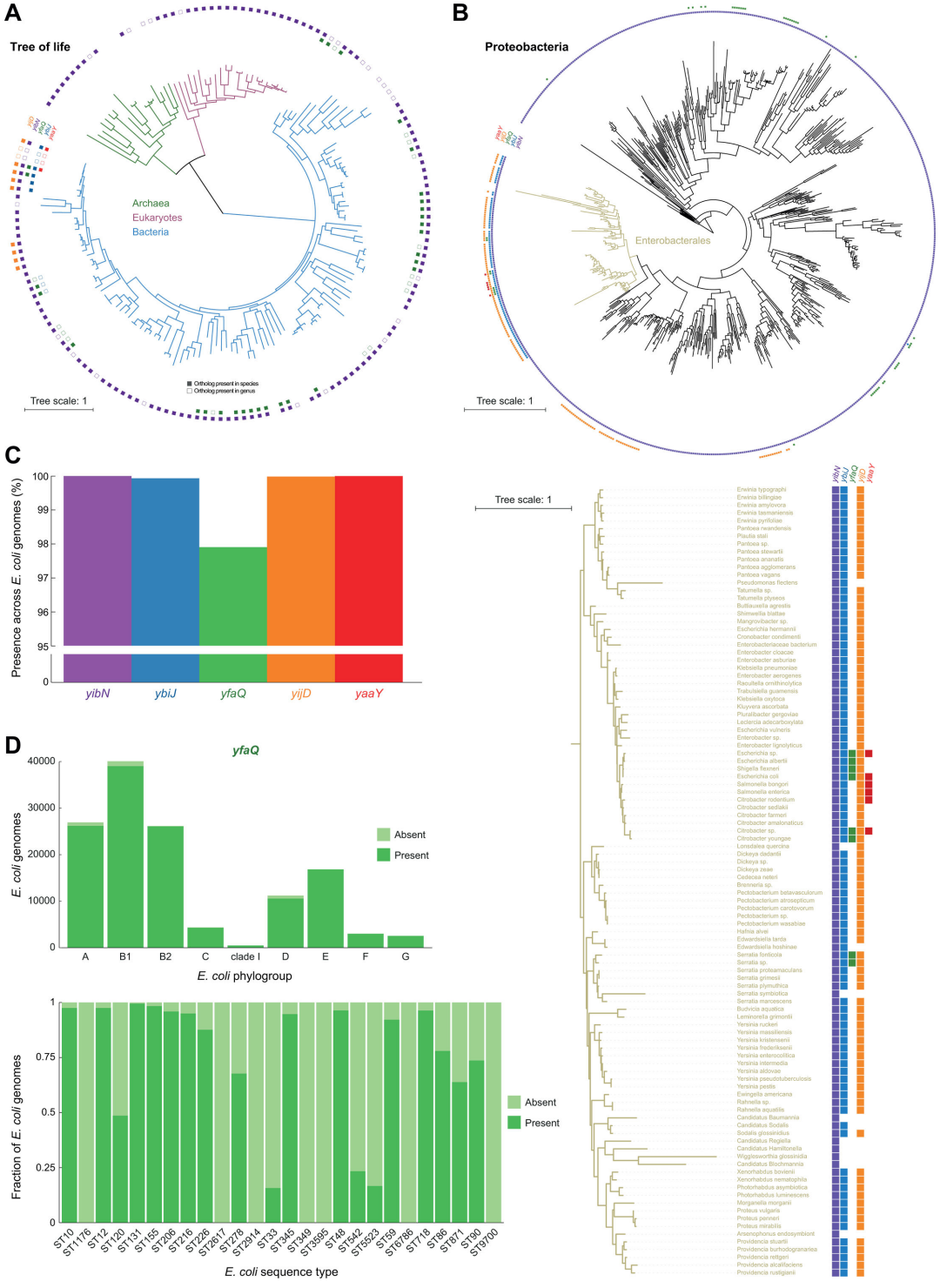


FIG 4 Conservation analysis of selected y-genes. (A) The presence of the indicated y-gene orthologs, determined using the EggNOG database, was mapped onto the tree of life. Full squares indicate the presence of the respective ortholog in a specific species, while empty squares indicate the presence of the ortholog in the relevant genus of the species present in the tree (but not in that species itself). The main tree branches represent the domains of Archaea (green), Eukaryotes (purple), and Bacteria (blue). Species names can be found in Fig. S3. (B) The presence of the indicated y-gene orthologs was mapped to the tree of YibN orthologs in Proteobacteria. The Enterobacterales branch (yellow) is depicted separately in detail, including species names. The species names for the full tree can be found in Fig. S4. (C) Bar graph indicating the prevalence of the y-genes across a large-scale genomic collection of *E. coli* (consisting of 131,610 genomes). (D) Patterns of presence or absence of *yfaQ* across *E. coli* phylogenetic groups (top) and in 26 sequence types (relative distribution) that had at least 10 genomes in which *yfaQ* was absent (bottom).

successful for YfaQ-msfGFP, we could not detect a signal in cells producing this fusion across conditions. This could be a consequence of either low fusion protein levels in these conditions (y-genes often have lower expression levels [24]) or aberrant protein fusion production and/or functionality.

YibN is predicted to consist of a short N-terminal transmembrane domain, formed by a hydrophobic, non-polar, α-helical tail, and a larger cytoplasmic globular domain that shares homology with rhodanese domain proteins (Fig. 5A). Proteins harboring such rhodanese domains are ubiquitous across the tree of life (54), which likely explains the broad occurrence of *yibN* orthologs (Fig. 4A). Rhodanese domain proteins are implicated in a wide variety of cellular processes, mostly through their ability to bind and transfer sulfur atoms via a cysteine residue within their active site (54–56) (Fig. 5B and C). While the predicted catalytic domain of YibN contains such a canonical cysteine residue, this catalytic cysteine residue is on the opposite side of the enzymatic loop of the rhodanese domain in comparison to that of other rhodanese proteins (Fig. 5B and C). The YibN-msfGFP fusion protein displayed an envelope-associated localization across all conditions (Fig. 5D), which is in line with its predicted structure.

YbiJ is a small protein without notable predicted structural features, except that it contains a Sec secretion signal and likely localizes to the periplasm (Fig. 6A). YbiJ is a paralog of McbA, both containing a small domain of unknown function (DUF1471) (57). McbA is thought to be involved in the response to extracellular stress together with other DUF1471 proteins (57), although the exact role of these proteins remains unclear. Despite minimal sequence homology, YbiJ displays a distinct structural homology to RcsF (Fig. 6B and C; Table S2), the Bam complex substrate and outer membrane lipoprotein involved in envelope stress sensing (41, 42). YbiJ lacks the hydrophobic α-helical tail region of RcsF, suggesting it does not interact directly with the outer membrane. YbiJ-msfGFP displayed a faint envelope enrichment, but only in M9glu (as no signal could be detected in the three other nutrient conditions) (Fig. 6D). While the envelope-enriched localization pattern of the YbiJ-msfGFP fusion protein in M9glu is in line with the properties of YbiJ, we currently do not know why we did not observe this in other nutrient conditions, although nutrient condition-specific production of the protein could be involved.

YijD localizes in envelope-associated clusters enriched at the cell subpoles and the division septum

Examination of the predicted structure of YijD revealed properties in line with a membrane association, as YijD is predicted to be a small protein consisting of four hydrophobic and non-polar α-helices (Fig. 7A). Protein topology analysis revealed that helices can insert into and span the inner membrane (Fig. 7B). The YijD-msfGFP fusion protein displayed a consistent and specific localization across nutrient conditions (Fig. 7C). In each condition, YijD-msfGFP was clearly envelope-associated, in line with its structural predictions and reported inner membrane association (52), but did so in a diverse number of irregularly shaped clusters that appeared to be enriched at the (sub)polar regions of the cells and the division septum (Fig. 7C). To provide further insight into this localization pattern, we constructed demographs, which consist of linear representations of integrated fluorescence of cells sorted by their cell length (58). These representations further supported an enrichment of YijD-msfGFP clusters at the subpolar regions (instead of the cell pole), which became more pronounced in rich conditions (Fig. 7D). Near the very end of the cell division cycle, clusters also emerged at the dividing septum (where new cell poles are formed after completion of cell division) (Fig. 7D). While the structural properties of YijD explain its observed envelope-associated localization, the mechanisms and reasons underlying the clustering and localization currently remain enigmatic.

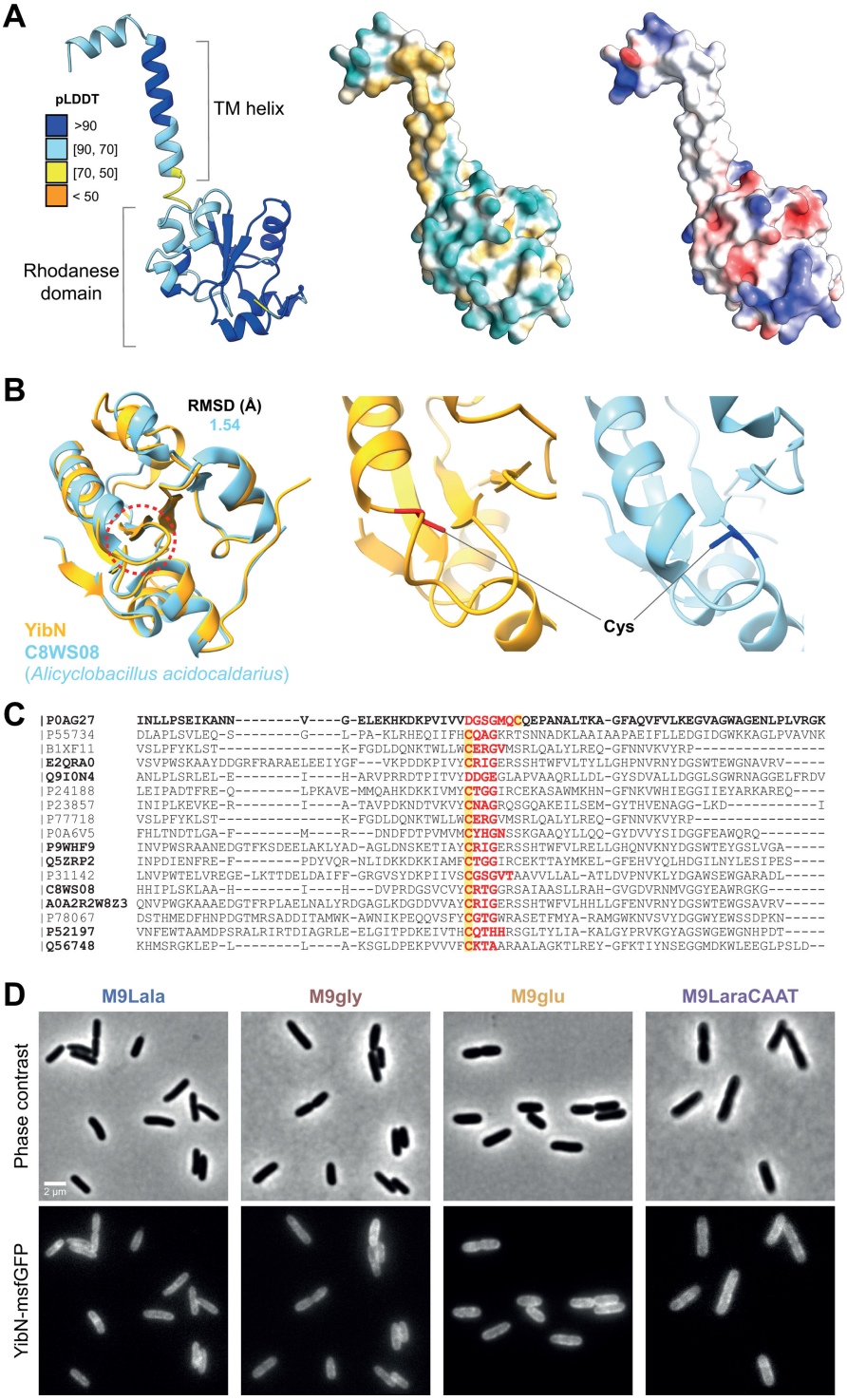


FIG 5 Structural properties and localization of YibN. (A) Structural depiction (ribbon) of YibN generated by AlphaFold2 and colored by pLDDT (left). Surface representations, colored depending on hydrophobicity (middle; yellow hydrophobic, cyan hydrophilic) or electrostatic surface potential (right; red negative, blue positive), are provided. (B) Aligned rhodanese domains from YibN (AF2; orange) and C8WS08 (PDB 3TP9; blue; a beta-lactamase and rhodanese domain protein from *Alicyclobacillus acidocaldarius*) (left). The enzymatic loop (i.e., the active site) has been indicated (red dashed circle). The cysteine residue and side chain are highlighted in red for YibN (middle) and blue for C8WS08 (right), showing that they lie on opposite ends of the enzymatic loop. The low root mean square deviation (RMSD) indicates the good quality of the alignment. (C) Aligned amino acid sequences of rhodanese domains. The first sequence is taken from YibN; other sequences depicted in bold were

(Continued on next page)

Fig 5 (Continued)

among the top 10 HHpred hits for YibN (Table S2). The residues comprising the enzymatic loop for each protein have been highlighted in red; the putative enzymatic cysteine is highlighted in yellow. (D) Representative phase contrast (top row) and epifluorescence (bottom row) images of the MG1655 *yibN::yibN-msfgfp* strain across the indicated nutrient conditions.

DISCUSSION

By further characterizing the deviating phenotypes of deletion strains across nutrient conditions, we were able to examine the nutrient dependency of phenotypic landscapes and identify three general principles. A first was that a spectrum of deviating phenotypes exists across nutrient conditions, ranging from nutrient- and/or feature-specific to more pleiotropic, with multiple phenotypic features being affected across nutrient conditions (Fig. 1B). A second finding was that phenotypic robustness was very limited across conditions, with the majority of deviating phenotypes not persisting across nutrient conditions (Fig. 1D). However, and this constitutes the third general principle, nutrient-independent phenotypes do exist and can be exploited to identify true genetic determinants of specific phenotypic features (Fig. 2). At the same time, we showed that similarities or dissimilarities in nutrient-dependent phenotypes of deletion strains, quantified by the correlation of their phenoprints, can be used to identify functional connections between genes and inform on their cellular function (Fig. 3). As such, microscopy-based phenoprint correlations provide an additional type of metric that enables investigation of protein interactions and function, complementary to existing databases (59, 60).

We further exemplified this by examining five specific genes of unknown function in *E. coli* (*yibN*, *yaaY*, *yfaQ*, *ybiJ*, and *yijD*) that led to notable phenotypes upon their deletion. Of these five, *yaaY* remains the most enigmatic. While deletion of *yaaY* led to a pronounced cell morphological and cell cycle phenotype, with more narrow cells and a later timing of FtsZ ring formation across nutrient conditions (Fig. 2C), the gene displayed very little conservation outside of *E. coli*. In addition, we failed to construct a fluorescent protein fusion, and its structural predictions yielded very poor results (Fig. S6). The protein has been computationally predicted to localize in the plasma membrane (61), yet this has not been experimentally verified. In contrast, we were able to experimentally verify the envelope-associated localization of the rhodanese domain-containing YibN (Fig. 5D). Structural predictions indicate it likely localizes there by inserting in the inner membrane using its N-terminal, hydrophobic, and non-polar α-helix (Fig. 5A). Recently, YibN has been identified as an interaction partner of YidC (62), modulating YidC's ability to insert proteins into the inner membrane and organize its lipids (63–66). This interaction and regulating role of YibN are in line with the upregulation of YibN upon YidC depletion (67), and *yibN* being in an operon together with *grxC*, *secB*, and *gpsA*, all proteins with envelope-related functions (68). While *yibN* deletion did not lead to observable growth defects across conditions (Fig. S2A) or temperatures (69), the deletion strain did display consistently wider cells across conditions (Fig. 2C). This provides further support for the importance of envelope integrity in determining cell width and highlights the potential of image-based profiling for uncovering phenotypes that previously remained undetected. At the same time, the exact molecular mechanisms and cellular role of YibN, together with the part played by its rhodanese domain and active site with cysteine residue (Fig. 5B and C), remain to be determined.

While YfaQ localization could not be experimentally verified using a fluorescent protein fusion, its predicted localization to the periplasm through a Sec signal peptide (Fig. S5), the phenoprint and connection network of its deletions strain (Fig. 3E; Fig. S2A), its homology determination (Table S2), and its structural predictions (Fig. S5) all point in the same direction: a role for the protein in maintaining envelope integrity by mediating peptidoglycan turnover or assembly. This is in line with its deletion leading to consistently wider cells (Fig. S2A). Another protein with a Sec signal sequence was YbiJ, although most notable insights into its cellular role came from its paralogues and

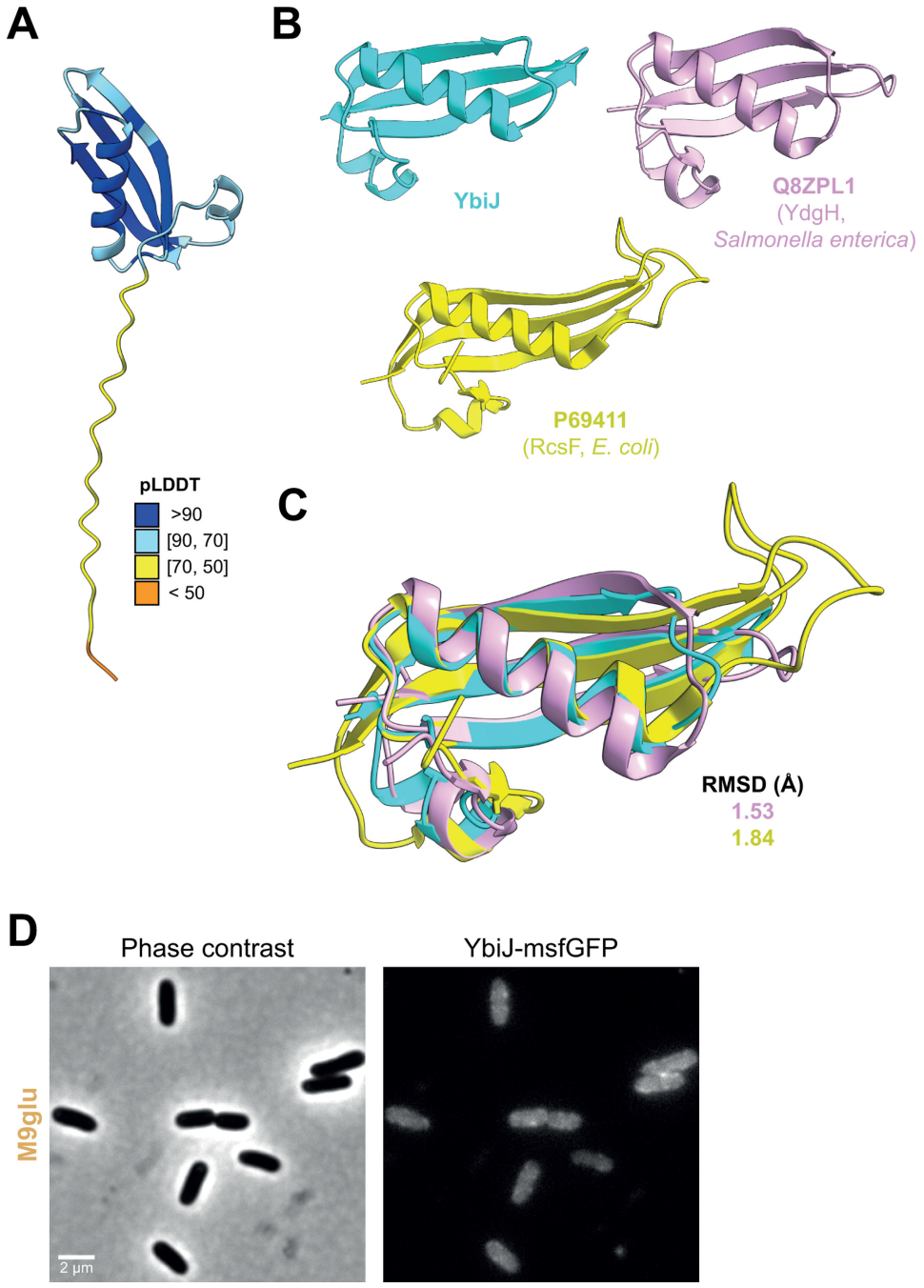


FIG 6 Structural properties and localization of YbiJ. (A) Structural depiction (ribbon) of YbiJ generated by AlphaFold2 and colored by pLDDT. The C-terminal peptide with pLDDT <70 comprises the secretion signal and cleavage site. (B) Structural depictions of the McbA-like domain of YbiJ (blue), YdgH (PDB 4EVU; pink) from *Salmonella enterica,* and RcsF (PDB 6T1W; yellow) from *E. coli*. Structural homologies were determined based on HHpred results. (C) Superimposed McbA-like domains showing structural homology. The low RMSDs indicate the good quality of the alignments. (D) Representative phase contrast (top) and epifluorescence (bottom) images of the MG1655 *ybiJ::ybiJ-msfgfp* strain in M9glu.

structural homology determination that revealed a striking structural homology with RcsF (Fig. 6A through C). Similar structural similarities between other DUF1471 proteins and RcsF have also been reported recently (70), and several of these proteins have been implicated in specific stress responses (71–73). In addition, the expression of *ybiJ* was also shown to be upregulated under acid stress (74) and exposure to ciprofloxacin (75),

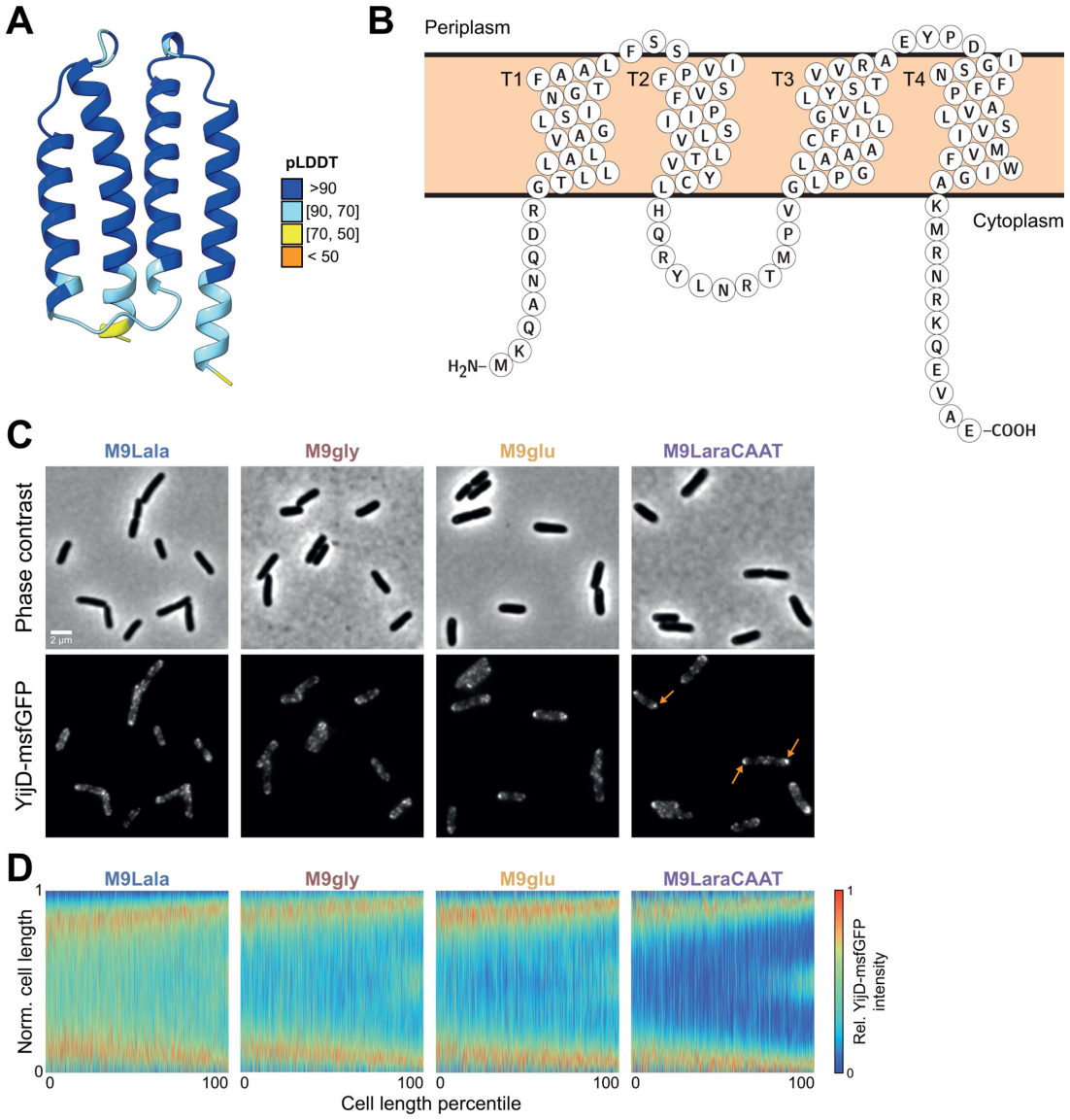


FIG 7 Localization and structural properties of YijD. (A) Structural representation (ribbon) of YijD protein generated using Alphafold2 and colored by pLDDT. (B) Protter representation of YijD homology illustrating transmembrane helices and cytoplasmic/periplasmic-exposed regions. (C) Representative phase contrast (top row) and epifluorescence (bottom row) images of the MG1655 *yijD::yijD-msfgfp* strain across the indicated nutrient conditions. Orange arrows highlight subpolar localization of YijD-msfGFP clusters in M9LaraCAAT. (D) Demographs of the YijD-msfGFP signal across the indicated nutrient conditions (n = 3,997, 3,121, 2,125, and 2,048 cells for M9Lala, M9gly, M9glu, and M9LaraCAAT, respectively). Cell length is normalized to show the relative cell length on the y-axis.

further hinting toward a role of this protein in stress sensing or mitigation. Although the precise contribution and molecular function of YbiJ and other DUF1471 proteins in these stress responses currently remain unclear, one intriguing hypothesis is that this could be mediated via the Rcs stress response, of which the activation is typically controlled by RcsF and its interaction with transmembrane outer membrane proteins (42).

YijD was the protein with the most distinct localization pattern, displaying a cluster-like organization that was enriched at the subpoles and the division septum (Fig. 7C and D). While displaying predicted structural properties that match an inner membrane localization, it remains to be determined how the heterogeneous and localized pattern is established. This could be an inherent property of the protein, or through interaction with other (membrane) proteins. The latter is in line with its phenoprint correlation network displaying connections to other proteins that are part of complexes that link

different parts of the cell envelope (Fig. 3E). Independently of how YijD localization is established, it is yet another factor contributing to the inherent intra-cellular heterogeneity in the composition of membranes (76). Additional factors that have been demonstrated to display a heterogeneous membrane-associated localization pattern similar to that of YijD (i.e., preferentially located in the cell poles and septum) are cardiolipin and other anionic phospholipid domains (77, 78). Moreover, in its genomic context, *yijD* is located in an operon together with *fabR*, a negative regulator of unsaturated fatty acid biosynthesis (79–81). Together, this indicates a potential role for YijD in fatty acid and membrane homeostasis, although the exact nature of this remains to be elucidated.

In general, our ability to assess the contribution of genetic variation to specific phenotypic features has improved throughout the years (82, 83). Yet, predicting phenotypes using genotypic information remains challenging. This is due to several factors including (i) pleiotropy, where a single genetic change affects two or more phenotypic traits (84), (ii) the polygenic or even omnigenic nature of phenotypic traits, where many genes (indirectly) contribute to variations in many phenotypic features (85), and (iii) the potential of the environment to modulate phenotypes (e.g., the expression of antibiotic resistance genes in bacteria or colony morphology in yeast [86, 87]). In this study, we highlight the potential of cross-condition image-based profiling, where many phenotypic readouts are considered simultaneously across multiple conditions, to identify general principles of cross-condition phenotypic variability and serve as a gateway for uncovering protein function.

MATERIALS AND METHODS

Data analysis

For our analyses, we started from an existing data set (22), which we expanded to 96 average population-level features by also including CVs for specific cellular properties (obtained by dividing the standard deviation for a feature by its mean). For all 96 features in a given nutrient condition, we calculated normalized scores (*s*) as before (22): $s = 1.35 \times [F_i - \text{median}\ (F_i^{\text{parent}})]/\text{iqr}(F_i^{\text{parent}})$. Here, F_i is the measured value of feature i in a nutrient condition, F_i^{parent} is the values of the parent strain replicates (≥115 replicates) for feature *i* in the same nutrient condition, and iqr is the interquartile range. As the interquartile range of normally distributed data is 1.35 times their standard deviation, we scaled the normalized scores by this factor to express them in terms of standard deviations from the median of the parent strain values.

Correlation coefficients

Spearman correlation coefficients (ρ) between phenoprints were calculated using MATLAB's built-in corr function. To determine a cutoff for correlation significance, we first tried a permutation-based approach as inreferences 13, 31 using 10,000 permutations to construct a background distribution of correlations. As this approach yielded cutoffs that were not sufficiently stringent, we chose to identify significant correlations between phenoprints of deletion strains using thresholds that exclude the middle 99.7% of all empirically obtained correlation coefficients: [−0.5154 0.6476] (Fig. 3B). This value corresponds to the data that falls within three standard deviations of the mean of an approximately normal distribution. For the generation of correlation networks, we used the built-in, edge-weighted, spring-embedded algorithm in Cytoscape (v3.10).

Strains and growth conditions

The parent strain for the existing data set was *E. coli* MG1655 *seqA::seqA-mcherry ftsZ::ftsZ-mvenus*^{SW}, containing chromosomally encoded, monomeric fluorescent protein fusions to a DNA replication marker (SeqA) and a divisome component (FtsZ) (22, 88). Bacterial strains and primers used in this study are listed in Tables S3 and S4, respectively. Strains were cultured in M9 medium, supplemented with 0.2% L-alanine (M9Lala),

glycerol (M9gly), glucose (M9glu), or L-arabinose, 0.1% casamino acids, 1 µg/mL thiamine (M9LaraCAAT). For microscopy, each strain was first grown to stationary phase in the appropriate nutrient condition in pre-culture tubes (3 mL of growth media) at 37°C under well-aerated conditions (200 rpm on an orbital shaker). A dilution (>1/10,000) was then used to inoculate fresh growth medium (3 mL). The resulting cultures were allowed to grow until they reached an optical density at 600 nm (OD_{600}) that corresponded to the early exponential phase (OD_{600} = 0.10–0.15) before being sampled for microscopy.

Construction of mutant strains

Mutant strains producing C-terminal translational fluorescent fusions to y-genes of interest were constructed by lambda Red-mediated recombination (89). Detailed construction procedures for each mutant strain are given in Table S3. All constructed mutants were initially confirmed by PCR with primer pairs attaching outside of the region where homologous recombination occurred (Table S4). Correct deletion or integration of PCR products was further verified by Sanger sequencing (LGC Genomics or Eurofins). Obtained reads were aligned to the predicted chromosomal sequences on Benchling using default parameters (Multisequence alignment, MAFFT v7), cleaned up, and checked for mutations.

Microscopy

Cells were imaged on 1.5% agarose (UltraPure Agarose, Invitrogen, Thermo Fisher Scientific, USA) pads, supplemented with M9 buffer. For each sample, 0.5 µL of cell culture was spotted on the pad, which was then covered with a cover glass (Fisherbrand Glass Square Coverslips, Fisher Scientific, Belgium). Per sample, 10 or 15 positions were imaged to ensure capture of a sufficient number of cells ($n \geq 2{,}048$ cells).

Imaging was performed on a Nikon Eclipse Ti2-E inverted epi-fluorescence microscope controlled by NIS-Elements AR software and equipped with a Plan Apo λ 100× DM oil objective (NA = 1.45; Nikon, the Netherlands), Ph3 condenser annulus (Nikon, the Netherlands), and a Kinetix A22J723021 sCMOS camera (Teledyne Photometrics, USA). For fluorescence excitation, a pE-800 light source (CoolLED, UK) with eight separate LED channels allowed narrowband excitation, and filter cubes with BrightLine Pinkel filter sets (penta-band set: LED-DA/FI/TR/Cy5/Cy7-5X-A; Semrock, IDEX Health and Science, USA) and OptoSpin filter wheels (Cairn GmbH, Germany) allowed capturing of narrowband fluorescent emission signals. The following optical configuration was used to acquire fluorescent images: GFP (excitation [CoolLED]: 470 nm, excitation filter: FF01-474/27, dichroic FF409/493/573/652/759-Di01, emitter filter cube: FF01-432/515/595/681/809, emitter filter wheel: FF01-515/30).

Conservation analysis of y-genes

In order to explore the prevalence of the selected y-genes and their orthologs, we used the EggNOG (v.5.0.0) (90) function Sequence search (http://eggnog5.embl.de/#/app/seqscan), in which we uploaded the reference proteins from *E. coli* MG1655 (YbiJ WP_206053394.1; YijD WP_000806411.1; YfaQ MCK2194123.1; YaaY MCN6283483.1; YibN WP_001156181.1). We curated the EggNOG outcomes for each of these proteins to examine whether an ortholog is reported for a given species (or genus). After this, we mapped the information about ortholog presence/absence to phylogenetic trees using the Interactive Tree of Life tool (iTOL v6) (91). For this, we adjusted the tree of life provided by iTOL (https://itol.embl.de/itol.cgi) to not contain multiple representatives of the same species. To visualize the orthologs' presence on a more nuanced level, we used another tree provided within EggNOG via the ETE toolkit implementation. The tree was calculated for YibN orthologs on a taxonomic level of the phylum of Proteobacteria (syn. Pseudomonadota). We filtered out cases of unspecific taxonomy (e.g., labeled only as alpha proteobacterium) and species that were represented by multiple strains to ensure only single species representatives. We chose the YibN-based tree as it was the

most abundant of the examined proteins. A consequence of this is that the visualized phylogenetic relations between species can differ from trees based on more data than a single protein, and that a YibN ortholog is automatically present in this visualization. Related to this, species associated with orthologs of one or more of the other four proteins, but not YibN, could have been omitted. The visualization in Fig. 4 does not depict the species name, yet this information can be found in Fig. S3 and S4.

The within-species conservation in *E. coli* was done on the level of genes corresponding to the reference proteins. The genes were implemented into a database, which was screened using ABRicate with a threshold of 85% for query coverage and identity against a large-scale *E. coli* genomic collection of 131,610 genomes (53). The outcomes were further processed in RStudio (v2023.06.0) with R (v4.4.1) to merge them with the appropriate genome metadata, to summarize the results, and to create figures. As *yfaQ* was the only gene that showed a slight decrease in prevalence among *E. coli* (i.e., present in ~98% of genomes), we additionally examined its absence patterns on the level of *E. coli* sequence types (STs). From 332 different STs that had at least one genome with *yfaQ* absence, we focused on 26 STs that had minimally 10 genomes without *yfaQ* (Fig. 4D).

Homology and structural predictions

Sequence similarity searches and homology detections were performed with the MPI Bioinformatics Toolkit (92), using HHpred (93, 94) comparison methods and the y-gene protein sequences provided by UniProt. Searches were performed against the PDB_mmCIF70 (HMM-based search and alignment generation against PDB chains; filtered for a maximum sequence identity of 70%) or Pfam-A databases. For more accurate prediction of transmembrane domains and signal peptides, DeepTMHMM 1.0 (95) and SignalP 6.0 (96) were used. The schematic representation of YijD topology was generated using Protter (97).

Structural predictions of y-gene protein products were generated using AlphaFold2 (AF2) through the web version of ColabFold v1.5.5 (98). Query protein sequences were again taken from the UniProt database, as above. Predictions were run for a maximum of 12 recycles. Structural similarity of predicted protein structures with homologous protein structures (as identified by HHpred searches) was quantified using the pairwise structural alignment tool from RCSB PDB (99).

Image analysis

Phase contrast images were processed using the "bact_phase_omni" model from Omnipose (100). This model was employed to identify cell outlines. To eliminate incorrectly detected cells, a support vector machine model was trained to distinguish single cells from erroneous detections (101). This approach ensured that only single cells were included in the analysis, thereby minimizing potential biases. From the correctly identified cell outlines, demographs, which are linear representations of the integrated fluorescence signal across the population of cells, sorted by cell length, were constructed.

ACKNOWLEDGMENTS

We thank Dr. Christine Jacobs-Wagner for inspiring this work, Dr. Seung Hyun Cho for scientific insights and discussions related to the structural similarity of YbiJ with RcsF, and Anjali Kiran Paranjape for performing preliminary experiments with fluorescent y-gene fusions. We would also like to thank the members of the Govers laboratory for fruitful discussions and critical reading of the manuscript.

K.S. was funded by a doctoral fellowship from the Research Foundation—Flanders (file number 1110925N). K.N. was funded by the European Union under Horizon Europe: MSCA Actions (Project 101105027) and by a junior postdoctoral fellowship from the Research Foundation—Flanders (file number 1251624N). This work was also supported

by a start-up grant (STG/21/068) from the KU Leuven Research Fund and by the European Union (ERC Starting Grant BacterialBlueprint 101040800).

AUTHOR AFFILIATION

[1]Department of Biology, KU Leuven, Leuven, Flanders, Belgium

AUTHOR ORCIDs

Kaat Sondervorst http://orcid.org/0000-0002-3787-7696
Sander K. Govers http://orcid.org/0000-0003-0260-4054

FUNDING

Funder	Grant(s)	Author(s)
Fonds Wetenschappelijk Onderzoek	1110925N	Kaat Sondervorst
Fonds Wetenschappelijk Onderzoek	1251624N	Kristina Nesporova
HORIZON EUROPE Marie Sklodowska-Curie Actions	101105027	Kristina Nesporova
Onderzoeksraad, KU Leuven	STG/21/068	Sander K. Govers
European Research Council	101040800	Kaat Sondervorst Kristina Nesporova Matthew Herdman Sander K. Govers

AUTHOR CONTRIBUTIONS

Kaat Sondervorst, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing | Kristina Nesporova, Formal analysis, Investigation, Visualization, Writing – review and editing | Matthew Herdman, Formal analysis, Investigation, Visualization, Writing – review and editing | Bart Steemans, Formal analysis, Investigation, Visualization, Writing – review and editing | Joëlle Rosseels, Formal analysis, Investigation | Sander K. Govers, Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All data sets used in this article are available in the supplemental information (Data set S1). All original code has been deposited at https://github.com/Govers-Lab/Sondervorst_et_al_2025.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Data set S1 (mSystems00206-25-s0001.xlsx). Extracted features for and normalized scores for deletion strains across nutrient conditions.
Supplemental Material (mSystems00206-25-s0002.pdf). Supplemental figures and tables.

REFERENCES

1. Raser JM, O'Shea EK. 2004. Control of stochasticity in eukaryotic gene expression. Science 304:1811–1814. https://doi.org/10.1126/science.1098641
2. Elowitz MB, Levine AJ, Siggia ED, Swain PS. 2002. Stochastic gene expression in a single cell. Science 297:1183–1186. https://doi.org/10.1126/science.1070919

3. Jorgensen P, Nishikawa JL, Breitkreutz B-J, Tyers M. 2002. Systematic identification of pathways that couple cell growth and division in yeast. Science 297:395–400. https://doi.org/10.1126/science.1070850
4. Smits WK, Kuipers OP, Veening J-W. 2006. Phenotypic variation in bacteria: the role of feedback regulation. Nat Rev Microbiol 4:259–271. https://doi.org/10.1038/nrmicro1381
5. Bock C, Datlinger P, Chardon F, Coelho MA, Dong MB, Lawson KA, Lu T, Maroc L, Norman TM, Song B, Stanley G, Chen S, Garnett M, Li W, Moffat J, Qi LS, Shapiro RS, Shendure J, Weissman JS, Zhuang X. 2022. High-content CRISPR screening. Nat Rev Methods Primers 2:8. https://doi.org/10.1038/s43586-022-00098-7
6. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006. https://doi.org/10.1038/msb4100050
7. Giaever G, Nislow C. 2014. The yeast deletion collection: a decade of functional genomics. Genetics 197:451–465. https://doi.org/10.1534/genetics.114.161620
8. Rousset F, Cui L, Siouve E, Becavin C, Depardieu F, Bikard D. 2018. Genome-wide CRISPR-dCas9 screens in *E. coli* identify essential genes and phage host factors. PLoS Genet 14:e1007749. https://doi.org/10.1371/journal.pgen.1007749
9. Yasir M, Turner AK, Bastkowski S, Baker D, Page AJ, Telatin A, Phan M-D, Monahan L, Savva GM, Darling A, Webber MA, Charles IG. 2020. TraDIS-Xpress: a high-resolution whole-genome assay identifies novel mechanisms of triclosan action and resistance. Genome Res 30:239–249. https://doi.org/10.1101/gr.254391.119
10. Coward C, Dharmalingham G, Abdulle O, Avis T, Beisken S, Breidenstein E, Carli N, Figueiredo L, Jones D, Khan N, Malara S, Martins J, Nagalingam N, Turner K, Wain J, Williams D, Powell D, Mason C. 2020. High-density transposon libraries utilising outward-oriented promoters identify mechanisms of action and resistance to antimicrobials. FEMS Microbiol Lett 367:fnaa185. https://doi.org/10.1093/femsle/fnaa185
11. Kazi MI, Schargel RD, Boll JM. 2020. Generating transposon insertion libraries in gram-negative bacteria for high-throughput sequencing. J Vis Exp 61612. https://doi.org/10.3791/61612
12. Inoue T, Shingaki R, Hirose S, Waki K, Mori H, Fukui K. 2007. Genome-wide screening of genes required for swarming motility in *Escherichia coli* K-12. J Bacteriol 189:950–957. https://doi.org/10.1128/JB.01294-06
13. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, Lee S, Kazmierczak KM, Lee KJ, Wong A, Shales M, Lovett S, Winkler ME, Krogan NJ, Typas A, Gross CA. 2011. Phenotypic landscape of a bacterial cell. Cell 144:143–156. https://doi.org/10.1016/j.cell.2010.11.052
14. Qimron U, Marintcheva B, Tabor S, Richardson CC. 2006. Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. Proc Natl Acad Sci U S A 103:19039–19044. https://doi.org/10.1073/pnas.0609428103
15. Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C, Kost-Alimova M, Gustafsdottir SM, Gibson CC, Carpenter AE. 2016. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc 11:1757–1774. https://doi.org/10.1038/nprot.2016.105
16. Caicedo JC, Singh S, Carpenter AE. 2016. Applications in image-based profiling of perturbations. Curr Opin Biotechnol 39:134–142. https://doi.org/10.1016/j.copbio.2016.04.003
17. Tran T-A, Sridhar S, Reece ST, Lunguya O, Jacobs J, Van Puyvelde S, Marks F, Dougan G, Thomson NR, Nguyen BT, Bao PT, Baker S. 2024. Combining machine learning with high-content imaging to infer ciprofloxacin susceptibility in isolates of *Salmonella* Typhimurium. Nat Commun 15:5074. https://doi.org/10.1038/s41467-024-49433-4
18. Seal S, Trapotsi M-A, Spjuth O, Singh S, Carreras-Puigvert J, Greene N, Bender A, Carpenter AE. 2025. Cell Painting: a decade of discovery and innovation in cellular imaging. Nat Methods 22:254–268. https://doi.org/10.1038/s41592-024-02528-8
19. Campos M, Govers SK, Irnov I, Dobihal GS, Cornet F, Jacobs-Wagner C. 2018. Genomewide phenotypic analysis of growth, cell morphogenesis, and cell cycle events in *Escherichia coli*. Mol Syst Biol 14:e7573. https://doi.org/10.15252/msb.20177573
20. Ursell T, Lee TK, Shiomi D, Shi H, Tropini C, Monds RD, Colavin A, Billings G, Bhaya-Grossman I, Broxton M, Huang BE, Niki H, Huang KC. 2017. Rapid, precise quantification of bacterial cellular dimensions across a genomic-scale knockout library. BMC Biol 15:17. https://doi.org/10.1186/s12915-017-0348-8
21. French S, Côté J-P, Stokes JM, Truant R, Brown ED. 2017. Bacteria getting into shape: genetic determinants of *E. coli* morphology. mBio 8:e01977-16. https://doi.org/10.1128/mBio.01977-16
22. Govers SK, Campos M, Tyagi B, Laloux G, Jacobs-Wagner C. 2024. Apparent simplicity and emergent robustness in the control of the *Escherichia coli* cell cycle. Cell Syst 15:19–36. https://doi.org/10.1016/j.cels.2023.12.001
23. Tong M, French S, El Zahed SS, Ong WK, Karp PD, Brown ED. 2020. Gene dispensability in *Escherichia coli* grown in thirty different carbon environments. mBio 11:e02259-20. https://doi.org/10.1128/mBio.02259-20
24. Ghatak S, King ZA, Sastry A, Palsson BO. 2019. The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. Nucleic Acids Res 47:2446–2454. https://doi.org/10.1093/nar/gkz030
25. Hu P, Janga SC, Babu M, Díaz-Mejía JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasseri NK, Musso G, Ali M, Nazemof N, Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G, Emili A. 2009. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. PLoS Biol 7:e96. https://doi.org/10.1371/journal.pbio.1000096
26. Rudd KE. 1998. Linkage map of *Escherichia coli* K-12, edition 10: the physical map. Microbiol Mol Biol Rev 62:985–1019. https://doi.org/10.1128/MMBR.62.3.985-1019.1998
27. Chang Y-C, Hu Z, Rachlin J, Anton BP, Kasif S, Roberts RJ, Steffen M. 2016. COMBREX-DB: an experiment centered database of protein function: knowledge, predictions and knowledge gaps. Nucleic Acids Res 44:D330–D335. https://doi.org/10.1093/nar/gkv1324
28. Deutschbauer A, Price MN, Wetmore KM, Tarjan DR, Xu Z, Shao W, Leon D, Arkin AP, Skerker JM. 2014. Towards an informative mutant phenotype for every bacterial gene. J Bacteriol 196:3643–3655. https://doi.org/10.1128/JB.01836-14
29. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, Carlson HK, Esquivel Z, Sadeeshkumar H, Chakraborty R, Zane GM, Rubin BE, Wall JD, Visel A, Bristow J, Blow MJ, Arkin AP, Deutschbauer AM. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature 557:503–509. https://doi.org/10.1038/s41586-018-0124-0
30. Deutschbauer A, Price MN, Wetmore KM, Shao W, Baumohl JK, Xu Z, Nguyen M, Tamse R, Davis RW, Arkin AP. 2011. Evidence-based annotation of gene function in *Shewanella oneidensis* MR-1 using genome-wide fitness profiling across 121 conditions. PLoS Genet 7:e1002385. https://doi.org/10.1371/journal.pgen.1002385
31. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo B-M, Marta E, Shiver AL, Whitehead EH, Weissman JS, Brown ED, Qi LS, Huang KC, Gross CA. 2016. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell 165:1493–1506. https://doi.org/10.1016/j.cell.2016.05.003
32. Szczepaniak J, Webby MN. 2024. The Tol Pal system integrates maintenance of the three layered cell envelope. NPJ Antimicrob Resist 2:46. https://doi.org/10.1038/s44259-024-00065-0
33. Egan AJF. 2018. Bacterial outer membrane constriction. Mol Microbiol 107:676–687. https://doi.org/10.1111/mmi.13908
34. Prieto AI, Kahramanoglou C, Ali RM, Fraser GM, Seshasayee ASN, Luscombe NM. 2012. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. Nucleic Acids Res 40:3524–3537. https://doi.org/10.1093/nar/gkr1236
35. Rice PA, Yang S, Mizuuchi K, Nash HA. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell 87:1295–1306. https://doi.org/10.1016/s0092-8674(00)81824-3
36. Freundlich M, Ramani N, Mathew E, Sirko A, Tsui P. 1992. The role of integration host factor in gene expression in *Escherichia coli*. Mol Microbiol 6:2557–2563. https://doi.org/10.1111/j.1365-2958.1992.tb01432.x
37. Mitchell AM, Srikumar T, Silhavy TJ. 2018. Cyclic enterobacterial common antigen maintains the outer membrane permeability barrier of *Escherichia coli* in a manner controlled by YhdP. mBio 9:e01321-18. https://doi.org/10.1128/mBio.01321-18
38. Konovalova A, Kahne DE, Silhavy TJ. 2017. Outer membrane biogenesis. Annu Rev Microbiol 71:539–556. https://doi.org/10.1146/annurev-micro-090816-093754

39. Voulhoux R, Bos MP, Geurtsen J, Mols M, Tommassen J. 2003. Role of a highly conserved bacterial protein in outer membrane protein assembly. Science 299:262–265. https://doi.org/10.1126/science.1078973
40. Wu T, Malinverni J, Ruiz N, Kim S, Silhavy TJ, Kahne D. 2005. Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. Cell 121:235–245. https://doi.org/10.1016/j.cell.2005.02.015
41. Laloux G, Collet J-F. 2017. Major tom to ground control: how lipoproteins communicate extracytoplasmic stress to the decision center of the cell. J Bacteriol 199:e00216-17. https://doi.org/10.1128/JB.00216-17
42. Cho S-H, Dekoninck K, Collet J-F. 2023. Envelope-stress sensing mechanism of Rcs and Cpx signaling pathways in Gram-negative bacteria. J Microbiol 61:317–329. https://doi.org/10.1007/s12275-023-00030-y
43. Tata M, Konovalova A. 2019. Improper coordination of BamA and BamD results in bam complex jamming by a lipoprotein substrate. mBio 10:e00660-19. https://doi.org/10.1128/mBio.00660-19
44. Cho S-H, Szewczyk J, Pesavento C, Zietek M, Banzhaf M, Roszczenko P, Asmar A, Laloux G, Hov A-K, Leverrier P, Van der Henst C, Vertommen D, Typas A, Collet J-F. 2014. Detecting envelope stress by monitoring β-barrel assembly. Cell 159:1652–1664. https://doi.org/10.1016/j.cell.2014.11.045
45. Konovalova A, Perlman DH, Cowles CE, Silhavy TJ. 2014. Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of β-barrel proteins. Proc Natl Acad Sci U S A 111:E4350–E4358. https://doi.org/10.1073/pnas.1417138111
46. Bougdour A, Cunning C, Baptiste PJ, Elliott T, Gottesman S. 2008. Multiple pathways for regulation of σ^S (RpoS) stability in *Escherichia coli* via the action of multiple anti - adaptors. Mol Microbiol 68:298–313. https://doi.org/10.1111/j.1365-2958.2008.06146.x
47. Battesti A, Hoskins JR, Tong S, Milanesio P, Mann JM, Kravats A, Tsegaye YM, Bougdour A, Wickner S, Gottesman S. 2013. Anti-adaptors provide multiple modes for regulation of the RssB adaptor protein. Genes Dev 27:2722–2735. https://doi.org/10.1101/gad.229617.113
48. Gravina F, Sanchuki HS, Rodrigues TE, Gerhardt ECM, Pedrosa FO, Souza EM, Valdameri G, de Souza GA, Huergo LF. 2018. Proteome analysis of an *Escherichia coli ptsN*-null strain under different nitrogen regimes. J Proteomics 174:28–35. https://doi.org/10.1016/j.jprot.2017.12.006
49. Sánchez-Cañizares C, Prell J, Pini F, Rutten P, Kraxner K, Wynands B, Karunakaran R, Poole PS. 2020. Global control of bacterial nitrogen and carbon metabolism by a PTS^{Ntr}-regulated switch. Proc Natl Acad Sci U S A 117:10234–10245. https://doi.org/10.1073/pnas.1917471117
50. Paradis-Bleau C, Kritikos G, Orlova K, Typas A, Bernhardt TG. 2014. A genome-wide screen for bacterial envelope biogenesis mutants identifies a novel factor involved in cell wall precursor metabolism. PLoS Genet 10:e1004056. https://doi.org/10.1371/journal.pgen.1004056
51. Asmar AT, Collet J-F. 2018. Lpp, the Braun lipoprotein, turns 50—major achievements and remaining issues. FEMS Microbiol Lett 365. https://doi.org/10.1093/femsle/fny199
52. Daley DO, Rapp M, Granseth E, Melén K, Drew D, von Heijne G. 2005. Global topology analysis of the *Escherichia coli* inner membrane proteome. Science 308:1321–1323. https://doi.org/10.1126/science.1109730
53. Nesporova K, Govers SK. 2025. Large-scale genomic analysis reveals the distribution and diversity of Type VI Secretion Systems in *Escherichia coli*. bioRxiv. https://doi.org/10.1101/2025.01.23.634537
54. Cipollone R, Ascenzi P, Visca P. 2007. Common themes and variations in the rhodanese superfamily. IUBMB Life 59:51–59. https://doi.org/10.1080/15216540701206859
55. Alsohaibani R, Claudel A-L, Perchat-Varlet R, Boutserin S, Talfournier F, Boschi-Muller S, Selles B. 2023. Rhodanese-fold containing proteins in humans: not just key players in sulfur trafficking. Antioxidants (Basel) 12:843. https://doi.org/10.3390/antiox12040843
56. Bordo D, Bork P. 2002. The rhodanese/Cdc25 phosphatase superfamily. Sequence-structure-function relations. EMBO Rep 3:741–746. https://doi.org/10.1093/embo-reports/kvf150
57. Eletsky A, Michalska K, Houliston S, Zhang Q, Daily MD, Xu X, Cui H, Yee A, Lemak A, Wu B, Garcia M, Burnet MC, Meyer KM, Aryal UK, Sanchez O, Ansong C, Xiao R, Acton TB, Adkins JN, Montelione GT, Joachimiak A, Arrowsmith CH, Savchenko A, Szyperski T, Cort JR. 2014. Structural and functional characterization of DUF1471 domains of *Salmonella* proteins SrfN, YdgH/SssB, and YahO. PLoS One 9:e101787. https://doi.org/10.1371/journal.pone.0101787
58. Hocking J, Priyadarshini R, Takacs CN, Costa T, Dye NA, Shapiro L, Vollmer W, Jacobs-Wagner C. 2012. Osmolality-dependent relocation of penicillin-binding protein PBP2 to the division site in *Caulobacter crescentus*. J Bacteriol 194:3116–3127. https://doi.org/10.1128/JB.00260-12
59. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, von Mering C. 2023. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 51:D638–D646. https://doi.org/10.1093/nar/gkac1000
60. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roechert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H. 2007. IntAct—open source resource for molecular interaction data. Nucleic Acids Res 35:D561–D565. https://doi.org/10.1093/nar/gkl958
61. Peabody MA, Laird MR, Vlasschaert C, Lo R, Brinkman FSL. 2016. PSORTdb: expanding the bacteria and archaea protein subcellular localization database to better reflect diversity in cell envelope structures. Nucleic Acids Res 44:D663–D668. https://doi.org/10.1093/nar/gkv1271
62. Zhao Z, Yamamoto N, Young JW, Solis N, Fong A, Al-Seragi M, Kim S, Aoki H, Phanse S, Le H-T, Overall CM, Nishikawa H, Babu M, Nishiyama K-I, Duong van Hoa F. 2025. YibN, a bona fide interactor of the bacterial YidC insertase with effects on membrane protein insertion and membrane lipid production. J Biol Chem 301:108395. https://doi.org/10.1016/j.jbc.2025.108395
63. Serdiuk T, Steudle A, Mari SA, Manioglu S, Kaback HR, Kuhn A, Müller DJ. 2019. Insertion and folding pathways of single membrane proteins guided by translocases and insertases. Sci Adv 5:eaau6824. https://doi.org/10.1126/sciadv.aau6824
64. van der Laan M, Bechtluft P, Kol S, Nouwen N, Driessen AJM. 2004. F_1F_0 ATP synthase subunit c is a substrate of the novel YidC pathway for membrane protein biogenesis. J Cell Biol 165:213–222. https://doi.org/10.1083/jcb.200402100
65. Li D, Rocha-Roa C, Schilling MA, Reinisch KM, Vanni S. 2024. Lipid scrambling is a general feature of protein insertases. Proc Natl Acad Sci U S A 121:e2319476121. https://doi.org/10.1073/pnas.2319476121
66. Dalbey RE, Kuhn A, Zhu L, Kiefer D. 2014. The membrane insertase YidC. Biochim Biophys Acta 1843:1489–1496. https://doi.org/10.1016/j.bbamcr.2013.12.022
67. Wickström D, Wagner S, Simonsson P, Pop O, Baars L, Ytterberg AJ, van Wijk KJ, Luirink J, de Gier J-WL. 2011. Characterization of the consequences of YidC depletion on the inner membrane proteome of *E. coli* using 2D blue native/SDS-PAGE. J Mol Biol 409:124–135. https://doi.org/10.1016/j.jmb.2011.03.068
68. Crane JM, Randall LL. 2017. The sec system: protein export in *Escherichia coli*. EcoSal Plus 7. https://doi.org/10.1128/ecosalplus.esp-0002-2017
69. Zhao Z, Yamamoto N, Young JW, Solis N, Fong A, Al-Seragi M, Kim S, Aoki H, Phanse S, Le H-T, Overall CM, Nishikawa H, Babu M, Nishiyama K, Duong van Hoa F. 2024. YibN, a bona fide interactor of YidC with implications in membrane protein insertion and membrane lipid production. bioRxiv. https://doi.org/10.1101/2024.12.13.628402
70. Meredith JD, Gray MJ. 2024. Genome-wide characterization of hypothiocyanite stress response of *Escherichia coli* reveals a potential role for RclB and RclC in lowering cell envelope permeability. bioRxiv. https://doi.org/10.1101/2024.12.06.627277
71. Sargentini NJ, Gularte NP, Hudman DA. 2016. Screen for genes involved in radiation survival of *Escherichia coli* and construction of a reference database. Mutat Res 793–794:1–14. https://doi.org/10.1016/j.mrfmmm.2016.10.001
72. Lee J, Hiibel SR, Reardon KF, Wood TK. 2010. Identification of stress - related proteins in *Escherichia coli* using the pollutant *cis* - dichloroethylene. J Appl Microbiol 108:2088–2102. https://doi.org/10.1111/j.1365-2672.2009.04611.x
73. Zhang X-S, García-Contreras R, Wood TK. 2008. *Escherichia coli* transcription factor YncC (McbR) regulates colanic acid and biofilm formation by repressing expression of periplasmic protein YbiM (McbA). ISME J 2:615–631. https://doi.org/10.1038/ismej.2008.24

74. Schumacher K, Gelhausen R, Kion-Crosby W, Barquist L, Backofen R, Jung K. 2023. Ribosome profiling reveals the fine-tuned response of *Escherichia coli* to mild and severe acid stress. mSystems 8:e01037-23. https://doi.org/10.1128/msystems.01037-23
75. Goswami M, Narayana Rao A. 2018. Transcriptome profiling reveals interplay of multifaceted stress response in *Escherichia coli* on exposure to glutathione and ciprofloxacin. mSystems 3:e00001-18. https://doi.org/10.1128/mSystems.00001-18
76. Matsumoto K, Hara H, Fishov I, Mileykovskaya E, Norris V. 2015. The membrane: transertion as an organizing principle in membrane heterogeneity. Front Microbiol 6:572. https://doi.org/10.3389/fmicb.2015.00572
77. Mileykovskaya E, Dowhan W. 2009. Cardiolipin membrane domains in prokaryotes and eukaryotes. Biochim Biophys Acta 1788:2084–2091. https://doi.org/10.1016/j.bbamem.2009.04.003
78. Renner LD, Weibel DB. 2011. Cardiolipin microdomains localize to negatively curved regions of *Escherichia coli* membranes. Proc Natl Acad Sci U S A 108:6264–6269. https://doi.org/10.1073/pnas.1015757108
79. Zhang Y-M, Marrakchi H, Rock CO. 2002. The FabR (YijC) transcription factor regulates unsaturated fatty acid biosynthesis in *Escherichia coli*. J Biol Chem 277:15558–15565. https://doi.org/10.1074/jbc.M201399200
80. Zhu K, Zhang Y-M, Rock CO. 2009. Transcriptional regulation of membrane lipid homeostasis in *Escherichia coli*. J Biol Chem 284:34880–34888. https://doi.org/10.1074/jbc.M109.068239
81. Feng Y, Cronan JE. 2011. Complex binding of the FabR repressor of bacterial unsaturated fatty acid biosynthesis to its cognate promoters. Mol Microbiol 80:195–218. https://doi.org/10.1111/j.1365-2958.2011.07564.x
82. Power RA, Parkhill J, de Oliveira T. 2017. Microbial genome-wide association studies: lessons from human GWAS. Nat Rev Genet 18:41–50. https://doi.org/10.1038/nrg.2016.132
83. Richardson TG, Harrison S, Hemani G, Davey Smith G. 2019. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. Elife 8:e43657. https://doi.org/10.7554/eLife.43657
84. Stearns FW. 2010. One hundred years of pleiotropy: a retrospective. Genetics 186:767–773. https://doi.org/10.1534/genetics.110.122549
85. Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell 169:1177–1186. https://doi.org/10.1016/j.cell.2017.05.038
86. Hughes D, Andersson DI. 2017. Environmental and genetic modulation of the phenotypic expression of antibiotic resistance. FEMS Microbiol Rev 41:374–391. https://doi.org/10.1093/femsre/fux004
87. Granek JA, Magwene PM. 2010. Environmental and genetic determinants of colony morphology in yeast. PLoS Genet 6:e1000823. https://doi.org/10.1371/journal.pgen.1000823
88. Gray WT, Govers SK, Xiang Y, Parry BR, Campos M, Kim S, Jacobs-Wagner C. 2019. Nucleoid size scaling and intracellular organization of translation across bacteria. Cell 177:1632–1648. https://doi.org/10.1016/j.cell.2019.05.017
89. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645. https://doi.org/10.1073/pnas.120163297
90. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148
91. Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res 52:W78–W82. https://doi.org/10.1093/nar/gkae268
92. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007
93. Hildebrand A, Remmert M, Biegert A, Söding J. 2009. Fast and accurate automatic structure prediction with HHpred. Proteins 77:128–132. https://doi.org/10.1002/prot.22499
94. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. 2019. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinform 20:473. https://doi.org/10.1186/s12859-019-3019-7
95. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh A, Winther O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. https://doi.org/10.1101/2022.04.08.487609
96. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40:1023–1025. https://doi.org/10.1038/s41587-021-01156-3
97. Omasits U, Ahrens CH, Müller S, Wollscheid B. 2014. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 30:884–886. https://doi.org/10.1093/bioinformatics/btt607
98. Kim G, Lee S, Levy Karin E, Kim H, Moriwaki Y, Ovchinnikov S, Steinegger M, Mirdita M. 2025. Easy and accurate protein structure prediction using ColabFold. Nat Protoc 20:620–642. https://doi.org/10.1038/s41596-024-01060-5
99. Bittrich S, Segura J, Duarte JM, Burley SK, Rose Y. 2024. RCSB protein Data Bank: exploring protein 3D similarities via comprehensive structural alignments. Bioinformatics 40:btae370. https://doi.org/10.1093/bioinformatics/btae370
100. Cutler KJ, Stringer C, Lo TW, Rappez L, Stroustrup N, Brook Peterson S, Wiggins PA, Mougous JD. 2022. Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. Nat Methods 19:1438–1448. https://doi.org/10.1038/s41592-022-01639-4
101. Steemans B, Govers SK. 2024. Protocol to train a support vector machine for the automatic curation of bacterial cell detections in microscopy images. STAR Protoc 5:102868. https://doi.org/10.1016/j.xpro.2024.102868