International Journal of
Molecular Sciences

MDPI

Article

Sulfoquinovose Catabolism in *E. coli* Strains: Compositional and Functional Divergence of *yih* Gene Cassettes

Anna D. Kaznadzey [1,2,*,†], Anna A. Rybina [1,†], Tatiana A. Bessonova [2,3], Dmitriy S. Korshunov [4], Maria N. Tutukina [1,2,‡] and Mikhail S. Gelfand [1,‡]

[1] Center for Molecular and Cellular Biology, Moscow 121205, Russia; rybinaann@gmail.com (A.A.R.); masha306@gmail.com (M.N.T.); mikhail.gelfand@gmail.com (M.S.G.)
[2] Vavilov Institute for General Genetics of Russian Academy of Sciences, Moscow 117971, Russia; tatianabessonova66@gmail.com
[3] Institute of Cell Biophysics of Russian Academy of Sciences, Pushchino 142290, Russia
[4] Biology Faculty, Lomonosov Moscow State University, Moscow 119991, Russia; korshunds@yandex.ru
* Correspondence: kaznadzey@vigg.ru
† These authors contributed equally to this work.
‡ These authors also contributed equally to this work.

Abstract

The sulfo-Embden–Meyerhof–Parnas (sulfo-EMP) pathway enables *Escherichia coli* to utilize sulfoquinovose, (SQ) a sulfonated sugar derived from plant sulfolipids, as a carbon source. This pathway is encoded by the *yih* gene cassette. However, structural and functional diversity of this cassette across *E. coli* strains has not been fully characterized. We identified two structural variants of the *yih* cassette across *E. coli* and *Shigella* strains: a long form (*ompL-yihOPQRSTUVW*) and a truncated short form (*yihTUVW*). Both forms occupy the same genomic location but differ in orientation and are scattered across the phylogenetic tree, suggesting frequent recombination events. Transcriptome analyses revealed that only the long cassette, as found in *E. coli* K-12 MG1655, is transcriptionally induced during growth on sulfoquinovose. The short cassette, found in *E. coli* Nissle 1917 and other host-adapted strains, showed no differential expression. Despite this, both strains grew comparably on sulfoquinovose, indicating different metabolic adaptation strategies. Gene expression profiling revealed shared stress responses but distinct central metabolic patterns. Electrophoretic mobility shift assays further demonstrated that the transcription factor YihW from Nissle 1917 binds its DNA targets with lower affinity than the homolog from K-12 and shows weaker sulfoquinovose-dependent dissociation.

Keywords: sulfoquinovose; *Escherichia coli*; Nissle 1917; RNA-seq; bacterial adaptation; *yih* cassette; operon evolution; sulfur metabolism

check for updates

Received: 14 September 2025
Revised: 21 October 2025
Accepted: 22 October 2025
Published: 24 October 2025

Citation: Kaznadzey, A.D.; Rybina, A.A.; Bessonova, T.A.; Korshunov, D.S.; Tutukina, M.N.; Gelfand, M.S. Sulfoquinovose Catabolism in *E. coli* Strains: Compositional and Functional Divergence of *yih* Gene Cassettes. *Int. J. Mol. Sci.* **2025**, *26*, 10351. https://doi.org/10.3390/ijms262110351

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Sulfoquinovose is a sulfonated derivative of glucose that constitutes the hydrophilic head group of sulfoquinovosyl diacylglycerol (SQDG), a major sulfolipid synthesized by photosynthetic organisms such as algae, higher plants, and cyanobacteria [1,2]. An estimated ten billion tons of SQDG are produced annually, representing a significant reservoir of organic sulfur. Bacteria are the main degraders of SQDG and SQ, with six major catabolism pathways currently known [3,4]. These include two glycolytic-like pathways, sulfo-Embden–Meyerhof–Parnas and sulfo-Entner–Doudoroff; two pentose phosphate pathway-like routes, the sulfo-transaldolase and sulfo-transketolase pathways; and two

non-glycolytic catabolic pathways, sulfo-ASMO and sulfo-ASDO, which use mono- or dioxygenases to cleave the C–S bond of SQ via alkanesulfonate monooxygenase or dioxygenase, respectively. Among these, the sulfo-EMP pathway is best characterized and serves as a model for SQ degradation [5], particularly in *Escherichia coli* [6].

The sulfo-EMP pathway converts SQ to sulfofructose and then to sulfolactaldehyde and dihydroxyacetone phosphate, the latter entering central glycolysis. Sulfolactaldehyde is further reduced to 2,3-dihydroxypropane-1-sulfonate, completing dissimilation of the sulfonate moiety without desulfonation. A cluster of up to ten genes is responsible for these functions in *E. coli*, referred to as the *yih* cassette. The complete *yih* cassette includes *ompL* (encoding a porin), *yihO* (MFS transporter), *yihP* (MFS transporter), *yihQ* (sulfoquinovosidase), *yihR* (SQ epimerase), *yihS* (SQ isomerase), *yihV* (sulfofructose kinase), *yihT* (sulfofructose-1-phosphate aldolase), *yihU* (sulfolactaldehyde reductase), and *yihW*, alternatively named *csqR* (transcriptional regulator). The respective enzymatic functions allow for conversion of SQ into intermediates of the central carbon metabolism, as described above, ultimately enabling the organism to exploit sulfosugars as a source of carbon and energy. Unlike pathways found in *Pseudomonas putida* and *Paracoccus pantotrophus*, which can mineralize the sulfonate group [5,7], the *E. coli* pathway retains the sulfur in a stable end product (2,3-dihydroxypropane-1-sulfonate), which is subsequently excreted [6].

In total, the sulfo-EMP pathway yields only one pyruvate and one ATP per SQ molecule, providing half the energy and carbon building blocks compared to glycolysis [6]. The NADH/NAD^+ balance remains unaffected due to the reaction catalyzed by YihU (sulfolactaldehyde reductase). Moreover, the sulfo-EMP pathway does not produce glucose-6-phosphate (G6P) and fructose-6-phosphate (F6P), which are essential for anabolic processes and the pentose phosphate pathway. During growth on SQ as the sole source of carbon, this imposes an additional metabolic burden on the cell, which has to find alternative ways to generate G6P and F6P. In a recent study, the key role of gluconeogenesis in this regard was proposed [8]. The gluconeogenesis pathway compensates for the deficiency of G6P or F6P and might also be accompanied by a shift in metabolic flux away from the tricarboxylic acid cycle.

YihW (CsqR), a transcriptional regulator encoded within the cassette, was identified as a local repressor of the *yih* cassette genes [9,10]. YihW belongs to the DeoR subfamily of GntR-family transcription factors, many of which regulate carbohydrate metabolism in response to small-molecule ligands [11]. Electrophoretic mobility shift assays (EMSA) established that YihW binds to the intergenic regions upstream of its own gene (*yihW*) and of the divergent promoters for *yihV* and *yihU*. SQ inhibits this DNA-binding activity [10], consistent with a model in which SQ acts as an effector that alleviates transcriptional repression in response to substrate availability. In [10], high concentrations of purified YihW protein were needed to bind its DNA targets in the EMSA assays, that possibly reflected the oligomeric nature of this class of regulators, although the precise oligomeric state of YihW remains unresolved.

While the sulfo-EMP pathway is well characterized in the laboratory *E. coli K-12 MG1655* strain, the extent of its variability across *E. coli* lineages and its functional regulation in distinct ecological or pathogenic contexts remain unclear. In particular, little is known about how differences in gene content and organization of the *yih* cassette may influence its transcriptional control and metabolic role. In this study, we compare *E. coli* strains representing distinct cassette architectures to determine how genomic organization affects transcriptional responsiveness to SQ and regulatory binding of the transcription factor YihW (CsqR).

2. Results

2.1. Long and Short yih Cassettes in E. coli Strains

To assess the variety of the *yih cassettes* across *E. coli* strains and *Shigella* species, we performed homology-based search complemented by gene neighborhood analysis to define cassette boundaries and structural variants. Two distinct types of cassettes were observed, a long variant (*ompL-yihOPQRSTUVW*, first described in *E. coli* K-12 [6] and a short variant (*yihTUVW*), lacking, in particular, genes encoding the initial steps of SQ catabolism, sulfoquinovosidase (*yihQ*), mutarotase (*yihR*), and isomerase (*yihS*). In all analyzed *E. coli* strains carrying the short variant, no homologs of other long cassette genes (*yihO*, *yihP*, *yihQ*, *yihR*, *yihS*) were detected elsewhere in the genome. The short variant is observed in a variety of strains, mostly pathogenic, and, interestingly, in a widely used probiotic strain *E. coli* Nissle 1917 (Figure 1). A significant association was found between the short variant and pathogenic strains (odds ratio = 2.67, $p = 1.1 \times 10^{-5}$), indicating enrichment in pathogens but not exclusivity. Its presence in *E. coli* Nissle 1917, a member of phylogroup B2 like many extraintestinal pathogenic *E. coli* (ExPEC) strains, may reflect underlying genomic similarity to ExPEC lineages [12].

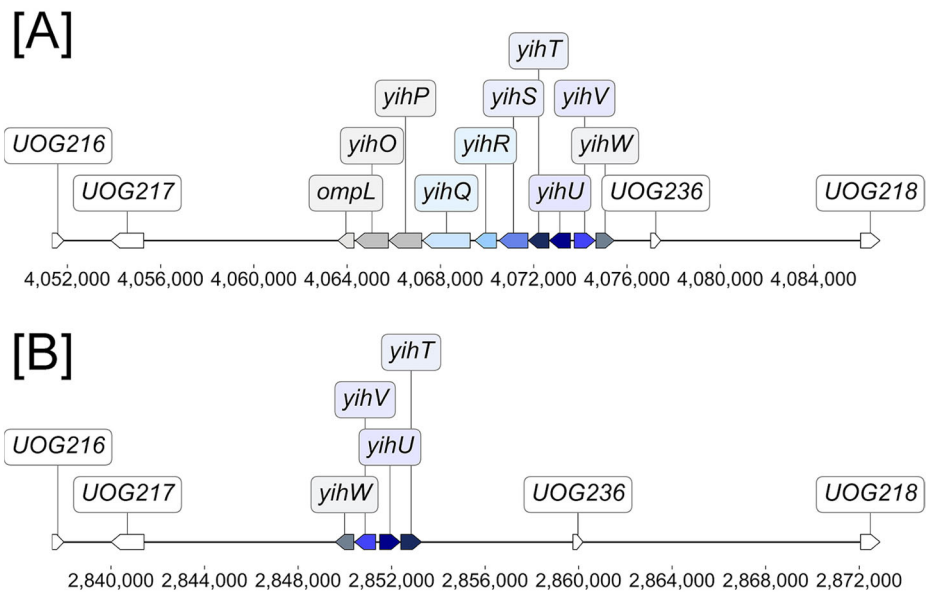


Figure 1. Position of *yih cassettes* relative to UOGs: (**A**) Most common configuration of the long cassette, represented by *E. coli* K-12 MG1655; (**B**) Most common configuration of the short cassette, represented by *E. coli* Nissle 1917.

We analyzed the genomic context of the *yih* cassette in 413 strains using positions of 238 universal single-copy orthologous groups (UOGs) as reference markers with positions highly conserved across most *E. coli* genomes [13]. Specifically, we identified the nearest flanking UOGs upstream and downstream of each *yih* cassette to evaluate the positional consistency across strains and detect possible insertions, deletions, or inversions.

The majority of *yih* cassettes, both short and long, is flanked by genes from UOG217 and UOG236 (Figure 1), indicating that the locus has a conserved genomic position across strains. Notably, the orientation of the short cassette is reversed relative to the long variant.

Analysis of the distribution of cassette types on a phylogenetic tree of *E. coli* and *Shigella* strains (Figure 2) demonstrated that strains with short and long cassettes were interspersed and both types of cassettes were observed in all but one phylogroups (Figure 2). This suggests that the *yih* cassette is an evolutionarily labile module, subject to recurrent rearrangement via homologous recombination or short-distance horizontal gene transfer (HGT) events.

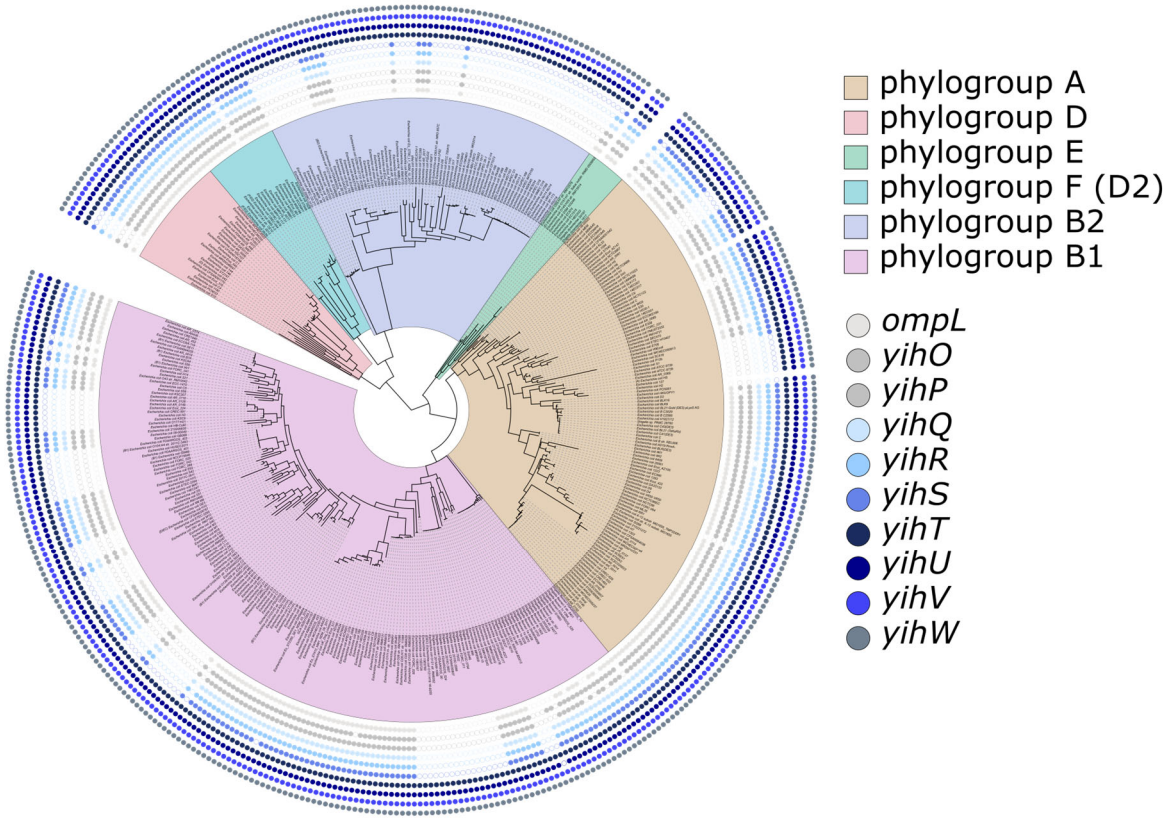


Figure 2. Phylogenetic tree of *Escherichia coli* and *Shigella* spp. Phylogroups are marked by different colors. The presence of *yih* genes is indicated by colored dots, revealing two distinct types: a long variant (e.g., *ompLyihOPQRTSUVW* in *E.coli* K12 MG1655) and a short variant (e.g., *yihTUVW* in *E. coli* Nissle 1917).

2.2. Transcriptional Changes in E. coli Strains with Short and Long yih Cassettes During Growth on Sulfoquinovose

To characterize the transcriptional response of the long and short *yih* cassette genes to SQ, we cultivated *E. coli* K-12 MG1655 and *E. coli* Nissle 1917 in minimal media with SQ or glucose as a sole carbon source to mid-exponential phase and subjected total mRNA to RNA-seq analysis.

We identified differentially expressed genes (DEGs); data obtained with growth on glucose was used as a control (Figure 3). In total, during growth on SQ, *E. coli* K-12 MG1655 had 979 DEGs, of which 236 were upregulated and 743 were downregulated, while *E. coli* Nissle 1917 had 991 DEGs, of which 225 were upregulated and 766 were downregulated. The biological replicates showed consistent expression patterns across the samples (Supplementary Figure S1).

Among the top DEGs in *E. coli* K-12 MG1655 growing on SQ, we observed genes from the *yih* cassette with a Log2FC (Log2 fold change) five or greater, including transporter genes (*ompL*, *yihO*, and *yihP*), genes encoding sulfo-EMP enzymes isomerase *yihS*, aldolase *yihT*, reductase *yihU*, kinase *yihV*, SQ mutarotase *yihR*, and the transcriptional regulator, *yihW* (*csqR*) (Figure 3, Supplementary Table S4). In contrast, *yih* genes of the short cassette in *E. coli* Nissle 1917 did not demonstrate any differential expression in response to SQ (Figure 3, Supplementary Table S3).

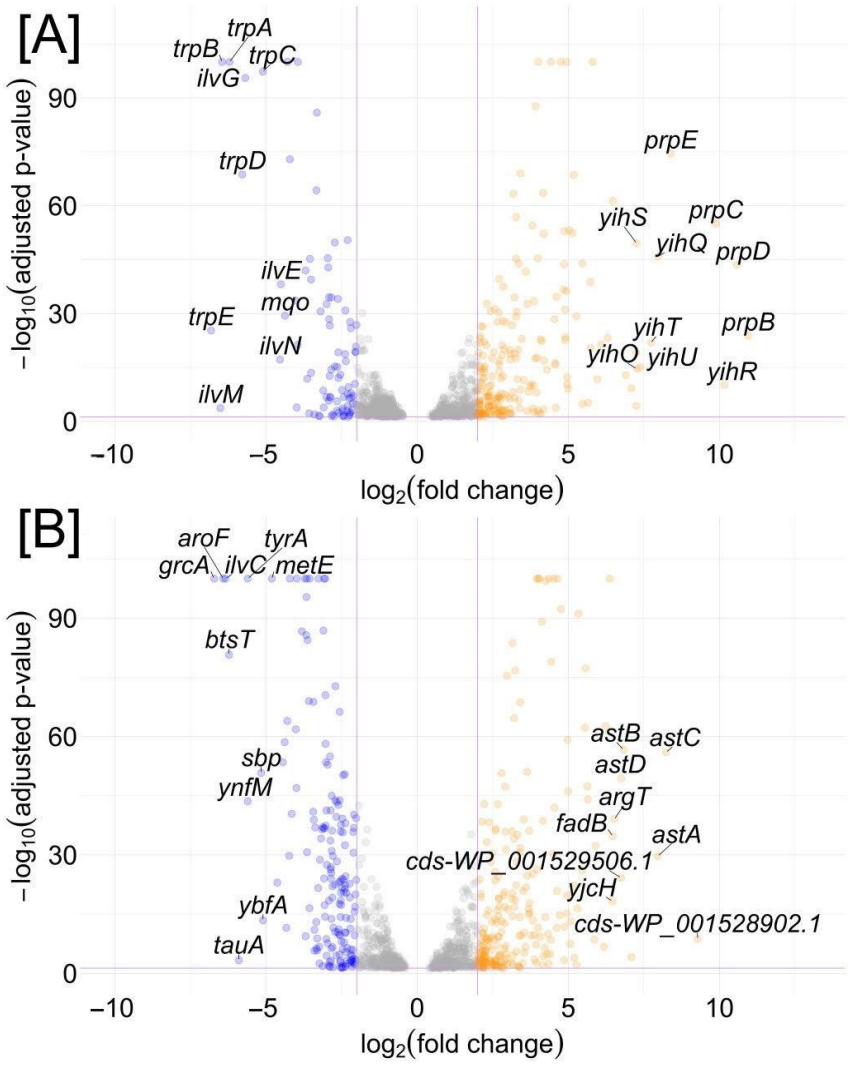


Figure 3. Distribution of DEGs in *E. coli* K-12 (**A**) and *E. coli* Nissle 1917 (**B**) during growth on SQ compared to growth on glucose. Upregulated genes are marked orange, downregulated genes are marked blue, genes without significant changes in expression are marked gray. The top ten DEGs are labeled. The x-axis represents $\log_2$(fold change), the y-axis represents the $-\log_{10}$(adjusted *p*-value). Thresholds for significance are indicated by violet dashed lines at $\log_2$(fold change) = ± 2 and $-\log_{10}$(adjusted *p*-value) ≈ 1.3 (corresponding to an adjusted *p*-value of 0.05).

To validate the RNA-seq results, expression of the *yih cassette* genes was assessed by qRT-PCR after four hours of growth on SQ or glucose for both *E. coli* strains (Figure 4). In *E. coli* K-12 MG1655, the *yihP*, *yihQ*, *yihR*, *yihS*, *yihT*, *yihU*, *yihV*, and *yihW* (*csqR*) genes were significantly upregulated on SQ. No activation of expression was observed for *yihV*, *yihT*, and *yihW* (*csqR)* in *E. coli* Nissle 1917 (Figure 4), while *yihU* was apparently repressed.

Both *E. coli* strains exhibit low growth rates on SQ (Figure 5). As previously reported [6], the biomass of *E. coli* K-12 grown on SQ during mid-log phase was approximately half that of cultures grown on glucose under the same carbohydrate concentrations (OD_{600}~0.3 vs. OD_{600}~0.6 after four hours of growth) (Figure 5). This can be partially attributed to the fact that the sulfo-EMP pathway generates only half the ATP and carbon yield compared to glycolysis [6].

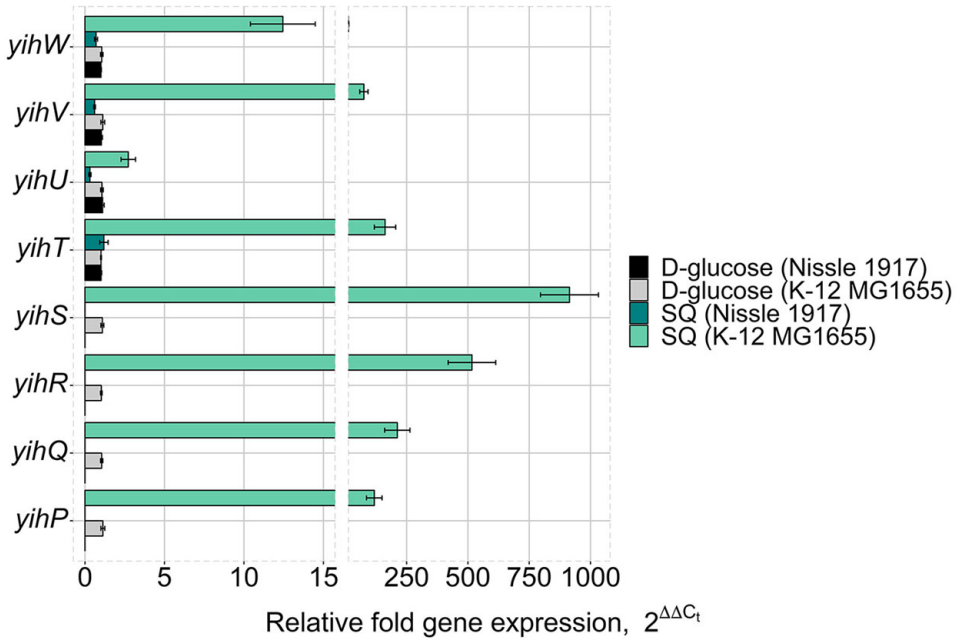


Figure 4. Expression of the *yih* cassette genes in *E. coli* K-12 MG1655 and *E. coli* Nissle 1917 after four hours of growth on D-glucose or SQ as a sole carbon source. Standard deviations were calculated based on three biological and four technical replicates.

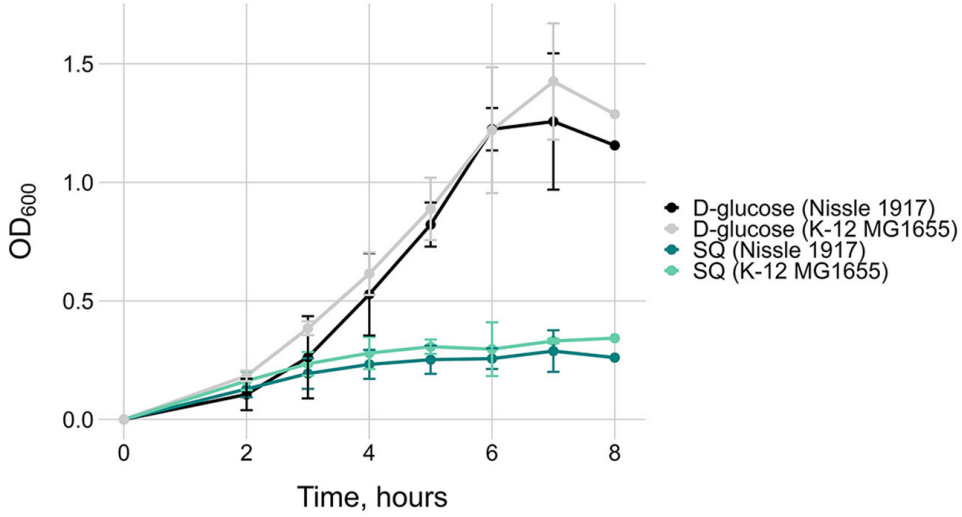


Figure 5. Growth of *E. coli* K-12 MG1655 and *E. coli* Nissle 1917 strains in the presence of 0.2% D-glucose or SQ as sole carbon source. Mean values and 95% confidence intervals were obtained on three biological replicates.

To assess whether the *E. coli* Nissle 1917 strain used other genes to utilize SQ using, we identified 25 DEGs whose protein products were similar to known bacterial SQ catabolism enzymes. However, these candidates were never co-localized, thus not forming operons or cassettes, and the sequence identity between respective proteins was never higher than 37% suggesting a metabolic function other than utilization of SQ (Supplementary Table S2).

Despite transcriptional inactivity of the short *yih* cassette genes, *E. coli* Nissle 1917 maintained growth rates on SQ similar to those of K-12 MG1655, suggesting that it relies on a different catabolic mechanism. Our next objective was to identify similar or contrasting gene expression patterns between the *E. coli* K-12 and Nissle 1917 strains by first identifying putative orthologs using reciprocal best-hit search, followed by gene set enrichment analysis.

2.3. Shared and Divergent Gene Expression Patterns in E. coli K-12 and Nissle 1917 in Response to SQ

Both *E. coli* strains grown on sulfoquinovose exhibited transcriptomic signatures consistent with adaptation to nutrient-limited conditions. These included repression of anabolic pathways such as amino acid and vitamin biosynthesis; induction of catabolic programs for alternative carbon sources, notably fatty acids, acetate, and propionate; up-regulation of nitrogen-scavenging systems, including arginine and putrescine degradation; and activation of stationary-phase responses, reflected in the expression of genes regulated by RpoS [14] (Figure 6, Supplementary Table S4).

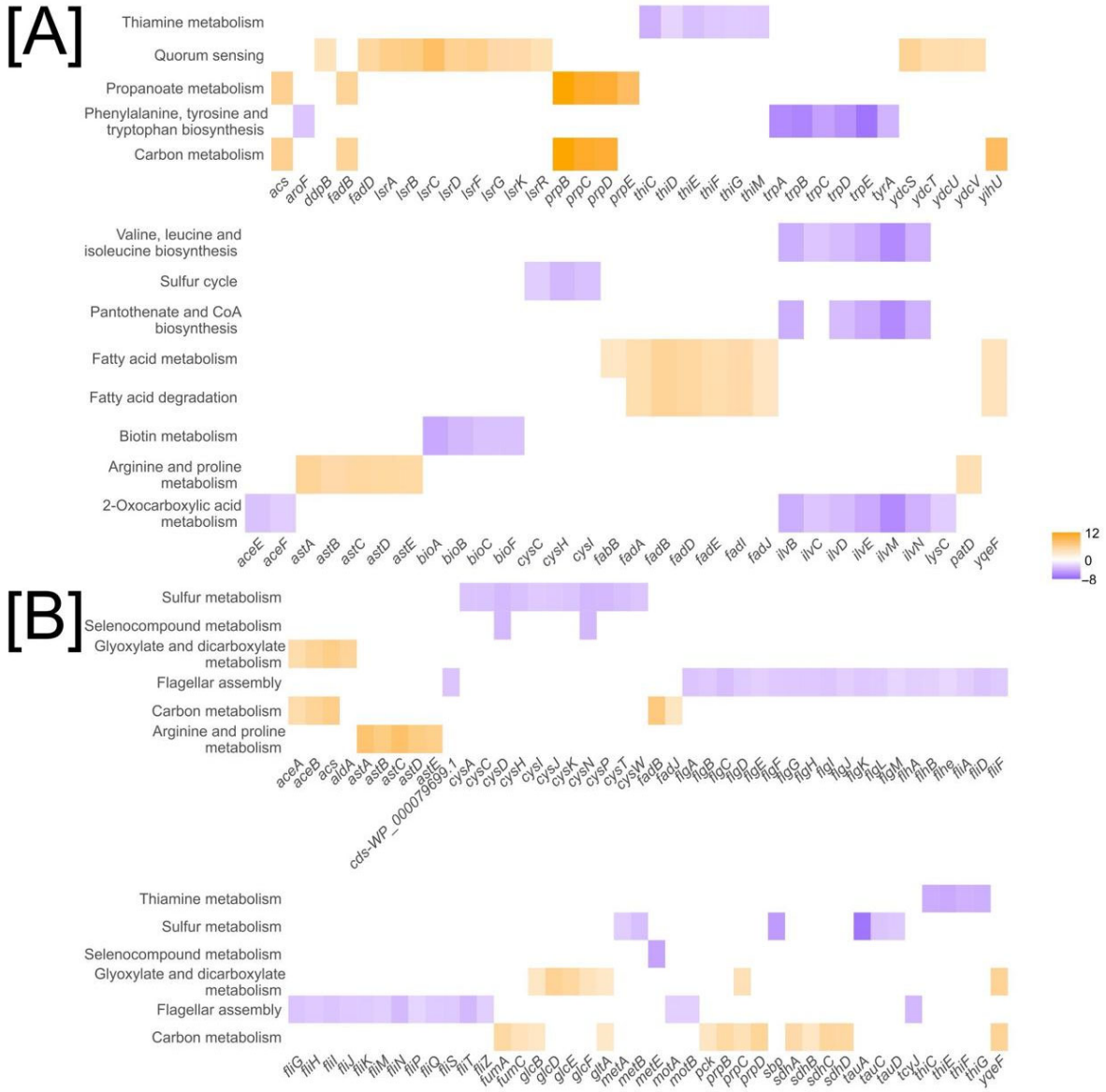


Figure 6. Enriched KEGG pathways in *E. coli* K-12 MG1655 (**A**) and *E. coli* Nissle 1917 (**B**) grown on SQ. Gene categories with the most enriched pathways are colored based on their $\log_2$(fold change) values, where blue indicates downregulation, orange indicates upregulation, and gray represents mid-range expression levels.

More specifically, both strains showed strong upregulation of genes involved in fatty acid (*fadAB*, *fadH*, *fadL*, *fadJI*, *fadD*, *fadM*, *fadE*) and propionate (*prpBCDE*) degradation (Figure 6). Concomitantly, the *aceEF* operon encoding the pyruvate dehydrogenase complex was downregulated.

Nitrogen metabolism was similarly reprogrammed. The *astCADBE* and *ydcSTUV-patD* operons, belonging to the Ntr-Nac regulon, were highly upregulated (Log_2FC 3–9), indicating reliance on alternative nitrogen sources. Biosynthetic genes for glutamate (*gltBD*), serine (*serA*), aromatic amino acids (*aroL/F*, *tyrA*), and branched-chain amino acids (*ilv*) were downregulated. Sulfur metabolism (*cysPUWAM*, *cysCND*, *cysHIJ*), thiamine biosynthesis (*thiCEFSGH*), and, in *E. coli* Nissle 1917, taurine transport (*tauA*, *tauC*, *tauD*) were suppressed.

Responses of the central carbon metabolism genes differed between the strains.In *E. coli* K-12, genes coding for aconitase (*acnAB*) and fumarase (*fumC)* were upregulated, while malate:quinone oxidoreductase (*mqo*), isocitrate dehydrogenase (*icd*), and dihydrolipoamide dehydrogenase (*lpd*) were downregulated (Supplementary Figure S2). We also observed downregulation of cytochrome o oxidase (*cyoABCD*) and upregulation of fumarate reductase (*frdABCD*), as well as glycerol catabolism genes (*glpABC*, *glpFKX*), gluconeogenesis marker *glpX* [15], and *treC*.

In contrast, *E. coli* Nissle 1917 featured upregulation of central carbon metabolism genes, including *icd*, and moderate upregulation of *cyoABCDE* (Supplementary Figure S3 and Table S4). We observed high expression of glyoxylate shunt (*aceAB*, *glcB*) and glycolate metabolism genes (*glcEFGBA*), as well as upregulation of phosphoenolpyruvate carboxykinase (*pckA*), fructose-1,6-bisphosphatase (*fbp*), and phosphoenolpyruvate synthase (*ppsA*).

In K-12, genes of curli biosynthesis (*csgDEFG*) and quorum-sensing regulators (*lsrABCD*, *lsrRK*) were strongly upregulated, (Supplementary Figure S4). Nissle 1917 showed no differential expression of these genes but demonstrated strong downregulation of flagellar assembly genes (*fliC*, *flgK*, *motAB*, Supplementary Figure S5).

2.4. Binding of YihW (CsqR) to Intergenic Regions of Long and Short yih Cassettes

To determine the binding properties of YihW from Nissle 1917 and its responsiveness to SQ, we performed EMSA with the *yihU/V* and *yihV/W* intergenic regions from *E. coli* K-12 MG1655 (long cassette) and Nissle 1917 (short cassette).

Similarly to the results previously obtained by Shimada et al. [10], YihW from both *E. coli* Nissle 1917 and K-12 MG1655 formed high molecular weight complexes with its DNA targets, possibly due to protein oligomerization, and these complexes dissociated in the presence of SQ acting as effector (Figures 7 and 8). In our experiment, molar excess of the protein required to bind all DNA at 30 °C for the *yihU/V* intergenic region was 24:1 (Figure 7A). At 37 °C, binding was less effective (Figure 7B); even at 48x molar excess of the protein, some DNA remained free. Interestingly, at 30 °C, YihW from Nissle 1917 interacted with *yihU/V* from the parent strain and *E. coli* K-12 MG1655 with equal efficiency (Figure 7A), but at 37 °C, binding to the K-12 fragment was stronger (Figure 7B). YihW (CsqR) from *E. coli* K-12 MG1655 bound all *yihU/V* DNA at 16x molar excess independent of the strain (Figure 7C), suggesting that in Nissle 1917, YihW is less involved in the *yih* cassette regulation compared to K-12, especially at 37 °C.

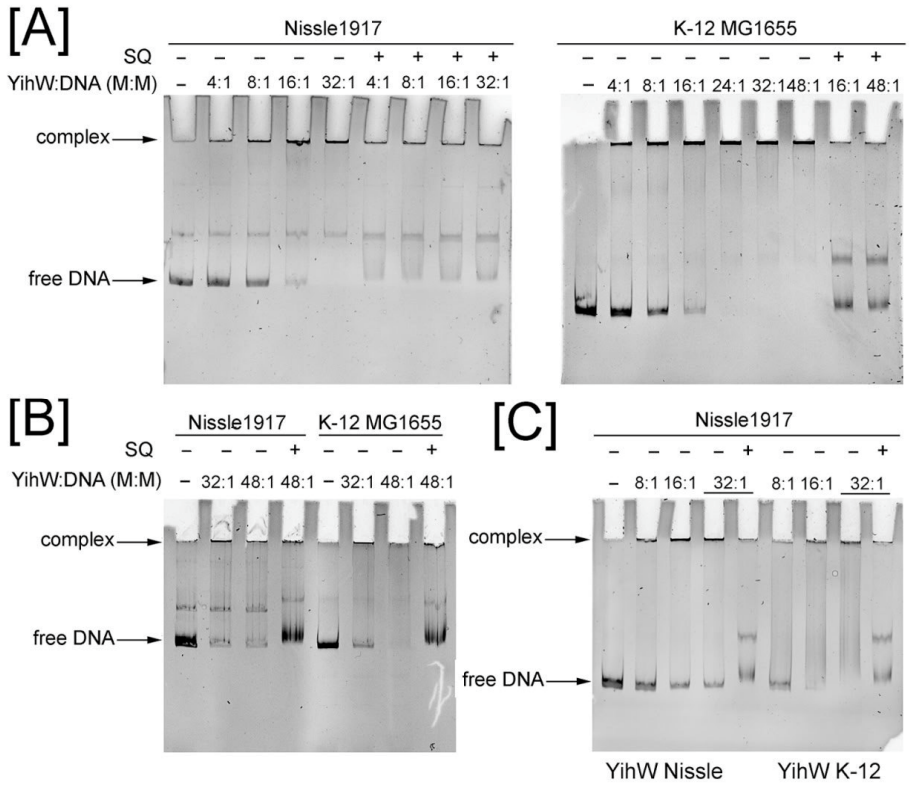


Figure 7. Binding of YihW from *E. coli* Nissle 1917 to the *yihU/V* intergenic region from *E. coli* Nissle 1917 and *E. coli* K-12 MG1655. (**A**) 30 °C (**B**) 37 °C. The YihW:DNA molar ratio, template strains, and the presence of SQ are indicated above the lanes. (**C**) Comparison of binding properties of YihW from *E. coli* Nissle 1917 and YihW (CsqR) from *E. coli* K-12 MG1655 with the *yihU/V* intergenic region from *E. coli* Nissle 1917 at 37 °C. The protein origin is indicated below the lanes.

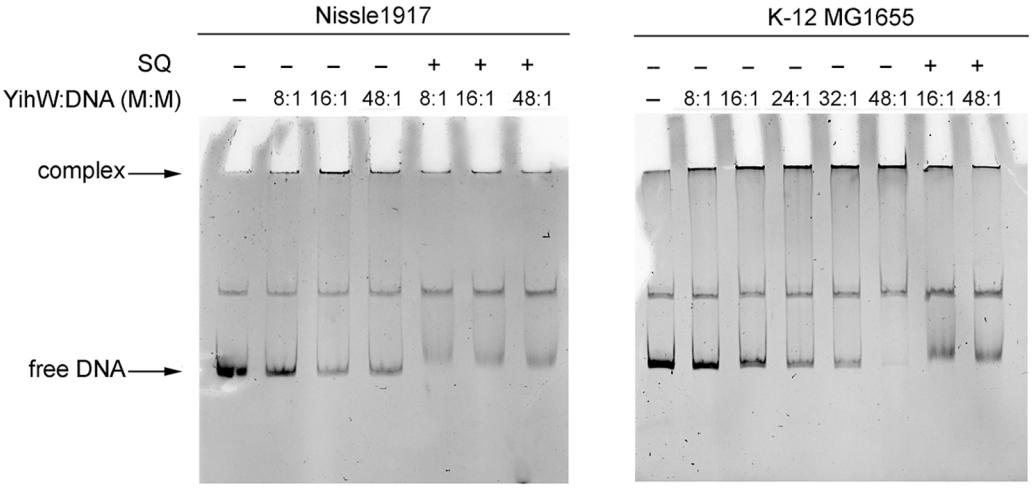


Figure 8. Binding of YihW from *E. coli* Nissle 1917 to the *yihV/W* intergenic region from *E. coli* Nissle1917 and *E. coli* K-12 MG1655 at 30 °C. The YihW:DNA molar ratio, template strains, and the presence of SQ are indicated above the lanes.

This assumption is in line with the saturation effect of binding with the *yihV/W* region from the parent strain (Figure 8), suggesting that autoregulatory binding may be inherently weaker or more tightly controlled. It is worth noting that the effect of SQ on complex dissociation between YihW from Nissle 1917 and both intergenic fragments from the parent strain was lower as compared to the fragments from K-12 (Figures 7A and 8).

3. Discussion

The *yih* cassette encodes the sulfo-EMP pathway responsible for SQ degradation in *E. coli* K-12 [6]. Here, we report variability in gene content and organization within the cassette across environmental and clinical *E. coli* isolates. The cassette has two major organizational variants: a long form (*ompL-yihOPQRSTUVW*), characteristic of laboratory strains such as *E. coli* K-12 MG1655, and a short form (*yihTUVW*), found in many strains, the majority of which are pathogenic. Interestingly, one of the non-pathogenic strains, containing a short *yih cassette*, is *E. coli* Nissle 1917, widely marketed as a probiotic.

Analyses of phylogeny and synteny (Figures 1 and 2) demonstrate that both variants share a conserved genomic location, flanked by universal orthologous groups, indicating homologous origin. The interspersed distribution of cassette types across phylogenetic clades supports the possibility of horizontal gene transfer and recombination events. The presence of both cassette types in closely related strains further underscores the evolutionary lability of the *yih* cassette and suggests that, in particular, recombination hotspots may facilitate cassette remodeling driven by niche-specific metabolic needs [16].

The short cassettes, which are often found in pathogenic or host-adapted strains, appear in a reversed orientation compared to the long cassettes. This suggests that most likely an inversion and further homologous recombination events have occurred at this locus. Whether short cassettes confer the ability to catabolize SQ remained unclear.

RNA-seq and qRT-PCR analyses (Figures 3 and 4) demonstrated that genes of the long cassette in K-12 MG1655 were transcriptionally responsive to SQ, with significant upregulation of all genes, including those encoding key enzymes of the SQ catabolism (*yihS*, *yihQ*, *yihR*, *yihT*, *yihV*, and, to lower extent, *yihU*) and associated transporters (*yihO*, *yihP*, *ompL*). The gene of transcriptional regulator *yihW (csqR)* was also upregulated, consistent with substrate-induced derepression. These results align with previous studies that identified YihW as a repressor of the *yih* genes [9,10].

In contrast, genes of the short cassette in *E. coli* Nissle 1917 showed no differential expression during growth on SQ. Neither *yihV*, *yihT*, *yihU*, nor *yihW* were significantly upregulated, and no alternative operons with homologous genes and relevant expression patterns were identified to suggest compensation via a different SQ catabolism pathway. This transcriptional inactivity suggests that the short cassette in *E. coli* Nissle 1917 is no longer involved in SQ degradation. Yet, both *E. coli* K-12 MG1655 and *E. coli* Nissle 1917 showed low, but comparable growth on SQ (Figure 5), suggesting alternative metabolic adaptations in *E. coli* Nissle 1917, possibly relying on gluconeogenesis, fatty acid catabolism, or other pathways. Retention of the long *yih* cassette in environmental strains such as *E. coli* K-12 MG1655, coupled with its active transcriptional response to SQ, supports the hypothesis that these strains exploit plant-derived sulfosugars like SQ, nutrients that are abundant in soil or plant-associated niches.

Gene-set enrichment analysis revealed shared signatures of nutrient stress in both strains during SQ utilization. Biosynthetic genes for amino acids (e.g., *serA*, *gltBD*, *aroL*, *ilv*, *cys*) and cofactors (e.g., *thi* operon) were consistently downregulated, while catabolic pathways for fatty acids (fad genes), propionate (*prpBCDE*), and nitrogen scavenging via arginine and putrescine degradation (*astCADBE*, *ydcSTUV-patD*) were activated. These patterns are consistent with RpoS and Ntr-Nac regulon activation and reflect a shift toward endogenous resource recycling under carbon and nitrogen limitation [14,15]. Suppression of *cys* and taurine transport genes may further indicate sulfur reallocation under stress [17].

Despite these similarities, the strains diverged in their central metabolic responses. *E. coli* K-12 MG1655 showed markers of anaerobic adaptation, gluconeogenesis, and biofilm initiation (e.g., *frdABCD*, *glpX*, *csgDEFG*, *lsr* operon) [14,15], while *E. coli* Nissle 1917 maintained TCA cycle activity, upregulated glyoxylate and glycolate shunts (*aceAB*, *glcEFGBA*),

sialic acid degradation genes (*nanK*, *nanE*), and stress-linked membrane proteases [18–25]. Motility was repressed in both strains via distinct routes: upregulation of curli biosynthesis genes (*csgDEFG*) and quorum-sensing regulators (*lsrABCD*, *lsrRK*) in K-12 and downregulation of flagellar genes in Nissle 1917 [21,26].

The regulatory differences were further explored via electrophoretic mobility shift assays, using YihW proteins from both strains (which have the amino acid sequence identity of 72%) and intergenic regions from both cassette types. It was previously demonstrated that YihW (CsqR) from K-12 MG1655 binds to its main target, the *yihU/V* intergenic region, and its own promoter (*yihV/W*) with clear SQ-induced dissociation of complexes [10]. We confirmed this binding pattern, although at much lower protein excess (8×–16× compared to 600×). In contrast, YihW from Nissle 1917 showed significantly weaker binding to its targets at a physiological temperature of 37 °C (Figure 6B for *yihU/V*, and almost no binding for *yihV/W*, Supplementary Figure S7). Moderate binding was observed at 30 °C—all *yihU/V* DNA was bound at 32× molar excess of YihW (Figure 7), while for *yihV/W* a saturation effect was demonstrated with no difference in binding between 32× and 48× excess of the regulator with free DNA still visible. One possible analogy is the "defensive oversaturation" observed in the *E. coli* GalS system, where high repressor levels lead to autoregulatory binding that buffers against hyper-repression [27]. Notably, the saturation effect was not detected when YihW from *E. coli* Nissle1917 was exposed to the *E. coli* K-12 MG1655-derived DNA fragments (Figure 8). Addition of SQ led to evident dissociation of the YihW-DNA complexes (Figures 7 and 8). However, for both fragments from *E. coli* Nissle 1917, the dissociation effect was weaker than for fragments from *E. coli* K-12 MG1655. Together with the RNA-seq results, this may indicate a less important role of the *yih* cassette for SQ catabolism in this strain. Thus, we suggest that the transcriptional regulator YihW from *E. coli* Nissle 1917 should not be called CsqR (as it is referred to in K-12 MG1655), due to lack of evidence of its involvement in SQ metabolism.

Regulatory rewiring might also explain why the short cassette is retained in probiotic and pathogenic strains despite apparent loss of the SQ catabolism function. In the broader context of transcriptional regulation, it was found that deletion of *yihW* in MG1655 derepresses not only the *yih* cassette genes, but also *ydfG*, an oxidoreductase of unknown relation to the SQ metabolism [10]. This implies that YihW may act as a node in a broader sulfur-responsive regulatory network. This is in line with models of operon evolution where modular gene clusters undergo selective retention, degeneration, or recombination in response to environmental demands [28–31].

4. Materials and Methods

4.1. Distribution and Composition of yih Cassettes in E. coli Strains

The genomic and phylogenetic data for *Escherichia coli* and *Shigella* spp. were taken from Seferbekova et al. [13]. The dataset included genomic and coding DNA sequences of *E. coli* and *Shigella* spp., universal single-copy orthologous groups (UOGs). The phylogenetic tree was constructed using a concatenated alignment of 238 universal single-copy orthologous groups with MAFFT (the linsi mode), followed by the Maximum Likelihood inference using RAxML with the GTR+Gamma model and 100 bootstrap replicates.

For the homology search, sequences of the *yih* cassette genes from *E. coli* K-12 MG1655 (GenBank ID: CP032667.1 [32]) were used as queries. Homologs were identified with MMseqs2 v14-7e284 [33], using the mmseqs2 search command with the default parameters against coding DNA sequences of bacterial strains from the considered dataset. The identified homologs of *yih* cassette genes were considered co-localized if the distance between their respective midpoint genomic positions was less than 5000 base pairs.

The phylogenetic tree of *E. coli* strains with resulting *yih* cassettes was visualized using the iTOL web server [34]. To analyze positioning and orientation of the cassettes across different *E. coli* strains, the start coordinate of the first gene and the end coordinate of the last gene in each *yih* cassette was compared with the positions of adjacent UOGs on the bacterial chromosome. Most common compositions of *yih* cassettes flanked by UOGs were visualized using the DNA Features Viewer package version 3.1.3 [35]. To test for an association between the type of *yih* cassettes and pathogenicity, we used metadata on pathotype status from Seferbekova et al. [13]. The distribution of *yih* cassette types among pathogenic and non-pathogenic strains was analyzed using Fisher's exact test.

4.2. Strains and Growth Conditions

For transcriptomics studies, two strains of *E. coli* were used, *E. coli* K-12 MG1655 (Wild type F- lambda- ilvG- rfb-50 rph-1) [36] and *E. coli* Nissle 1917 (Mutaflor®, DSM 6601, serotype O6:K5:H1, Herdecke, Germany) [37]. Cell cultures were grown in the minimal medium M9 supplemented with 5% (*v*/*v*) LB and 0.2% (*w*/*v*) of a carbon source, either D-glucose or sulfoquinovose. Bacterial cultures were grown aerobically at 37 °C till mid-exponential phase (OD_{600} = 0.2–0.3) under constant shaking. OD_{600} of each sample was measured using UV–VIS Spectrophotometer (Thermo Scientific, Waltham, MA, USA). Two biological replicates per condition were used for the RNA-seq analysis, and three biological replicates per condition were used for the qRT-PCR validation.

4.3. RNA Extraction and qRT-PCR

RNA was extracted using TRIZol (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol and then treated with DNAse I (New England Biolabs, Ipswich, MA, USA) for 1 h at 37 °C. Reverse transcription was made using 1 µg of total RNA, gene-specific primers, and MMul-V RevertAid reverse transcriptase (Thermo Fisher Scientific, Vilnius, Lithuania) according to the manufacturer's protocol. Primers used for reverse transcription (−RT) and amplification (−PCR) are listed in Supplementary Table S1. As a control, *hns* was used [37]. The DT-lite thermocycler (DNA-Technology, Moscow, Russia) and qPCR-HS SYBR mix (Evrogen, Moscow, Russia) were used for quantitative PCR (qRT-PCR). Data obtained from three biological samples and no less than three technical replicates were analyzed for statistical significance using the $2^{-\Delta\Delta Ct}$ method. Error bars indicate standard deviations of the respective mean values.

4.4. RNA Library Preparation and Sequencing

RNA-seq was performed on two biological replicates per condition. One µg of total RNA from exponentially growing cells was taken for rRNA depletion using the Illumina RiboZero kit (Illumina, San Diego, CA, USA). 100 ng of rRNA-depleted RNA was further taken for library preparation. Sequencing libraries were prepared using the NEBNext Ultra II Directional RNA library prep kit for Illumina (New England Biolabs, Ipswich, MA, USA) ccording to the manufacturer's protocol and indexed with the NebNext primers (New England Biolabs, Ipswich, MA, USA). Libraries were quantified using the Qubit HS RNA kit (Thermo, Waltham, MA, USA) and qualified on Bioanalyzer (Agilent, Santa Clara, CA, USA). Sequencing was performed as 150 + 150 PE on the Illumina NextSeq 500 machine at the Skoltech Core Genomics and Bioimaging facility. All raw sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) and can be accessed under the BioProject accession number PRJNA1338229.

4.5. Analysis of RNA-Sequencing Data

Quality of reads was assessed using FastQC v0.11.9 and MultiQC v1.12.dev0 [38]. Reads were trimmed to remove adapters using BBduk v35.85 [39]. The genome assemblies

of *E. coli* K-12 MG1655 (GenBank accession: GCF_000005845.2) and *E. coli* Nissle 1917 (GenBank accession: GCF_019967895.1) were used as references. Reads were mapped to the reference data using Bowtie2 v2.2.1 [40]. Reads were decontaminated of rRNA. All processing of read alignment data was performed using Samtools v1.11 [41]. Read counts were obtained using featureCounts v2.0.1 [42].

Differential expression analysis was performed on two biological replicates for each growth condition (growth on either SQ or glucose) using DESeq2 v1.38.3, with growth on glucose as a baseline condition [43]. Genes with an adjusted *p*-value of <0.05, calculated using the Benjamini–Hochberg method, were selected for further analysis. For visualization purposes, adjusted *p*-values smaller than 1×10^{-100} were set to a minimum value of 1×10^{-100} to improve readability in the volcano plot and avoid compression of moderately significant data points. All raw values, including uncapped adjusted *p*-values, are available in Supplementary Table S4. To compare differential expression data between the two strains, homologs of *E. coli* Nissle 1917 genes were identified in the *E. coli* K-12 MG1655 genome by reciprocal best-hit search with MMseqs2 v13-45111 [43]. Functional analysis of differentially expressed genes (DEGs) was performed using the clusterProfiler v4.6.2 [44], pathview v1.38.0 [45], KEGGREST, v1.38.0 and DOSE v3.24.2 [46] R packages as follows: To determine which pathways respond most prominently to growth on SQ, DEGs from each strain were mapped to the KEGG genome data of *E. coli* K-12 MG1655 and gene set enrichment analysis was conducted on DEGs within specific KEGG pathways (*p*-value < 0.05, the Benjamini–Hochberg adjusted *p*-value method). Functional annotation of DEGs was cross-validated using the DAVID web-server [47]. Operons containing the DEGs in the reference genome of each strain were predicted using Operon-mapper [48].

4.6. YihW Expression and Purification

Both protein variants were expressed under control of T7 promoter and own Shine-Dalgarno and regulatory regions in the pGEMEX-based vectors [49,50]. YihW from *E. coli* K-12 MG1655 was produced as described in Rybina et al. [49] and purified using the HIStrap column (GE Healthcare, Chicago, IL, USA) according to the manufacturer's protocol. YihW from Nissle 1917 was produced in BL21(DE3)-Codon Plus-RIL cells, induced with 50 µM IPTG at OD600 = 0.2 and incubated for 5 h at 37 °C in LB (Luria–Bertani) media. Then cells from 200 mL of culture were harvested, washed with 1× BS (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4), resuspended in 1× PBS containing 1 mM DTT and 1 mM PMSF, and sonicated on the Q125 Sonicator (QSonica, LLC, Newtown, CT, USA) for 15 min (10 s pulse, 20 s pause) at 45% intensity. Cleared lysate was removed, the pellet was again washed, resuspended in 2 mL of Loading buffer (50 mM NaH2PO4, 300 mM NaCl, 5 mM imidazole, 8 M urea), and then the protein was purified on the Ni-NTA spin columns (Qiagen, Hilden, Germany) using modified buffers containing both urea and imidazole (5 mM for equilibration, 20 mM for washes, 500 mM for elution). Purified protein was then dialyzed against buffers containing 50 mM NaH2PO4 and 300 mM NaCl with lowering urea concentrations and used for band-shift assays (Supplementary Figure S6).

4.7. Electrophoretic Mobility Shift Assays

Electrophoretic mobility shift assays (EMSA, band-shift assays) were used to test the efficiency of the YihW proteins binding to the regulatory regions of genes from the *yih* cassette. DNA fragments containing intergenic regions *yihUV* and *yihVW* were PCR amplified (primers marked as R and F for each intergenic region, Supplementary Table S1), extracted from the gel and used for subsequent experiments. 0.5 pmol of DNA was incubated at 30 °C or 37 °C in 1× Binding buffer (50 mM Tris-Cl, 100 mM KCl, 50% glycerol, 1mM EDTA) and appropriate amount of SQ [10] synthesized as described in [31] for 10 min,

then 2 to 24 pmol of purified YihW was added. Free DNA fragments were loaded on separate lanes as controls. After further 30 min incubation, complexes were loaded onto a 5% polyacrylamide gel that had been pre-warmed to 30 °C or 37 °C, respectively. Gels were run in 1xTBE buffer under 200V, stained with ethidium bromide and visualized on the Vilber gel-documentation system (Collégien, France).

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms262110351/s1.

Author Contributions: Conceptualization, A.D.K. and M.N.T.; methodology, A.A.R., A.D.K., M.N.T.; software, A.A.R.; validation, T.A.B., D.S.K. and A.A.R.; formal analysis, A.A.R.; investigation, A.D.K., M.N.T. and A.A.R.; resources, M.N.T. and M.S.G.; data curation, A.D.K., A.A.R., T.A.B., D.S.K. and M.N.T.; writing—original draft preparation, A.D.K. and A.A.R.; writing—review and editing, A.D.K., M.N.T. and M.S.G.; visualization, A.A.R. and T.A.B.; supervision, M.N.T. and M.S.G.; project administration, M.S.G.; funding acquisition, M.S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Russian Science Foundation (RSCF, Moscow, Russia) under grant number 24-14-00276.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The RNA-seq data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA1338229. The code used for RNA-seq preprocessing, differential expression analysis, and phylogenetic annotation is available at: https://github.com/rybinaanya/for_yih_consideration (accessed on 21 October 2025).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Benson, A.A.; Daniel, H.; Wiser, R. A Sulfolipid in Plants. *Proc. Natl. Acad. Sci. USA* **1959**, *45*, 1582–1587. [CrossRef]
2. Harwood, J.L.; Nicholls, R.G. The Plant Sulpholipid—A Major Component of the Sulphur Cycle. *Biochem. Soc. Trans.* **1979**, *7*, 440–447. [CrossRef] [PubMed]
3. Wei, Y.; Tong, Y.; Zhang, Y. New Mechanisms for Bacterial Degradation of Sulfoquinovose. *Biosci. Rep.* **2022**, *42*, BSR20220314. [CrossRef]
4. Ye, Z.; Wei, Y.; Jiang, L.; Zhang, Y. Oxygenolytic Sulfoquinovose Degradation by an Iron-Dependent Alkanesulfonate Dioxygenase. *iScience* **2023**, *26*, 107803. [CrossRef]
5. Felux, A.-K.; Spiteller, D.; Klebensberger, J.; Schleheck, D. Entner-Doudoroff Pathway for Sulfoquinovose Degradation in Pseudomonas Putida SQ1. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E4298–E4305. [CrossRef]
6. Denger, K.; Weiss, M.; Felux, A.-K.; Schneider, A.; Mayer, C.; Spiteller, D.; Huhn, T.; Cook, A.M.; Schleheck, D. Sulphoglycolysis in *Escherichia coli* K-12 Closes a Gap in the Biogeochemical Sulphur Cycle. *Nature* **2014**, *507*, 114–117. [CrossRef] [PubMed]
7. Denger, K.; Huhn, T.; Hollemeyer, K.; Schleheck, D.; Cook, A.M. Sulfoquinovose Degraded by Pure Cultures of Bacteria with Release of C3-Organosulfonates: Complete Degradation in Two-Member Communities. *FEMS Microbiol. Lett.* **2012**, *328*, 39–45. [CrossRef]
8. Mui, J.W.-Y.; De Souza, D.P.; Saunders, E.C.; McConville, M.J.; Williams, S.J. Remodeling of Carbon Metabolism during Sulfoglycolysis in *Escherichia coli*. *Appl. Environ. Microbiol.* **2023**, *89*, e0201622. [CrossRef]
9. Kaznadzey, A.; Shelyakin, P.; Belousova, E.; Eremina, A.; Shvyreva, U.; Bykova, D.; Emelianenko, V.; Korosteleva, A.; Tutukina, M.; Gelfand, M.S. The Genes of the Sulphoquinovose Catabolism in *Escherichia coli* Are Also Associated with a Previously Unknown Pathway of Lactose Degradation. *Sci. Rep.* **2018**, *8*, 3177. [CrossRef] [PubMed]
10. Shimada, T.; Yamamoto, K.; Nakano, M.; Watanabe, H.; Schleheck, D.; Ishihama, A. Regulatory Role of CsqR (YihW) in Transcription of the Genes for Catabolism of the Anionic Sugar Sulfoquinovose (SQ) in *Escherichia coli* K-12. *Microbiology* **2019**, *165*, 78–89. [CrossRef]

11. Pérez-Rueda, E.; Collado-Vides, J. The Repertoire of DNA-Binding Transcriptional Regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **2000**, *28*, 1838–1847. [CrossRef] [PubMed]
12. Grozdanov, L.; Raasch, C.; Schulze, J.; Sonnenborn, U.; Gottschalk, G.; Hacker, J.; Dobrindt, U. Analysis of the Genome Structure of the Nonpathogenic Probiotic *Escherichia coli* Strain Nissle 1917. *J. Bacteriol.* **2004**, *186*, 5432–5441. [CrossRef]
13. Seferbekova, Z.; Zabelkin, A.; Yakovleva, Y.; Afasizhev, R.; Dranenko, N.O.; Alexeev, N.; Gelfand, M.S.; Bochkareva, O.O. High Rates of Genome Rearrangements and Pathogenicity of *Shigella* spp. *Front. Microbiol.* **2021**, *12*, 628622. [CrossRef]
14. Hua, Q.; Yang, C.; Oshima, T.; Mori, H.; Shimizu, K. Analysis of Gene Expression in *Escherichia coli* in Response to Changes of Growth-Limiting Nutrient in Chemostat Cultures. *Appl. Environ. Microbiol.* **2004**, *70*, 2354–2366. [CrossRef]
15. Martínez-Gómez, K.; Flores, N.; Castañeda, H.M.; Martínez-Batallar, G.; Hernández-Chávez, G.; Ramírez, O.T.; Gosset, G.; Encarnación, S.; Bolivar, F. New Insights into *Escherichia coli* Metabolism: Carbon Scavenging, Acetate Metabolism and Carbon Recycling Responses during Growth on Glycerol. *Microb. Cell Fact.* **2012**, *11*, 46. [CrossRef]
16. Lawrence, J.G.; Roth, J.R. Selfish Operons: Horizontal Transfer May Drive the Evolution of Gene Clusters. *Genetics* **1996**, *143*, 1843–1860. [CrossRef]
17. Kertesz, M.A. Riding the Sulfur Cycle—Metabolism of Sulfonates and Sulfate Esters in Gram-Negative Bacteria. *FEMS Microbiol. Rev.* **2000**, *24*, 135–175. [CrossRef] [PubMed]
18. Pavoncello, V.; Barras, F.; Bouveret, E. Degradation of Exogenous Fatty Acids in *Escherichia coli*. *Biomolecules* **2022**, *12*, 1019. [CrossRef]
19. Brauer, M.J.; Yuan, J.; Bennett, B.D.; Lu, W.; Kimball, E.; Botstein, D.; Rabinowitz, J.D. Conservation of the Metabolomic Response to Starvation across Two Divergent Microbes. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19302–19307. [CrossRef]
20. Bouillet, S.; Bauer, T.S.; Gottesman, S. RpoS and the Bacterial General Stress Response. *Microbiol. Mol. Biol. Rev.* **2024**, *88*, e00151-22. [CrossRef]
21. Schink, S.J.; Biselli, E.; Ammar, C.; Gerland, U. Death Rate of *E. coli* during Starvation Is Set by Maintenance Cost and Biomass Recycling. *Cell Syst.* **2019**, *9*, 64–73.e3. [CrossRef]
22. Cornforth, D.M.; Foster, K.R. Competition Sensing: The Social Side of Bacterial Stress Responses. *Nat. Rev. Microbiol.* **2013**, *11*, 285–293. [CrossRef]
23. Kuhar, I.; Zgur-Bertok, D. Transcription Regulation of the Colicin K Cka Gene Reveals Induction of Colicin Synthesis by Differential Responses to Environmental Signals. *J. Bacteriol.* **1999**, *181*, 7373–7380. [CrossRef]
24. Matin, A. The Molecular Basis of Carbon-starvation-induced General Resistance in *Escherichia coli*. *Mol. Microbiol.* **1991**, *5*, 3–10. [CrossRef]
25. Jozefczuk, S.; Klie, S.; Catchpole, G.; Szymanski, J.; Cuadros-Inostroza, A.; Steinhauser, D.; Selbig, J.; Willmitzer, L. Metabolomic and Transcriptomic Stress Response of *Escherichia coli*. *Mol. Syst. Biol.* **2010**, *6*, 364. [CrossRef]
26. Li, J.; Attila, C.; Wang, L.; Wood, T.K.; Valdes, J.J.; Bentley, W.E. Quorum Sensing in *Escherichia coli* Is Signaled by AI-2/LsrR: Effects on Small RNA and Biofilm Architecture. *J. Bacteriol.* **2007**, *189*, 6011–6020. [CrossRef]
27. Semsey, S.; Krishna, S.; Erdőssy, J.; Horváth, P.; Orosz, L.; Sneppen, K.; Adhya, S. Dominant Negative Autoregulation Limits Steady-State Repression Levels in Gene Networks. *J. Bacteriol.* **2009**, *191*, 4487–4491. [CrossRef]
28. Price, M.N.; Dehal, P.S.; Arkin, A.P. Horizontal Gene Transfer and the Evolution of Transcriptional Regulation in *Escherichia coli*. *Genome Biol.* **2008**, *9*, R4. [CrossRef] [PubMed]
29. Rodionov, D.A.; Dubchak, I.; Arkin, A.; Alm, E.; Gelfand, M.S. Reconstruction of Regulatory and Metabolic Pathways in Metal-Reducing Delta-Proteobacteria. *Genome Biol.* **2004**, *5*, R90. [CrossRef] [PubMed]
30. Rodionov, D.A.; Vitreschak, A.G.; Mironov, A.A.; Gelfand, M.S. Comparative Genomics of the Methionine Metabolism in Gram-Positive Bacteria: A Variety of Regulatory Systems. *Nucleic Acids Res.* **2004**, *32*, 3340–3353. [CrossRef] [PubMed]
31. Leyn, S.A.; Suvorova, I.A.; Kholina, T.D.; Sherstneva, S.S.; Novichkov, P.S.; Gelfand, M.S.; Rodionov, D.A. Comparative Genomics of Transcriptional Regulation of Methionine Metabolism in Proteobacteria. *PLoS ONE* **2014**, *9*, e113714. [CrossRef] [PubMed]
32. Sayers, E.W.; Beck, J.; Bolton, E.E.; Brister, J.R.; Chan, J.; Connor, R.; Feldgarden, M.; Fine, A.M.; Funk, K.; Hoffman, J.; et al. Database Resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res* **2025**, *53*, D20–D29. [CrossRef] [PubMed]
33. Steinegger, M.; Söding, J. MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [CrossRef]
34. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef] [PubMed]
35. Zulkower, V.; Rosser, S. DNA Features Viewer: A Sequence Annotation Formatting and Plotting Library for Python. *Bioinformatics* **2020**, *36*, 4350–4352. [CrossRef]
36. Blattner, F.R.; Plunkett, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; et al. The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1462. [CrossRef]

37. Tutukina, M.N.; Potapova, A.V.; Cole, J.A.; Ozoline, O.N. Control of Hexuronate Metabolism in *Escherichia coli* by the Two Interdependent Regulators, ExuR and UxuR: Derepression by Heterodimer Formation. *Microbiology* **2016**, *162*, 1220–1231. [CrossRef]
38. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef]
39. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*; Lawrence Berkeley National Lab. (LBNL): Berkeley, CA, USA, 2014.
40. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
41. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup the Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
42. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]
43. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef] [PubMed]
44. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters. *OMICS* **2012**, *16*, 284–287. [CrossRef]
45. Luo, W.; Brouwer, C. Pathview: An R/Bioconductor Package for Pathway-Based Data Integration and Visualization. *Bioinformatics* **2013**, *29*, 1830–1831. [CrossRef] [PubMed]
46. Yu, G.; Wang, L.-G.; Yan, G.-R.; He, Q.-Y. DOSE: An R/Bioconductor Package for Disease Ontology Semantic and Enrichment Analysis. *Bioinformatics* **2015**, *31*, 608–609. [CrossRef]
47. Sherman, B.T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M.W.; Lane, H.C.; Imamichi, T.; Chang, W. DAVID: A Web Server for Functional Enrichment Analysis and Functional Annotation of Gene Lists (2021Update). *Nucleic Acids Res.* **2022**, *50*, W216–W221. [CrossRef]
48. Taboada, B.; Estrada, K.; Ciria, R.; Merino, E. Operon-Mapper: A Web Server for Precise Operon Identification in Bacterial and Archaeal Genomes. *Bioinformatics* **2018**, *34*, 4118–4120. [CrossRef] [PubMed]
49. Rybina, A.A.; Glushak, R.A.; Bessonova, T.A.; Dakhnovets, A.I.; Rudenko, A.Y.; Ozhiganov, R.M.; Kaznadzey, A.D.; Tutukina, M.N.; Gelfand, M.S. Phylogeny and Structural Modeling of the Transcription Factor CsqR (YihW) from *Escherichia coli*. *Sci. Rep.* **2024**, *14*, 7852. [CrossRef]
50. Igarashi, K.; Ishihama, A. Bipartite Functional Map of the *E. coli* RNA Polymerase Alpha Subunit: Involvement of the C-Terminal Region in Transcription Activation by cAMP-CRP. *Cell* **1991**, *65*, 1015–1022. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.