nature communications

Article

https://doi.org/10.1038/s41467-024-50957-y

Structure, biosynthesis and regulation of the T1 antigen, a phase-variable surface polysaccharide conserved in many *Salmonella* serovars

Received: 5 March 2024

Accepted: 22 July 2024

Published online: 02 August 2024

Check for updates

Steven D. Kelly[1], Mikel Jason Allas[2,3], Lawrence D. Goodridge[4], Todd L. Lowary[2,3,5] & Chris Whitfield[1]

The bacterial genus *Salmonella* includes diverse isolates with multiple variations in the structure of the main polysaccharide component (O antigen) of membrane lipopolysaccharides. In addition, some isolates produce a transient (T) antigen, such as the T1 polysaccharide identified in the 1960s in an isolate of *Salmonella enterica* Paratyphi B. The structure and biosynthesis of the T1 antigen have remained enigmatic. Here, we use biophysical, biochemical and genetic methods to show that the T1 antigen is a complex linear glycan containing tandem homopolymeric domains of galactofuranose and ribofuranose, linked to lipid A–core, like a typical O antigen. T1 is a phase-variable antigen, regulated by recombinational inversion of the promoter upstream of the T1 genetic locus through a mechanism not observed for other bacterial O antigens. The T1 locus is conserved across many *Salmonella* isolates, but is mutated or absent in most typhoidal serovars and in serovar Enteritidis.

Bacteria belonging to the genus *Salmonella* are ubiquitous. They cause a range of human infections including gastroenteritis and bacteremia (usually by foodborne, non-typhoidal strains), and enteric fever (by typhoidal strains)[1]. The *Salmonella* lineage contains two species: *enterica* and *bongori*. The latter almost exclusively infect cold-blooded animals, although occasional human outbreaks are reported[2], while *S. enterica* contains the predominant human pathogens and is made up of six subspecies (*enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI). The subspecies are further categorized into serotypes via the Kauffmann–White scheme, based on different combinations of lipopolysaccharide (LPS) O-antigens and flagellar filament H-antigens. These define more that 2600 serovars often named after the geographic location where they were isolated[3].

LPS O-antigens (O-polysaccharides; OPSs) are the major surface polysaccharide structures present in *Salmonella* (and in many other Gram-negative genera). They are composed of repeat units whose variable structures provide distinct antigenic epitopes defining the serotype specific portion of LPS[4]. Currently, 46 structurally characterized OPS structures have been reported for *S. enterica* serogroups and the genetic regions driving their biosynthesis have been identified[5]. However, a recent pangenome study suggests further diversity may exist[5,6]. The majority of *Salmonella* O-antigens are synthesized by the Wzx/Wzy-dependent assembly strategy, where OPS repeat units are built on an undecaprenyl phosphate (und-P) acceptor at the interface of the cytoplasm and the cell membrane (reviewed in[4]). These are subsequently exported by a Wzx flippase and polymerized by the Wzy polymerase at the periplasmic face of the membrane, before being ligated to the conserved portion of the LPS glycolipid (consisting of lipid A and a core oligosaccharide containing the OPS attachment site). Serotype O67 is the only example that uses a different assembly pathway, where the und-linked glycan is polymerized in the cytoplasm and exported to the periplasm via an ATP-binding

[1]Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada. [2]Department of Chemistry, University of Alberta, Edmonton, AB, Canada. [3]Institute of Biological Chemistry, Academia Sinica, Nangang, Taipei, Taiwan. [4]Department of Food Science, University of Guelph, Guelph, ON, Canada. [5]Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan. e-mail: tlowary@gate.sinica.edu.tw; cwhitfie@uoguelph.ca

cassette (ABC) transporter for ligation[5]. The completed LPS is then translocated to the cell surface.

While OPSs are distinguishing polysaccharide antigens in *Salmonella* serotypes, they can be co-expressed with other glycans. All isolates produce Enterobacterial Common Antigen, which they share with other members of the Enterobacteriaceae; the precise cellular role(s) for this polysaccharide is uncertain[7]. *Salmonella* also make a colanic acid exopolysaccharide, regulated by the Rcs regulon[8]. This is absent in human-adapted *Salmonella* Typhi and Paratyphi C, which produce the host immune response-shielding Vi antigen capsular polysaccharide[9]. Finally, some isolates also produce T antigens (where T denotes their apparent transient nature) but the structure(s) and function(s) of these antigens is enigmatic. The T1 antigen was the first representative identified, found in a strain of Paratyphi B that was not typable by the available O-antisera[10]. A subsequent study employed methylation analysis to describe an incomplete T1 structure containing a complex mix of galactofuranose (Gal*f*) and ribofuranose (Rib*f*) residues and it was hypothesized that the T1 glycan might be linked to the LPS lipid A–core[11]. Further studies described preliminary but inconclusive attempts to identify the precursors for T1-antigen biosynthesis, as well as the consequences of T1 expression on virulence and phage-sensitivity[12–15].

Surprisingly, investigation of the T1 antigen ended in the early 1970s without resolving its complete structure or understanding its relationship to serotype-specific OPS. Furthermore, the genetic locus responsible for its production and the molecular basis for its transient nature was never identified. The goal of this study was to address these open questions to provide a platform for investigating the biological functions of this unusual glycan.

Results

Identification of the Salmonella T1 biosynthetic cluster encoding an ABC transporter-dependent system

We recently reported the first structure and function of a family of ribofuranosyltransferase (Rib*f*-GT) enzymes involved in polysaccharide bacterial biosynthesis[16]. The prototypical dual catalytic domain Rib*f*-GTs contain a C-terminal glycan phosphoribosyltransferase (gPRT), which transfers ribose-5-phosphate from the activated donor, phosphoribosyl-5-phospho-D-ribosyl-α−1-diphosphate (PRPP), to a sugar acceptor. In a second reaction step, the 5′ phosphate is then removed by an N-terminal phosphoribose phosphatase (PRP) catalytic module. Because the T1 antigen was previously reported to contain Rib*f*[11], the biochemically defined Rib*f*-transferase prototype from *Klebsiella pneumoniae*[16] was exploited in a BLAST search to identify a homolog in *Salmonella*. A candidate Rib*f*-transferase gene was found in serovar Paratyphi B, the original T1 source[10] but, surprisingly, additional candidates were identified in genomes from a wide range of *Salmonella* serovars. The candidate Rib*f*-transferases share 97–100% identity to one another (with 100% coverage) and are encoded by a gene in a well conserved locus located between *nei* (encoding endonuclease VIII) and *pxpA* (5-oxoprolinase subunit) (Fig. 1a). The locus also contained genes associated with production of

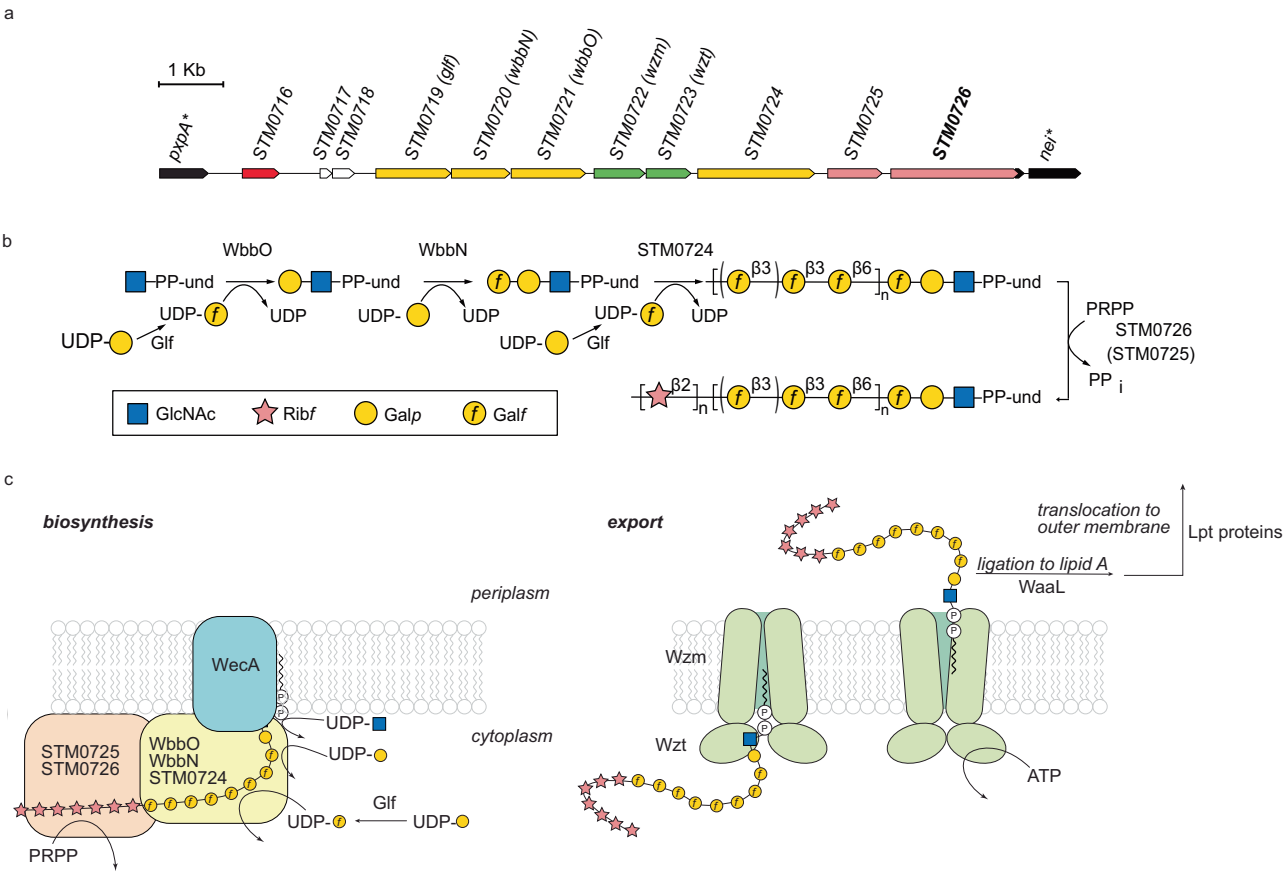

Fig. 1 | Genetic region and biosynthesis of the *Salmonella* T1 antigen. a The T1 antigen biosynthesis cluster in *Salmonella enterica* serovar Typhimurium SL3770 (LT2) (NC_003197.2). The cluster is a highly conserved ~12,500 base pair region between *pxpA* and *nei* in *Salmonella*. *STM0726* (in bold) encodes the Rib*f*-transferase used to identify the cluster. Genes responsible for galactofuranose addition are colored yellow, and genes responsible for ribofuranose addition are shown in pink. The ABC transporter components are shown in green and the tyrosine recombinase (characterized in Fig. 5) is shown in red. **b** The biosynthetic pathway for T1 antigen biosynthesis established in this study. **c** Cartoon representation of the biosynthesis and export steps of the T1 antigen. Note that biosynthesis and export are separated for simplicity, but in the cell these processes are expected to be obligatorily coupled.

https://doi.org/10.1038/s41467-024-50957-y

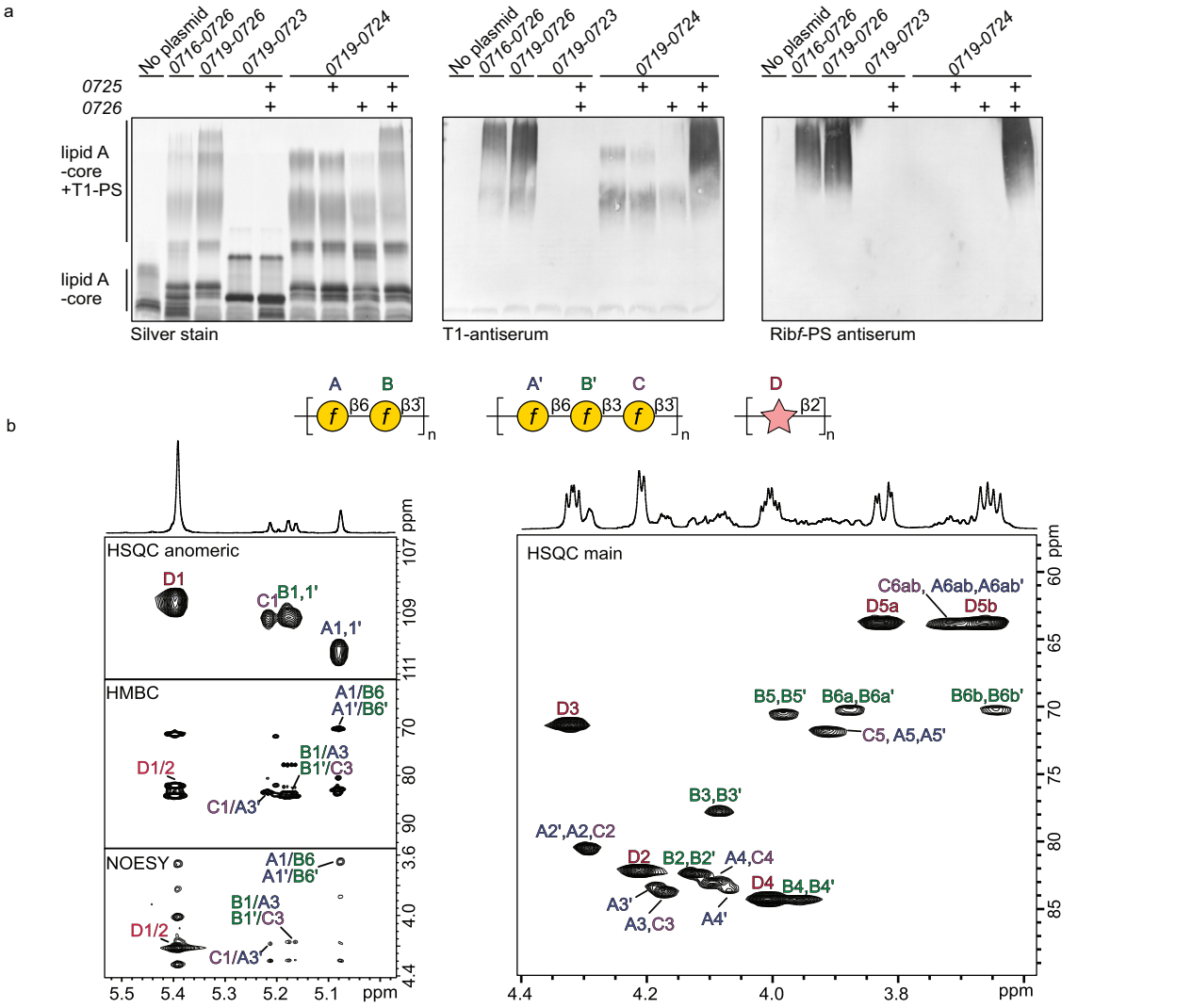

Fig. 2 | Production of the T1 antigen in *E. coli* K-12. a Silver stained SDS-PAGE gel and western immunoblots demonstrating recombinant expression of the T1 antigen in *E. coli* DH5α, which lack their native OPS due to mutations in the ancestral O16 gene cluster[17]. Samples were prepared as whole-cell lysates and similar results were acquired in triplicate. Rabbit T1 antiserum was raised against *E. coli* DH5α transformed with pWQ1128 (*STM0719-STM0726*). Rib*f*-PS antiserum was produced by adsorbing the T1 antiserum against *E. coli* DH5α expressing pWQ1130 (*STM0719-STM0724*) to remove antibodies recognizing Gal*f*-PS and therefore also shows no reactivity to the core-oligosaccharide. Source data are provided as a Source Data file. **b** NMR spectroscopic analysis of the T1 antigen. The spectra reveal a mixture of two Gal*f*-PS species and one Rib*f*-PS species, with the relevant structures shown above the spectra. Repeat unit assignments are shown on the HSQC spectrum and sugar linkage points are indicated on the HMBC and NOESY spectra. Chemical shifts are reported in Supplementary Table 1.

Gal*f*-containing polysaccharides in other bacteria (see below), lending confidence that it was responsible for T1-antigen production. Unfortunately, we were unable to obtain the original T1-producing Paratyphi B isolate, so the 11-*orf* locus from the well-studied *S. enterica* serovar Typhimurium LT2 (strain SL3770) was used as a prototype for subsequent investigation. At the start of this study the molecular basis for the transient T1 expression in *Salmonella* was unknown, thus precluding the use of the natural host for the primary investigations. We therefore chose to perform the analysis in a recombinant *E. coli* K-12 background with the 11-*orf* locus introduced on a plasmid. The host glycosylation machineries in *E. coli* K-12 are clearly understood, allowing confident data interpretation.

To establish the function of the *Salmonella* locus, the 11-*orf* region was cloned into the pACYC184 vector, with the genes downstream of the *tetR* promoter, permitting expression from the *tetR* promoter, or from any native promoters within the cloned region. The recombinant plasmid pWQ1128 was transformed into *E. coli* DH5α, which (like other K-12 strains) lacks a native OPS[17]. Examination of LPS profiles of whole cell lysates in silver-stained SDS-PAGE (Fig. 2a) revealed that the expression of the 11-*orf* gene region diminished the amount of lipid A-core species and led to the production of high-molecular-weight molecules with a heterogenous size distribution, typical for the addition of glycan chains (and seen typically with OPS). As described below, eight of the eleven *orfs* in this region (beginning with *STM0719*) encode putative proteins sharing similarity to characterized participants in glycan biosynthesis and export in other bacteria (Fig. 1b, c). A DNA fragment containing *STM0719–STM0726* (cloned in pWQ1129) produced a similar profile to pWQ1128 (Fig. 2a), so pWQ1129 was used as the basis for other constructs. Addition of OPS to lipid A-core depends on WaaL, the O-antigen ligase enzyme, which attaches OPS to the core oligosaccharide in the periplasm (i.e., post transport)[4]. The high-molecular-weight product was lacking when the locus was expressed in a *waaL* mutant (Supplementary Fig. 1a), confirming the T1 antigen is presented like conventional OPS and does not possess a different lipid terminus like that used in the *Salmonella* Vi antigen glycolipid[18].

STM0722 and *STM0723* are homologs of *wzm* and *wzt* (Fig. 1a), encoding the transmembrane and nucleotide-binding polypeptides for an ABC transporter, defining one of the major strategies for surface

polysaccharide assembly in bacteria[4] (Fig. 1c). These ABC transporters export undecaprenyl diphosphate-linked glycans (including some OPSs) to the periplasm. In most examples, the synthesis of the lipid intermediate is initiated by a homolog of WecA, a UDP-N-acetylglucosamine (GlcNAc)-dependent phosphoglycosyltransferase that transfers GlcNAc-1-P to undecaprenyl-phosphate[4] (Fig. 1c). WecA is encoded by a gene in the enterobacterial common antigen cluster located elsewhere on the *Salmonella* genome[7]. Consistent with this, no high-molecular-weight products are made in an *E. coli wecA* mutant transformed with pWQ1129 (Supplementary Fig. 1a). Six genes in the *Salmonella* locus (*STM0719–STM0721* and *STM0724–STM0726*) encode enzymes required for the biosynthesis of the T1 antigen (Fig. 1b) and their functions are described below.

The T1 antigen contains galactofuranose and ribofuranose homopolymeric domains

To determine the structure of the T1 antigen, LPS containing T1 was produced in *E. coli* K-12 harboring plasmid pWQ1129 and purified. The T1 antigen was released from the LPS by mild acid hydrolysis and analyzed by a combination of 1- and 2- dimensional ^{1}H and ^{13}C NMR spectroscopic experiments (Fig. 2b, Supplementary Table 1). These data showed that T1 was composed of two homopolymeric components; one containing Gal*f* and the other Rib*f*, with integration of the anomeric peaks indicating Rib*f* is present in ~2:1 abundance of Gal*f*. The Gal*f*-PS was a heterogenous mixture of two species. The first contained a disaccharide repeat unit [→3)-β-D-Gal*f*-(1 → 6)-β-D-Gal*f*-(1 →] (residues **A** and **B**), while the second had an additional (1 → 3)-linked Gal*f* residue, resulting in a trisaccharide repeat unit [→3)-β-D-Gal*f*-(1 → 3)-β-D-Gal*f*-(1 → 6)-β-D-Gal*f*-(1 →] (residues **A'** and **B'** and **C**). The Rib*f*-PS polymer contained a single anomeric signal (residue **D**) and was determined to be a β-(1 → 2)-linked homopolymer. Briefly, data from ^{1}H–^{1}H-COSY and ^{1}H–^{1}H–TOCSY experiments were used to assign the individual residues. To assign the linkage between the residues, ^{1}H–^{13}C HMBC (Fig. 2 and ^{1}H–^{1}H NOESY data were employed (Fig. 2). The H1 of residue **A** showed a correlation at δH 5.08/δC 70.2 in the HMBC and δH 5.08/δH 4.21 in the NOESY, corresponding to a linkage at C6 of residue **B** and **B'**. In turn, H1 of residues **B** and **B'** showed correlations of δH 5.17/δC 83.6 and δH 5.18/δC 83.6 in the HMBC and δH 5.17/δH 4.17 in the NOESY, demonstrating the linkages to C3 of residue **A** and C3 of residue **C**, respectively. Finally, residue **C** showed a correlation at δH 5.21/δC 83.3 in the HMBC and δH 5.21/δH 4.17 in the NOESY, corresponding to a linkage at C3 of residue **A'**. Residue **D** showed no major correlations to any of the Gal*f*-PS residues but instead had a strong correlation between H1/C2 at δH 5.40/δC 82.1 in HMBC and H1/H2 δH 5.40/4.21 in NOESY. Furthermore, the downfield shift of H2/C2 compared to literature data for unsubstituted Rib*f* glycosides[19] confirmed that the Rib*f*-PS was a (1 → 2)-linked homopolymer. Comparison of all anomeric signals with literature data were consistent with 1,2-*trans* (β) configurations for all glycosidic linkages (i.e., all appeared as singlets)[20]. However, the NMR data could not distinguish between two organizational scenarios. In one, the two structural domains are arranged in tandem. Alternatively, the T1 antigen may contain two populations, one with a linked Rib*f*-PS and the other with Gal*f*-PS. This was resolved by a combination of genetic and biochemical analyses.

The Galf-PS is linked to lipid A-core via an adapter trisaccharide

The *STM0719* gene is predicted to be encode a Glf (UDP-galactopyranose mutase) enzyme. The protein shares 75% identity, 85% similarity (96% coverage) with the well characterized *K. pneumoniae* Glf that produces UDP-Gal*f* from UDP-Gal*p*[21]. STM0720 (WbbN; 59% identity, 75% similarity, 100% coverage) and STM0721 (WbbO; 70% identity, 85% similarity, 99% coverage) are predicted UDP-Gal*f*- and UDP-Gal*p*-dependent GTs involved in biosynthesis of the *K. pneumoniae* O2a antigen. In O2a biosynthesis, these three proteins cooperate with WecA to generate an undecaprenyl diphosphate-linked trisaccharide (Gal*f*-Gal*p*-GlcNAc-PP-und)[22]. The *Salmonella* proteins possess the same biochemical activities (shown in the pathway scheme in Fig. 1b,c), based on the ability of the *Salmonella* homologs to restore O2a-antigen biosynthesis in mutants with corresponding deletions in the *Klebsiella* operon (Supplementary Fig. 1b). Furthermore, transformation of *E. coli* DH5α with a plasmid containing the *STM0719–STM0723* fragment (pWQ1130) results in a new product whose migration in SDS-PAGE is consistent with the addition of the predicted trisaccharide to lipid A-core (Fig. 2a).

In *Klebsiella* O2a biosynthesis, a dual catalytic module polymerase (WbbM) assembles a glycan with a disaccharide repeat unit [→3)-β-D-Gal*f*-(1 → 3)-α-D-Gal*p*-(1 →] using Gal*f*-Gal*p*-GlcNAc-PP-und produced by WecA-WbbO-WbbN as an acceptor. The T1-antigen pathway apparently uses the same acceptor for a different structure, and STM0724 was predicted to provide the putative polymerase for the repeat-unit domain based on its similarity to the GlfT2 GT from *Mycobacterium tuberculosis*. GlfT2 polymerizes Gal*f* residues with alternating β-(1 → 5) and β-(1 → 6) linkages, to build the backbone of the cell wall arabinogalactan, which is essential for viability of mycobacteria[23]. The predicted STM0724 protein and GlfT2 share 27% sequence identity and 42% similarity (Supplementary Fig. 2). Furthermore, an AlphaFold model of STM0724 shares structural similarity with the experimentally derived structure of GlfT2 (Fig. 3a)[23]. Transformation of *E. coli* DH5α with a plasmid containing *STM0719–STM0724* resulted in an SDS-PAGE LPS profile that retained a cluster of intermediate molecular weight species (larger than the presumed trisaccharide formed in the absence of *STM0724*). Compared to transformants containing the full locus, this material showed reduced reactivity with antiserum raised against cells expressing the T1 antigen (Fig. 2a). This would be anticipated if the LPS contained only a partial T1 structure, lacking some antigenic epitopes. These observations led to the hypothesis that STM0724 is the polymerase for the biosynthesis of the Gal*f*-PS component of T1 antigen.

To provide conclusive evidence for this activity, STM0724 was purified, and its activity was investigated in vitro using synthetic acceptor **1**, which replicates part of the repeat unit structure of the T1 Gal*f*-PS, linked to a methoxybenzamide (MB) group for detection. The donor substrate, UDP-Gal*f*, was synthesized in situ by Glf. The reaction generated a range of polymeric products with differing HPLC elution times, accompanied by consumption of the input acceptor (Fig. 3b). The reaction mix was scaled up and the product was purified by size exclusion chromatography and analyzed by NMR spectroscopy. The resulting spectra revealed that the in vitro product matched the Gal*f*-PS portion of the polysaccharide produced in vivo (Fig. 3c, Supplementary Table 2).

Linkage of Ribf-PS to LPS is dependent on production of Galf-PS

The collective evidence indicates that the genes in the *STM0719–STM0724* fragment in the cluster are sufficient to produce Gal*f*-PS linked to the non-reducing terminus of the trisaccharide generated by WecA–WbbN–WbbO (Fig. 1b,c). *E. coli* DH5α cells producing LPS containing this part of the T1 antigen were used to adsorb the Gal*f*-specific antibodies from T1 antiserum, generating antiserum specific for the Rib*f*-PS component. The Rib*f*-specific antiserum recognized bacteria transformed with the *STM0719–STM0726* construct and reacted with the highest molecular weight LPS species (Fig. 2a). Deletion of either *STM0725* or *STM0726* from the T1 locus eliminated production of the putative Rib*f*-PS, as might be expected because STM0726 is the only identifiable Rib*f*-transferase (i.e., possessing sequence similarity to the known Rib*f*-transferase prototype) encoded by the locus. The wildtype profile and immunoreactivity were restored by co-transformation with a plasmid carrying *STM0725–STM0726*. However, the profile was not restored by these genes when the Gal*f*-PS polymerase (STM0724) was absent, confirming that STM0724-directed synthesis of the Gal*f*-PS was necessary for further extension

https://doi.org/10.1038/s41467-024-50957-y

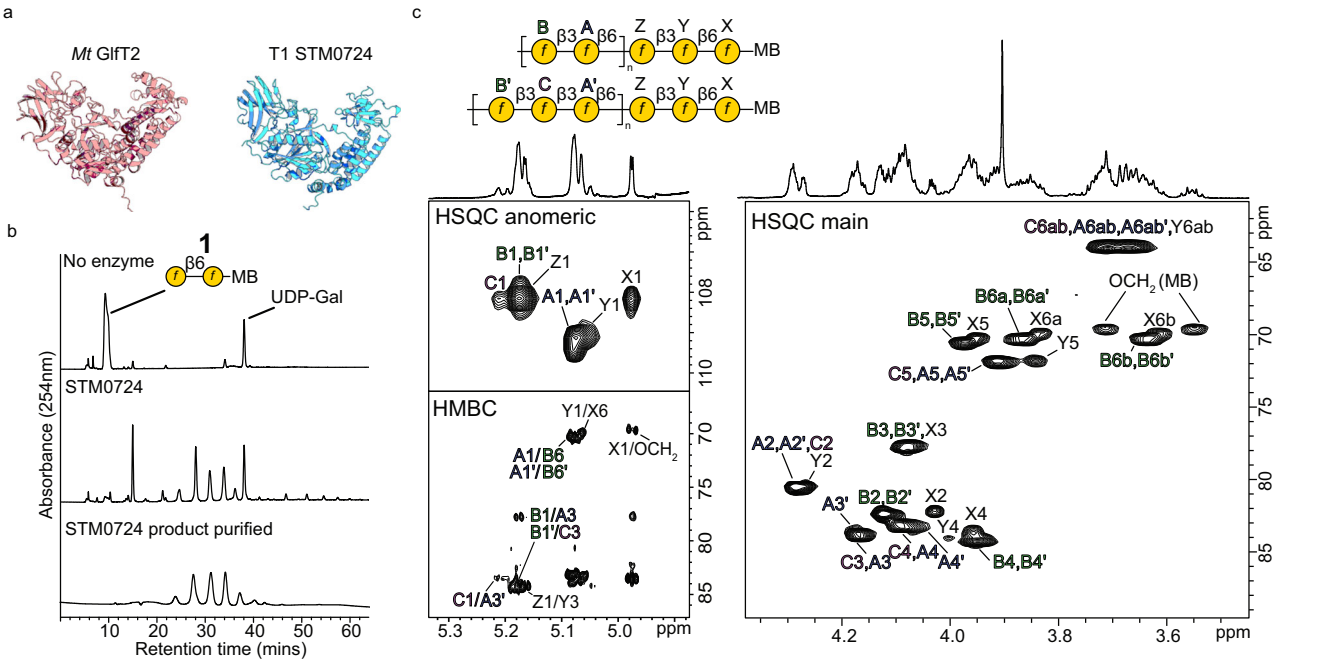

Fig. 3 | STM0724 is the T1 Gal*f*-PS polymerase and an ortholog of mycobacterial GlfT2. a Crystal structure of *Mycobacterium tuberculosis* GlfT2 (pink)(PDB ID: 4fix) and AlphaFold model of STM0724 (blue)(Uniprot Q8ZQV0). The proteins share 21% sequence identity and have a Z-score of 32.5, with an rmsd of 3.3 Å. **b** HPLC traces of in vitro GT reactions containing STM0724, UDP-Gal*f* (synthesized in situ by Glf from UDP-Gal*p*) and acceptor compound **1**. Reactions were performed for 1 h at 30 °C, stopped by the addition of an equal volume of acetonitrile, and the mixture analyzed by HPLC, using detection of the methoxybenzamide tag in the acceptor. The results from triplicate reactions were identical. Also shown is the product purified from large-scale reactions by gel filtration and SepPak elution. **c** NMR analysis of purified STM0724 product. The identity of the polymer is assigned on the HSQC spectrum and the linkage points of the sugars are shown on the HMBC spectrum. Sugars labeled X,Y and Z represent the reducing end and most of the chemical shifts converged onto the repeat unit. MB stands for methoxybenzamide. Chemical shifts are reported in Supplementary Table 2.

of the T1 antigen with the immunodominant non-reducing terminal Rib*f*-PS portion (completing the pathway scheme in Fig. 1b,c). These results eliminated the possibility of a separate population of LPS molecules substituted only with Rib*f*-PS. Further biochemical experiments described below identified the position within the Gal*f*-PS that provided the linkage site for Rib*f*-PS.

STM0725 and STM0726 are both required for Rib*f*-PS polymerization

Mutant complementation experiments with the construct lacking *STM0725–STM0726* showed that STM0725 and STM0726 are both required for restoring production of the Rib*f*-PS (Fig. 2a). A multiple sequence alignment revealed strong sequence conservation (including catalytic residues) between STM0726 and the characterized Rib*f*-transferase prototype (Supplementary Fig. 3). Consistent with the sequence similarity, an AlphaFold model of STM0726 contains the requisite N-terminal PRP module and a C-terminal gPRT module (Fig. 4a). Sequence data predicted that STM0725 is a GT and the corresponding AlphaFold model and multiple sequence alignment shows a single Rossmann-like fold and a DXD motif involved in the binding of a metal co-factor for coordination of nucleotide-diphosphate linked sugars, characteristic of the widespread GT-A fold (Supplementary Figs. 4, 5)[24].

To assign the precise activities of STM0725 and STM0726, six synthetic acceptors (**1**–**6**) (Supplementary Fig. 6), representing fragments of Gal*f*-PS with different lengths and terminal linkages, were employed. The acceptors were incubated with purified enzymes and different donors. Because Gal*f* and Rib*f* were the only sugars found in the T1 antigen carbohydrate structure (as described above)[11], UDP-Gal*f* and PRPP were reasoned to be the only donor candidates. The initial working hypothesis was that STM0725 would add a Gal*f* residue (as it lacks the characteristic architecture of Rib*f*-transferase) thus serving as a transition GT bridging the Gal*f*-PS and the Rib*f*-PS. However, this proved to be incorrect; in reactions containing STM0725 + STM0726 and UDP-Gal*f* only, none of the 6 acceptors were modified (Supplementary Fig. 7b). In contrast, in the presence of PRPP as the sole donor, compounds with a non-reducing terminal (1→3)-linked Gal*f* (**2, 4** and **5**) all served as acceptors for robust Rib*f*-transferase activity in reactions containing STM0725 and STM0726, giving a range of products with increasing retention times consistent with the transfer of multiple Rib*f* residues (Fig. 4b, Supplementary Fig. 7a,8). To validate the structure of the in vitro reaction products, a large-scale reaction was performed with **2** and the product was purified. NMR analysis (Fig. 4b) confirmed the presence of a β-(1→2) linked ribofuranose homopolymer linked to **2**, with no additional modifications, identical to the Rib*f*-PS domain in the authentic T1 antigen (Fig. 4c, Supplementary Table 3). Furthermore, an HMBC experiment identified the linkage between the reducing-end Rib*f* (C) and the 5-position of the non-reducing Gal*f* (B), by a correlation at δH 5.44/δC 77.8, and a characteristic downfield shift of C5 of B consistent with its substitution. Notably, re-analysis of the NMR spectra of T1 PS produced in recombinant *E. coli* showed the same linkage between the PS domains (Supplementary Fig. 9), but this assignment was only possible with the guidance provided by the data for the in vitro product due to the complexity of the spectra from the natural product. This reinforces the presence of a covalent linkage between the two structural domains and identifies the C5 hydroxyl group of (1→3)-linked Gal*f* residues as the attachment site for the Rib*f*-PS.

The data above also indicate that the combination of STM0725 and STM0726 transfer only Rib*f*, a result that is hard to reconcile with the GT-A-fold of STM0725. Therefore, in vitro reactions were performed in an attempt to dissect the individual roles of STM0725 and STM0726. STM0725 was unable to catalyze any Rib*f* additions to **2**, while STM0726 was capable of adding a small amount of a single Rib*f*

https://doi.org/10.1038/s41467-024-50957-y

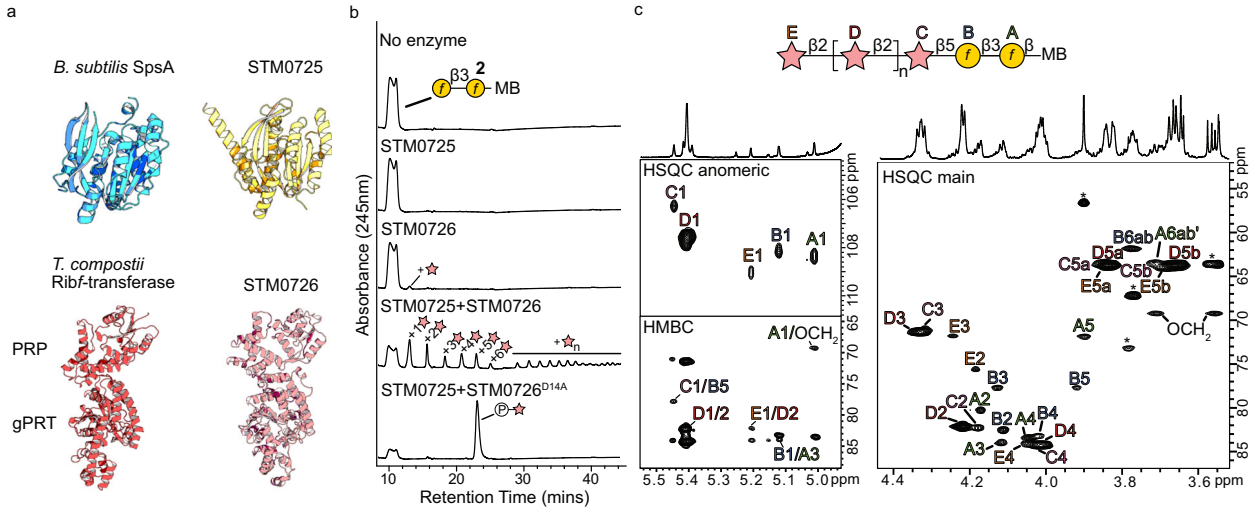

Fig. 4 | STM0725 and STM0726 are required for polymerization of T1 Rib*f*-PS. **a** AlphaFold models of STM0725 (Uniprot: Q8ZQU9) and STM0726 (Uniprot: Q8ZQU8) resemble the crystal structures of a prototypical GT-A-fold GT enzyme, *Bacillus subtilis* SpsA (PDB ID: 1h7q), and the *Thermobacillus compostii* Rib*f*-transferase (PDB ID: 7shg), respectively. STM0725 shares 15% identity with SpsA, and the structural comparison shows a Z-score of 17.9 and an rmsd of 3.2 Å. STM0726 shares 16% identity with the prototypical Rib*f*-transferase structure, and the structures have a Z-score of 21.4 with an rmsd of 6.5 Å. The lower level of three-dimensional similarity in this STM0726 comparison arises primarily from differences in the relative orientation of the PRP and gPRT domains in the two proteins. **b** HPLC traces of in vitro reactions containing STM0725 and STM0726 PRPP donor substrate and acceptor compound **2**. Reactions were incubated for 18 h at 4 °C to prevent protein precipitation, which occurred rapidly at warmer temperatures. Reactions were stopped by the addition of an equal volume of acetonitrile and the mixture analyzed by HPLC, which detected the methoxybenzamide tag in the acceptor. Products indicated on the chromatograms were confirmed by MS (Supplementary Fig. 8) and the results from triplicate reactions were identical. **c** NMR analysis of the purified STM0725/STM0726 product from a scaled-up reaction. The determined structure, including the reducing terminal synthetic acceptor, is shown above the spectra and assignments are shown on the HSQC spectrum. Linkage points of the monosaccharides were determined by an HMBC experiment and are shown. Chemical shifts are reported in Supplementary Table 3.

residue (Fig. 4b, Supplementary Figs. 7, 8). One possible interpretation of this data is that STM0726 is sufficient for both the initiating transfer of Rib*f* and further extension of the Rib*f*-PS. In this scenario, STM0725 performs an ancillary non-catalytic role in polymerization. To pursue this, we needed to determine whether STM0726 was capable of transferring Rib*f* to a terminal Rib*f* residue, in addition to the terminal Gal*f* shown above. An enzymatic strategy was established to trap the product of a single Rib*f* addition to **2**. PRP modules of Rib*f*-transferases possess an invariable Asp required for dephosphorylation[16]. The catalytic Asp14 in the PRP domain of STM0726 was predicted by multiple sequence alignment (Supplementary Fig. 3) and AlphaFold modeling and replaced with alanine. Reactions containing STM0725 and STM0726^{D14A} with **2** resulted in a new product in the chromatogram (Fig. 4b). This reaction product (**7**) was scaled up and purified for NMR and confirmed Rib*f*-5-P linked to the 5-position of the non-reducing Gal*f* in **2** (Supplementary Fig. 10, Supplementary Table 4). Incubation of **7** with wildtype STM0726 resulted in complete dephosphorylation and, in the presence of PRPP, products were observed with one or two additional Rib*f* residues (Supplementary Fig. 11). In contrast, multiple Rib*f* residues were added when **7** was incubated with both STM0726 and STM0725. To determine the linkage of the second Rib*f* residue, **7** was dephosphorylated by incubation with STM0726 in the absence of PRPP to generate the trisaccharide **8** (Supplementary Fig. 12, Supplementary Table 5), followed by installation of a single Rib*f*-5-P residue using STM0726^{D14A}. The NMR data of the resulting phosphorylated tetrasaccharide (**9**) confirmed the terminal Rib*f*-5-P was β-(1→2)-linked (Supplementary Fig. 13, Supplementary Table 6). Collectively, the results assign no required GT catalytic activity to STM0725 in T1 antigen biosynthesis, despite its apparent GT-A-fold, and implicates STM0726 as the sole (bifunctional) Rib*f*-transferase, capable of making both β-Rib*f*-(1→2)-β-Rib*f* and β-Rib*f*-(1→5)-β-Gal*f* linkages.

STM0725 forms a heterocomplex with STM0726

The data above is consistent with STM0725 performing a structural role as a stabilizing chaperone or stimulatory partner for STM0726, rather than being a *bone fide* GT. Fulfilling this role would require STM0725–STM0726 interaction, which was examined by co-expressing epitope-tagged versions of the proteins. Purification of His_6-STM0725 on Ni-NTA resin resulted in the co-elution of STM0725 and STM0726-FLAG, as evident from the corresponding western immunoblots (Supplementary Fig. 14a). The reciprocal purification, applying the Ni-NTA elution fraction to anti-FLAG beads also demonstrated co-elution of STM0725 and STM0726. Taken together, these data show that the two proteins form a heterocomplex. AlphaFold modeling of the complex indicates a high-confidence model in which STM0725 interacts with STM0726 at an interface between the PRP and gPRT domains (Supplementary Fig. 14b), consistent with the interaction of the two proteins. Unfortunately, STM0725 alone, and copurified with STM0726, was consistently unstable in vitro. This instability led to precipitation after purification that was only partially prevented with glycerol. Thus, this characteristic precluded more extensive investigations of isolated STM0725 and its complex with STM0726.

Distribution of the T1 antigen biosynthesis locus in *Salmonella*

Examination of several *Salmonella* genomes revealed that some isolates possessed an intact T1 antigen locus, while others apparently did not. To provide more context, we examined the broader distribution of the locus by probing the *Salmonella* Foodborne Syst-OMICS database (SalFoS)(https://salfos.ibis.ulaval.ca) as a dataset. SalFoS contains genomic sequences from 2850 diverse *Salmonella* isolates from the environment, plant, and animal food products, as well as from human infections[25].

Isolates distributed across the entire *Salmonella enterica* species were found with intact versions of each of the genes spanning *STM0719-STM026* (the raw data is given in Supplementary Data 1 and a summary of the distribution among species and serovars is in Supplementary Table 7). *Salmonella enterica* serovars Enteritidis and Typhimurium are the dominant causes of non-typhoidal gastrointestinal infections by this species and are closely related. Almost all *S. enterica* Typhimurium isolates (299/318) had complete T1 clusters, and

the loci were also highly prevalent in other infection relevant serovars[26–28] including Newport (94/107), Javiana (31/38), Derby (12/17) and the monophasic clade (33/34). In contrast, isolates of serovars Enteritidis and Infantis possessed none of the T1 locus genes and no sequence remnants were found between *nei* and *pxpA*. The locus in the prevalent Heidelberg serovar (22/75) consistently contained premature stop codons in either *glf, wzt* or *wbbO*. Among the human adapted serovars, all 17 Typhi isolates contain a premature stop codon in *STM0724* that is expected to abolish T1 production. T1 clusters were well represented and expected to be functional in Paratyphi B (48/67), and Paratyphi A (12/12), but never in Paratyphi C (0/3), which all showed no remnants of T1. Outside of subspecies enterica, T1 was highly present in the IV subspecies (20/28). Parts of the T1 cluster were also found in subspecies IIIB, but none contained complete clusters due to premature stop codons present in one or more genes, depending on the isolate (0/39). Finally, subspecies II (0/3), IIIA (0/14), VI (0/2) and *bongori* (0/2) showed no hits for any of the T1 genes. To validate the lack of T1 hits (or remnants) in subspecies II, IIIA, VI and species *bongori*, GenBank was also searched. This also revealed no full T1 clusters in subspecies VI, however unlike the SalFos isolates, remnants of the T1 genes could be detected in ancestral *bongori*, IIIA and II. The *bongori* isolates (CP035676.1, CP053417.1) contained premature stop codons in *STM0725* and *STM0726*, or *STM0719* respectively, and the IIIA isolates (CP082954.1, LR134156.1) contained premature stop codons in *STM0725* and *STM0726* or *STM0724* respectively. Some II isolates were found with intact T1 clusters (*i.e.*, CP034717.1 and CP034697.1), but other isolates contained premature stop codons in *STM0724, STM0725*, or *STM0726*. Notably, the isolates with mutations in *STM0725* and/or *STM0726* may still have the capacity to produce Gal*f*-PS, as the rest of the T1 cluster is intact.

T1 antigen expression is regulated by a recombinational phase switch

Although stable expression of the T1 antigen was observed with plasmid constructs in an *E. coli* host, the original transient description of the antigen led to the hypothesis that the cluster must be regulated in some manner in *Salmonella*. This was supported by varying degrees of anti-T1 immunoreactivity in western immunoblots of whole-cell lysates prepared from several *Salmonella* isolates with or without the T1 biosynthesis locus (Supplementary Fig. 15). Immunoreactivity was enhanced in a mutant lacking its conventional O antigen, but the lack of silver-stained higher molecular weight material in the PAGE profile indicated that T1 expression was still relatively low (Supplementary Fig. 15).

Upstream of the now established T1 biosynthesis genes are three additional genes: *STM0716, STM0717* and *STM0718* (Fig. 5a). As shown above, *STM0716–STM0718* are not required for T1 antigen biosynthesis in an *E. coli* background. *STM0716* encodes a predicted tyrosine recombinase sharing 60% identity/73% similarity (92% coverage) with FimB from *E. coli* MG1655. FimB directs the phase-variable transcriptional regulation of type 1 fimbrial adhesins by mediating the inversion of a DNA segment encompassing the *fim* promoter, thereby flipping the *fim* operon between OFF and ON positions (reviewed in ref. 29). The switch is composed of inverted nine base-pair repeats, flanking a 314-nucleotide segment that contains the *fim* promoter. Given the similarity shared by STM0716 and FimB, the region following *STM0716* was searched for inverted repeats. A 269 base-pair region flanked by nine base-pair inverted repeats was identified and hypothesized to be a potential phase switch (Fig. 5a).

To investigate the physiological consequences of the phase switch candidate for T1 antigen expression, we first sought to determine whether this region is invertible in its native genomic context. qPCR was performed on *S. enterica* serovar Typhimurium SL3770 using primers designed to probe the two states (Fig. 5b), subsequently determined to be the OFF and ON states (see below). qPCR detected both states, suggesting a dynamic process. Furthermore, quantitation was performed by comparison to a standard curve using synthesized ON or OFF state DNA templates, revealing that the OFF state was approximately 2–3 times more abundant than the ON state (Fig. 5b, Supplementary Fig. 16, Supplementary Table 8). This bias is supported by examination of sequence data for this region in *Salmonella* isolates deposited in GenBank; all but one isolate (CP037891.1) reflect the predominant OFF state.

Because the switch is located well upstream of the first T1 biosynthesis gene (*STM0719*) and the intervening region contains *STM0717* and *STM0718* genes that play no apparent role in T1 production, it was essential to confirm that the switch region directly regulated downstream genes responsible for T1 antigen production. To facilitate this, plasmid pWQ1128 containing *STM0716–STM0726* was modified by deleting the vector's *tetR* promoter and inserting a transcriptional terminator upstream of STM0716 to prevent any transcriptional read through from the vector. *STM0716* was then deleted to prevent DNA inversion, and the invertible promoter region was cloned in both orientations. In the OFF orientation, no T1 antigen could be detected in whole-cell lysates by either silver staining or western immunoblotting. However, when the switch region was in the ON state, robust T1 antigen production was observed (Fig. 5c). Co-transformation of these plasmids with a compatible plasmid carrying *STM0716* (encoding the putative recombinase) restored the phase-variation process in *E. coli*. T1 antigen production was now detected in bacteria containing the OFF-state construct and Sanger sequencing of the switch region showed a mixed population of DNA, consistent with part of the population changing to an ON-state (Fig. 5d). Conversely, when STM0716 was expressed in the presence of the ON-state plasmid, a reduction of the T1 antigen was detected in the silver stain and western immunoblot and Sanger sequencing again showed a mixed population of DNA (Fig. 5d). Together, these results show that the region proceeding *STM0716* is indeed a phase-variable promoter region that can be flipped in a bidirectional manner by STM0716 to regulate expression of downstream genes directing biosynthesis of the T1 antigen.

Discussion

In this study, the long-overlooked T1 antigen of *Salmonella* is reinvestigated and its distribution, structure and relationship to OPS was resolved. We show that the T1 antigen is a widely distributed lipid A–core-linked polysaccharide in *Salmonella* but its abundance is typically low (and frequently undetectable) compared to the serotype-specific O antigen when wildtype cells are grown under routine laboratory conditions. This is evident by the increased amount of T1 antigen in whole cell lysates of a *S. enterica* serovar Typhimurium mutant that lacks the serogroup B O antigen due to mutation of the gene for the phosphoglycosyltransferase (*wbaP*) (Supplementary Fig. 1a). The precise linkage of T1 to lipid A-core oligosaccharide was not pursued here, but the requirement of WaaL for T1 ligation is consistent with the attachment site being the same as native O-antigen to give the structure shown in Supplementary Fig. 18. The organization of a lipid A–core-linked glycan into tandem structural domains is a rare occurrence; the large majority of OPS are composed of a single repeat unit structure. Prototypes for this unusual tandem format are found in *K. pneumoniae* serotype O1, where the O1 antigen with a disaccharide repeat unit (→3)-α-D-Gal*p*-(1→3)-β-D-Gal*p*-(1→) is added to the non-reducing terminus of the O2 antigen (→3)-α-D-Gal*p*-(1→3)-β-D-Gal*f*-(1→)[30,31]. A similar two domain structure occurs in *Klebsiella* serotype O2ac[31,32]. In an intriguing coincidence, biosynthesis of these *Klebsiella* O antigens and the *Salmonella* T1 antigen are all assembled on the same undecaprenyl diphosphate linked trisaccharide generated by WecA–WbbN–WbbO (Fig. 1b). This structural motif apparently provides a versatile foundation for glycan diversification.

https://doi.org/10.1038/s41467-024-50957-y

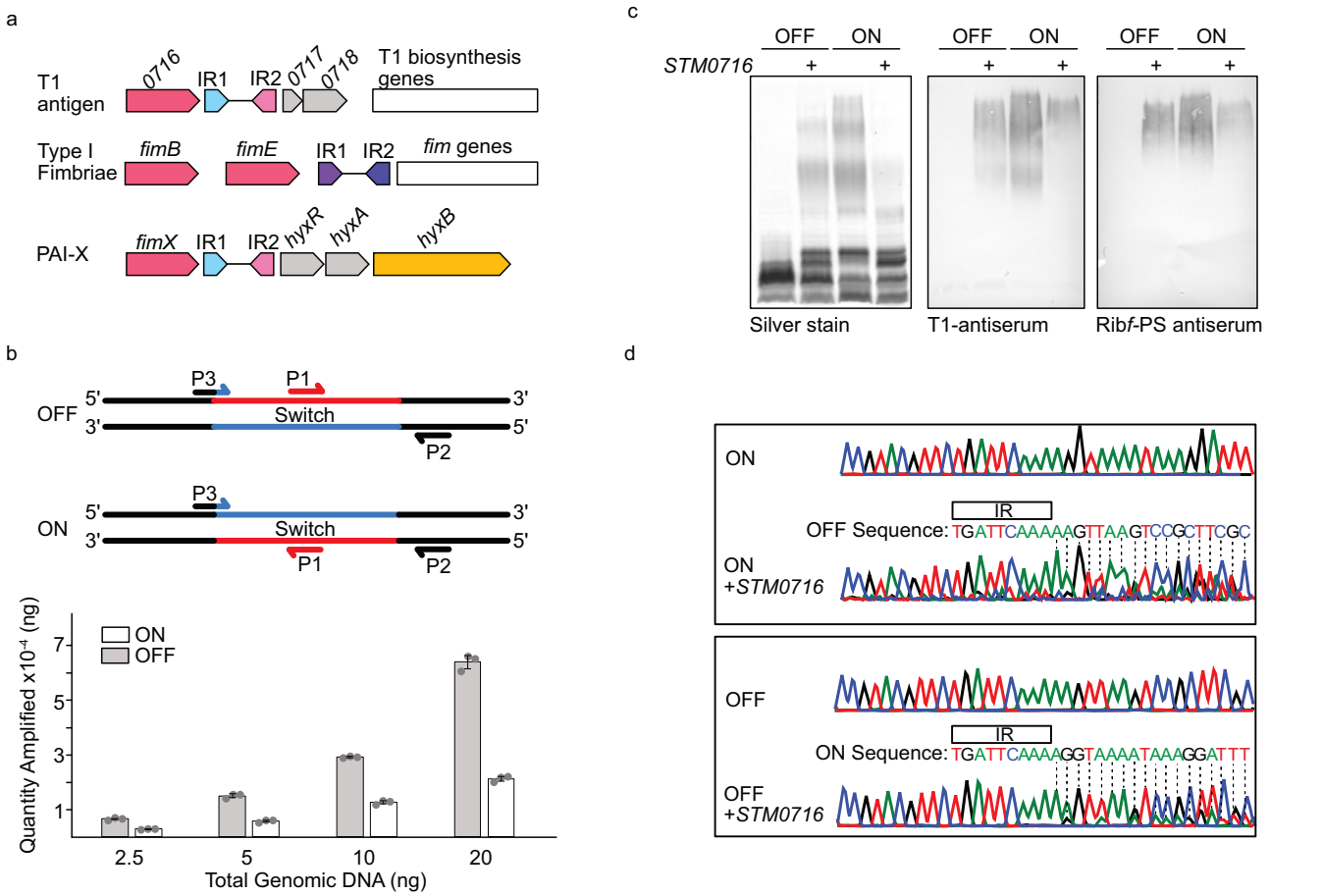

Fig. 5 | The T1 cluster is regulated by a recombination-mediated phase variation mechanism. a Genetic organization of the genomic regions encompassing genetic switches for T1 antigen biosynthesis, and the expression of *E. coli* type I fimbriae (shown from MG1655; U00096.3), and a putative autotransporter (HyxB) encoded on *E. coli* pathogenicity island PAI-X (from *E. coli* O157:H7; EDL933). STM0716 is an ortholog of FimB and FimX, which are tyrosine recombinases responsible for flipping a promoter region preceding the *fim* and *hyxRAB* operons. In each case, the invertible region is flanked by nine base-pair inverted repeats (IR1 and IR2). **b** Schematic showing the experimental strategy to identify ON and OFF states of the invertible switch using qPCR. Primer P1 (in the middle of the switch) serves as the forward primer or reverse primer, depending on the switch orientation and acts with one of the two primers (P2 or P3) located outside the switch to generate a diagnostic amplification product. Standard curves were constructed using synthesized (fixed) templates representing the OFF or ON states, and used to calculate the relative amounts of each state from genomic DNA templates. Bar graphs represent the average amounts detected, with individual data points plotted on top of the bar graph. Samples and standards were analyzed in triplicate and error bars indicating standard deviation is shown using the average DNA quantity detected in each sample as the center point. qPCR was performed on two separate genome preparations with comparable results. **c** T1 antigen phenotypes resulting from engineered ON and OFF states of the T1 phase switch. Whole-cell lysates were examined by SDS-PAGE and western immunoblotting. pWQ1128 (carrying *STM0716*-STM0716) was cloned to contain the switch region locked in either the ON or OFF state and *STM0716* was deleted to create pWQ1133 (ON) and pWQ1134 (OFF). Each plasmid was also co-transformed with a plasmid carrying *STM0716* in *trans*. The presence of the T1 antigen was then probed by silver staining and western blotting with the specific antisera. These analyses were performed in triplicate. Source data are provided as a Source Data file. **d** Sanger sequencing results for the switch regions from plasmids recovered from the cultures examined in **c**. No phase variation was detected in the absence of STM0716, evident in the uniform chromatographic species. Expression of STM0716 resulted in a mixed population of species reflecting both orientations of the switch. Most importantly, expression of STM0716 in the ON state led to a predominant OFF state, whereas expression of STM0716 in the OFF state resulted in a minor amount of ON state.

Biosynthesis of the Gal*f*-PS component of T1 antigen is performed by the STM0724 polymerase, an ortholog of GlfT2. GlfT2 forms the backbone of the mycobacterial cell wall arabinogalactan and possesses a single catalytic-site GT that builds an alternating β-(1→5) and β-(1→6) Gal*f* galactan[23]. A query of other Gal*f*-homopolymers was conducted to investigate the distribution of GlfT2-like enzymes. Examples were identified in a diverse set of organisms, and included one, two and three linkage Gal*f*-PSs (Supplementary Fig. 17). In all cases, genes were found for GlfT2 orthologs sharing identities ranging from 23–38% (38–53% similarity). AlphaFold modeling showed that each of the orthologs have the distinct GlfT2 fold indicating all operate with a single catalytic site, regardless of the number of linkage types in the products[33]. GlfT2 is a processive enzyme, where the glycan is retained by the enzyme during elongation until 30–40 residues have been added. Recent data has provided insight into the basis for dual-linkage (β-(1→5) and β-(1→5)) specificity[34,35] as well as a potential mechanism for translocating the chain through the active site during processive polymerization[36]. The mixed products in *Salmonella* T1 antigen suggests STM0724 exhibits less intrinsic fidelity than GlfT2. Establishing the extent to which the mechanistic details apply to other family members that generate products with 1–3 different linkage configurations will require detailed biochemistry and structures of enzyme–substrate complexes but the present study identifies important representatives to pursue.

STM0726 is the polymerase for the biosynthesis of the Rib*f*-PS. This enzyme possesses the dual gPRT–PRP activity and the prototypical fold found in canonical glycan Rib*f*-transferases[16], but is the first representative known to transfer more than a single Rib*f* sugar, or

possess the ability to make more than one type of glycosyl linkage. In this sense, it is analogous to the STM0724 Gal*f*-PS polymerase. However, STM0726 is able to transfer Rib*f* to acceptors with either terminal Gal*f* or Rib*f* residues and, in that regard, is also similar to GlfT1, another bifunctional enzyme involved in mycobacterial arabinogalactan biosynthesis. GlfT1 synthesizes Gal*f*–Gal*f* and Gal*f*–Rha*p* (rhamnopyranose) linkages, although GlfT1 does not continue polymerization[23]. How STM0726 synthesizes two distinct glycosidic linkages via a single catalytic site is a topic for further investigation, currently limited by the instability of the purified enzyme. An equally fascinating question is why optimal activity of the STM0726 polymerase depends on another GT-like protein that lacks GT catalytic activity. Further insight into the relationship between STM0725 and STM0726 was obtained by a tBLASTn search using STM0725 as a query. This returned 121 hits (excluding *Salmonella*) with identities ranging from 54% to 22%. Surprisingly, in 91 of these hits, the genes for the STM0725 orthologs occurred adjacent to a gene encoding an STM0726 ortholog. These were found in genomes of diverse species and genera, as well as in a bacteriophage belonging to the *Caudoviricetes*. Orthologous pairs were present in *Serratia marcescens* O16[37] and *Liberibacter crescens* BT-1[38], which are known to produce Rib*f*-homopolymers. Interestingly, the STM0725 orthologs were separated into two phylogenetically distinct clades of proteins (Supplementary Fig. 5b). STM0725 itself was grouped with a clade of proteins that all lacked a candidate catalytic base in the position expected in GT-A enzymes. In contrast, the second clade all possessed a conserved Asp typical of functional GT-A enzymes, suggesting these proteins may retain catalytic activity. *L. crescens* BT-1 produces a complex OPS with two structural components; one involves short spans of poly-Rib*f*, while the other is a disaccharide repeat composed of Gal*p* and Rib*f*[38]. The STM0726 ortholog (B488_06780) is the only identifiable Rib*f*-transferase encoded the BT-1 genome. Nevertheless, an STM0725 ortholog does not seem to be an essential requirement in all poly-Rib*f* biosynthesis systems. The *Helicobacter pylori* SS1 LPS was found to contain short poly-Rib*f*[39], and a single STM0726 ortholog (but no STM0725 ortholog) is encoded by the genome.

Salmonella isolates with intact versions of all the required T1 antigen biosynthetic genes were found distributed across the *Salmonella enterica* species. The distribution (Supplementary Table 7) implies that the T1 antigen is not required for typhoidal infections and, while widespread in isolates causing non-typhoidal gastrointestinal infections and in environmental isolates, possession of the genetic locus is not universal. The appearance of parts of the locus in *S. bongori* and *S. enterica* subspecies IIIA suggests it is not a recent acquisition during the evolution of the genus. A closer look at the *S. bongori* isolates revealed that both isolates from SalFoS contained a different polysaccharide cluster located between *nei* and *pxpA* downstream of copies of *STM0716–0718* and an identifiable phase switch (Supplementary Fig. 19). The genes in the cluster closely resemble the well characterized *E. coli* O9/O9a locus[40], which directs production of a polymannose OPS (Supplementary Table 9). The potential production of this glycan in *Salmonella* has not been investigated. Notably, this O9/O9a like cluster was also identified in 10 other isolates in the SalFoS database including representatives from subspecies *salamae*, and *diarizonae*. The number of hits for the O9/O9a-like cluster were substantially less than for the T1 antigen, but this could reflect the bias towards human-related *Salmonella* isolates (where the T1 antigen is predominant) in the collection.

One of the key questions we had when this study was initiated is why the T1 antigen was originally described as transient and why it has not been detected during the many investigations of O antigen structures in *Salmonella*. Part of this may be answered by the expected competition between the T1 antigen and the serotype-specific OPS for lipid A–core. However, the other factor is the recombination-mediated phase variation of T1 antigen expression, which is unprecedented for a lipid A-core-linked glycan. Conventional *Salmonella* O antigens are thought to result from (mostly) constitutive expression of biosynthesis genes. There is some modest fine-tuning of the products through OxyR regulator-mediated on/off transcriptional control of the *opvAB* locus, which alters the distribution of chain lengths of OPS chains in some way and affects sensitivity to bacteriophage and serum[41]. In addition, OxyR and the DNA methylase (Dam) are implicated in epigenetic regulation of an OPS-modification process in some Salmonella OPS[42]. This modification involves adding side chain glucose residues in a periplasmic post-polymerization reaction, directed by enzymes encoded by a lysogenic bacteriophage. A comparable recombinational promoter-inversion mechanism has been described in regulating the multiple cell surface exopolysaccharides produced by *Bacteroides fragilis*[43]. The process is directed by a serine site-specific recombinase[44] unrelated to STM0716. While the origin of the regulatory system for the T1 antigen is unknown, a similar locus organization is found within *E. coli* Pathogenicity Island X (PAI-X) (Fig. 5d). The locus in PAI-X possesses a gene encoding an STM0716 ortholog (FimX; 72% identity, 84% similarity, 93% coverage) upstream of a candidate phase switch possessing inverted repeats whose sequence is identical to those in the T1 antigen locus. The switch precedes the downstream genes: *hyxA, hyxR*, encoding putative transcriptional regulators and *hyxB* encoding a putative autotransporter adhesin. STM0718 shares 39% identity/60% similarity (81% coverage) with HyxA but the physiological significance (if any) is currently unknown.

The presence of a highly conserved T1-antigen locus with complex regulation in many *Salmonella* genomes, suggests a potentially important role in fitness. Transposon directed insertion sequencing (TraDIS) in *Salmonella* Typhimurium revealed preservation of an unknown polysaccharide locus (identified here as a T1 antigen biosynthesis)[45–47]. While this supports a role in fitness, some caution is needed in interpretation because the T1-biosynthesis gene cluster has higher AT content than the rest of the genome. This is a known region of H-NS binding[48], which can compromise TraDIS data[49]. Several transcriptomic and proteomic studies in *Salmonella* Typhimurium identify no major changes in the abundance of T1 biosynthesis transcripts or proteins under the conditions tested. One study indicates a modest elevation in transcription of the T1 - antigen biosynthesis genes in *S. Typhimurium* SL1344, following opsonization (12h) by macrophages. This is accompanied by a reduction in transcription of O-antigen biosynthesis genes, but the biological implications are uncertain because SL1344 is expected to be unable to produce T1 antigen due to a mutation in *glf*. Further investigation is therefore needed to provide insight[50] into possible conditions where a particular ON- or OFF-state would be beneficial[51–53]. Furthermore, it has yet to be established whether the conservation of T1-antigen production reflects importance in the pathogen lifestyle, or if it is required for fitness and survival in an environment outside the host. Nevertheless, the findings of the current study provide the essential foundation to investigate these broader questions about the physiological role of T1 antigen in *Salmonella*.

Methods

Bacterial strains and DNA methods

Bacterial strains and plasmids used in this study are listed in Supplementary Table 10. All bacteria were cultured in LB supplemented with 50 μg/ml kanamycin, 100 μg/ml ampicillin or 34 μg/ml chloramphenicol, where applicable. Genomic DNA was purified using the PureLink Genomic DNA mini kits from ThermoFisher (Invitrogen). DNA fragments were generated using KOD Hot Start DNA polymerase (Novagen) for cloning, site-directed mutagenesis and gene deletions/manipulations. The sequences of the oligonucleotide primers are provided in Supplementary Table 11. PCR products and plasmids were purified using GeneJET PCR or plasmid purification kit (ThermoScientific), respectively. Plasmid constructs were confirmed using

Sanger sequencing in the Advanced Analysis Center at the University of Guelph, or by whole plasmid sequencing from Plasmidsaurus (https://www.plasmidsaurus.com).

Bioinformatics analyses

Initial protein domain predictions were made by BLAST using the Conserved Domain Database. AlphaFold models were obtained from Uniprot or created manually in ColabFold[54] when not available in Uniprot. Manually created AlphaFold models are provided in figshare (see data availability). Multiple sequence alignments were performed using ClustalW v.2.0[55], and alignments were visualized using ESPript v.3.0[56]. 3D structure alignments were performed in Dali[57] and protein structures were visualized in Pymol v.2.1. Phylogenetic studies were conducted using Phylogeny.fr[58] using a bootstrap value of 100. The Salmonella Foodborne Syst-OMICS (SalFoS – https://salfos.ibis.ulaval.ca) database was used as a data source to investigate T1 antigen distribution across genus *Salmonella*. Sequencing files were downloaded and queried locally using tBLASTn with an e-value cut-off of $x10^{-20}$. BLAST searches were performed for each protein in the T1 cluster. Only those isolates that contained full-length homologs for all proteins for T1 production were considered T1 positive, while any isolate with mutation(s) resulting in truncations in one or more of the T1 proteins were considered T1 negative. Bit scores were used as an indicator for intact protein sequence. Bit scores lower than a determined threshold (cutoffs indicated for each protein in Supplementary Data 1) corresponded to a truncated protein product that was considered T1 negative. In a small number of cases premature stop codons resulted in truncated proteins that still had high bit scores (indicated in Supplementary Data 1), and these were manually assessed and designated as T1 negative.

LPS and T1 antigen purification

10 L cultures of *E. coli* DH5α harboring pWQ1129 were grown overnight at 37 °C. Cells were collected and LPS was purified using the hot-phenol method[59]. After phase separation, the aqueous layer containing LPS was dialyzed against water to remove residual phenol. The pH was then reduced to 4 to precipitate proteins and nucleic acids, which were removed by centrifugation at 12,000 × *g* for 20 mins. The supernatant was dialyzed against deionized water until neutral pH was achieved, and then concentrated by a rotovap and lyophilized. OPS was isolated by hydrolysis of purified LPS at 100 °C in 2 % (v/v) acetic acid for two hours, followed by centrifugation to remove the lipid precipitate. The carbohydrate-containing supernatant was then separated on a Sephadex G-50 superfine column (2.5 × 75 cm) in 50 mM pyridinium acetate buffer (pH 4.5) at a flow rate of 0.6 mL/min. Elution was monitored with a Smartline 2300 refractive index detector (Knauer) and fractions were collected at 10 min intervals. Fractions corresponding to peaks were pooled, frozen and lyophilized. The OPS fraction was determined to contain a substantial amount of contaminating enterobacterial common antigen (ECA). Due to the acidic nature of ECA, it was separated from the T1 OPS using anion exchange using a DEAE Fast Flow (GE Healthcare) column eluted with a 5–500 mm sodium phosphate gradient at pH 6.3. The flowthrough containing neutral T1 OPS was collected, concentrated and desalted with a G-50 PD-10 column eluting in water. The final yield was ~16 mg.

Antiserum production

Animal handling and immunization was performed at the University of Guelph Central Animal Facility. Rabbit anti-T1 was produced using formalin killed *E. coli* DH5α expressing T1 (pWQ1128). Killed cells were resuspended in 0.85% (w/v) NaCl at ~10^8 cfu/ml and mixed 1:1 with Freund's incomplete adjuvant (Sigma). A New Zealand white rabbit received an intramuscular injection every two weeks for six weeks total. Blood was collected after six weeks, and the serum was separated and stored at −80 °C. To adsorb antisera and remove antibodies against the T1 Gal*f*-PS, *E. coli* DH5α cells expressing pWQ1130 (Δ*STM0725–STM0726*) were grown overnight and 200 mL of cells were resuspended in 16 mL of sterile PBS. The cell pellet from 4 mL of culture was collected by centrifugation, resuspended in 5.2 mL of antiserum and incubated at 37 °C for 1 h. Cells were then removed, and the process was repeated three more times. All T1 antisera generated here are available upon request.

Expression and purification of proteins

Proteins expressed from pET-vector plasmids were expressed by induction with 0.5 mM 1-thio-β-D-galactopyranoside (IPTG) and those from pBAD-vector plasmids were induced with 0.2% L-arabinose. The induced cultures were grown overnight at 18 °C. The cells were harvested and resuspended in buffer A (50 mM Tris pH 7.5, 500 mM NaCl, 20% glycerol). Cell disruption was performed with an Avestin C3 Emulsiflex. Lysates were cleared by sequential centrifugation at 12,000 × *g* for 20 min, followed by 100,000 × *g* for 1 h to remove membranes. The supernatant was applied to Ni-NTA agarose with a 2 mL bed volume. The beads were washed with 10 bed volumes of buffer A containing 10 mM imidazole, followed by 10 bed volumes of buffer A containing 30 mM imidazole. Bound proteins were eluted with buffer A containing 250 mM imidazole. The eluates were concentrated with a 10,000 MWCO concentrator (Sartorius) and buffer exchanged into buffer A using PD-10 Sephadex G-25 columns (Cytiva). Purified protein concentrations were estimated using absorbance at 280 nm, with molecular weights and extinction coefficient estimations from ProtParam.

Protein interaction studies

Protein co-purification was performed by exploiting the epitope tags on His_6-STM0725 and STM0726-FLAG. Cultures expressing pWQ1142 were prepared and lysed as described above. The cell-free lysate was passed through a Ni-NTA agarose column and the column was washed as described above. The eluate was concentrated, buffer exchanged to buffer A, and then incubated with 50 μL of anti-FLAG M2 magnetic beads (Sigma) beads for 1 h. The beads were washed three times with 1 mL of buffer A and the protein was eluted by resuspending the beads in 40 μL 100 mM glycine buffer pH 2.5.

Quantitative PCR

Quantitative PCR was performed using the PowerTrack SYBR Green Master Mix (ThermoFisher). Genomic DNA was purified using the PureLink genomic DNA kit (Invitrogen). Amplification reactions were performed using the QuantStudio 3 standard program (Applied Biosystems) and analyzed using the Design and Analysis Software (ThermoFisher Connect). DNA fragments with locked ON or OFF switch orientations were synthesized by IDT. Three technical replicates were performed with each sample and standard, and values reported as the average with standard deviation. Results are representative of two biological replicates from separate genomic preparations.

SDS-PAGE and immunoblotting

LPS samples were analyzed using proteinase-digested whole-cell lysates as described in ref. 60. Cells were collected from a culture volume equivalent to 1 OD_{600nm} unit and resuspended in loading buffer, boiled for 10 min, and subsequently treated with proteinase K at 55 °C for 1 h. Samples were separated on 12% Tris-glycine gels and visualized using silver staining or western blotting after transfer to nitrocellulose membranes (Protran, GE Healthcare). SDS-PAGE of purified proteins was conducted by mixing an equal volume of protein with loading buffer and separating in the way described above and the SeeBlue Plus2 protein ladder provided molecular weight markers (Invitrogen). Coomassie-blue staining was performed using the Pierce Power Stainer. Western immunoblot transfer was conducted with a constant current of 200 mA for 45 min (LPS samples) or 350 mA for 60 min (protein) in buffer containing 25 mM Tris, 150 mM glycine, and

20% (v/v) methanol. The membranes were blocked in 5% skim milk for LPS analysis (BD Difco), or 5% bovine serum albumin for protein analysis (Cytiva); the blocking agent was prepared in TBST (10 mM Tris-Cl, pH 7.5, 150 mM NaCl, 0.005% (v/v) Tween 20). Rabbit T1 antiserum was used as a 1:500 dilution in skim milk-TBST, and mouse anti-His_5 (Qiagen - Catalog number: 34660 Lot: 172019951) antibody or mouse anti-FLAG (Sigma - Catalog number: F3165 Lot: SLCM4081) were each used at a 1:1000 dilution in BSA-TBST. The secondary antibodies were goat anti-rabbit-conjugated alkaline phosphatase (1:3000) (Cedarlane - Catalog Number: 111-055-003 Lot: 86-172-030422.) or goat anti-mouse-conjugated alkaline phosphatase (Jackson laboratories - Catalog Number: 115-055-003 Lot: 138204.), and detection was achieved using nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Roche Applied Science).

In vitro GT reactions

Standard reactions were performed in 20 µL volumes containing 50 mM HEPES pH 7.5, 20 mM $MgCl_2$, 10% glycerol, 1 mM acceptor, 5 mM PRPP, 5 mM UDP-Gal*p* and 50 µm Glf, and 5 mM NADH. Reaction mixtures contained 100 µM enzyme. STM0724 reactions were incubated at 30 °C for 1 h while STM0725 and STM0726 reactions were incubated at 4 °C for 18 h. Reactions were stopped by adding an equal volume of acetonitrile and the resulting protein precipitate was removed by centrifugation at 12,000 × *g* for 5 min. Large-scale reactions were performed using the same components and concentrations in 1.6 mL volumes, but glycerol was excluded. In the absence of glycerol, STM0724 and STM0725 precipitated quickly, so reactions were performed on freshly purified protein maintained at 4 °C for the entirety of the purification and reaction preparation. For NMR analysis, products with single-sugar additions were purified from the reaction mixture using a SepPak C8 column eluted in 50% methanol. In vitro polysaccharide products made by STM0724 and STM0725 + STM0726 were fractionated by size exclusion chromatography using a Superdex 200 Increase 10/300 column eluted with 100 mM ammonium acetate. The high molecular weight fractions were retained and purified using a SepPak C8 as above. Reactions analyzing STM0725 and STM0726 containing UDP-Gal*p* and NADH were also first applied to a SepPak C8 (as above) to eliminate overlapping peaks from NADH.

The reaction components were separated by HPLC using an Agilent 1260 Infinity II LC system. A 10 µL aliquot of the reaction mixture was injected and separated with a GLYCOSEP N column (4.6 × 250 mm, Prozyme). Buffer A contained 10 mM ammonium formate pH 4.4 in 80% acetonitrile, buffer B contained 30 mM ammonium formate pH 4.4 in 40% acetonitrile and solvent C contained 0.5% formic acid. Separation was accomplished with a linear gradient of 100% A to 100% B over 160 min (0.4 mL/min), followed by a 2 min gradient of 100% B to 100% C.; returning to 100% A over 2 min and holding for 15 min (1 mL/min); followed by 0.4 mL/min for 5 min in A. The column temperature was 30 °C and elution was monitored at 260 nm. All HPLC was analyzed using Agilent OpenLAB revision A.02.16.

Mass spectrometry

LC-MS was performed in the University of Guelph Advanced Analysis Center with an Agilent 1260 HPLC interfaced with an Agilent UHD 6530 Q-TOF mass spectrometer. LC separation was performed on a C18 column (Agilent Poroshell 120, EC-C18 50 × 3.0 mm, 2.7 µm), using water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B). A 10 µL aliquot of reaction mixture was injected and a gradient elution was performed with a flow rate of 0.4 mL/min starting with 10% B for 1 min; increasing to 100% B in 29 min; a column wash at 100% B for 5 min; followed by a 20 min re-equilibration. The mass spectrometer electrospray capillary voltage was maintained at 4.0 kV, and the drying gas temperature at 250 °C, with a flow rate of 8 L/min. Nebulizer pressure was 30 psi and the fragmentor was set to 160. Nozzle, skimmer, and octapole RF voltages were set at 1000 V, 65 V and 750 V, respectively. Nitrogen (purity >98%) was used as nebulizing, drying, and collision gas. Ionization of compounds was performed in negative mode. The mass axis was calibrated using the Agilent tuning mix HP0321 (Agilent technologies) prepared in acetonitrile. All mass spectrometry analysis was performed with Agilent MassHunter Qualitative Analysis version 10.0.

Nuclear magnetic resonance spectroscopy

NMR analyses of OPS and in vitro reaction products were conducted in the University of Guelph Advanced Analysis Center. Analysis was performed using a Bruker Avance III 600 MHz spectrometer, equipped with a cryoprobe. Samples were deuterium exchanged twice by resuspending in 99.0% D_2O and lyophilizing. Samples were prepared in 99.96% D_2O with chemical shifts referenced to a 3-trimethylsilylpropanoate-2,2,3,3-d_4 (δ_H 0 ppm, δ_C −1.6 ppm) internal standard. The mixing times for TOCSY and NOESY experiments were 80 ms and 200 ms, respectively. All spectra were recorded at 25 °C. NMR analysis was performed using Bruker TopSpin version 4.2.0.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw data for mass spectrometry, NMR and ColabFold generated AlphaFold models are available at figshare [https://figshare.com/projects/Structure_biosynthesis_and_regulation_of_the_T1_antigen_a_phase-variable_surface_polysaccharide_conserved_in_many_Salmonella_serovars/211810]. The genome accession used for cloning T1 was from *Salmonella enterica* SL3770 (NC_003197.2). All other genome, protein or PDB accessions used for bioinformatic purposes are presented in the relevant figures. Source data are provided with this paper for western blots, SDS-PAGE and qPCR, and are available in the Source Data File. Source data are provided with this paper.

References

1. Crump, J. A., Sjölund-Karlsson, M., Gordon, M. A. & Parry, C. M. Epidemiology, clinical presentation, laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive *Salmonella* infections. *Clin. Microbiol. Rev.* **28**, 901–937 (2015).
2. Lamas, A. et al. A comprehensive review of non-enterica subspecies of *Salmonella enterica*. *Microbiol. Res.* **206**, 60–73 (2018).
3. Issenhuth-Jeanjean, S. et al. Supplement 2008-2010 (no. 48) to the White-Kauffmanne-Le Minor scheme. *Res. Microbiol.* **165**, 526–530 (2014).
4. Whitfield, C., Williams, D. M. & Kelly, S. D. Lipopolysaccharide O-antigens-bacterial glycans made to measure. *J. Biol. Chem.* **295**, 10593–10609 (2020).
5. Liu, B. et al. Structural diversity in *Salmonella* O antigens and its genetic basis. *FEMS Microbiol. Rev.* **38**, 56–89 (2014).
6. Seif, Y., Monk, J. M., Machado, H., Kavvas, E. & Palsson, B. O. Systems biology and pangenome of *Salmonella* O-antigens. *mBio* **10**, 10–1128 (2019).
7. Rai, A. K. & Mitchell, A. M. Enterobacterial common antigen: synthesis and function of an enigmatic molecule. *mBio* **11**, 1–19 (2020).
8. Gunn, J. S., Bakaletz, L. O. & Wozniak, D. J. What's on the outside matters: the role of the extracellular polymeric substance of gram-negative biofilms in evading host immunity and as a target for therapeutic intervention. *J. Biol. Chem.* **291**, 12538–12546 (2016).
9. Keestra-Gounder, A. M., Tsolis, R. M. & Bäumler, A. J. Now you see me, now you don't: the interaction of *Salmonella* with innate immune receptors. *Nat. Rev. Microbiol.* **13**, 206–216 (2015).

10. Kauffmann, F. A new antigen of *Salmonella* paratyphi B and *Salmonella typhimurium*. *Acta Pathol. Microbiol. Scand.* **39**, 299–304 (1956).
11. Berst, M. et al. Structural investigations on T1 lipopolysaccharides. *Eur. J. Biochem* **11**, 353–359 (1969).
12. Lindberg, A. A., Sarvas, M. & Mäkelä, P. H. Bacteriophage attachment to the somatic antigen of *Salmonella*: effect of o-specific structures in leaky R mutants and s, t1 hybrids. *Infect. Immun.* **1**, 88–97 (1970).
13. Sarvas, M. & Nikaido, H. Biosynthesis of T1 antigen in *Salmonella*: origin of D-galactofuranose and D-ribofuranose residues. *J. Bacteriol.* **105**, 1063–1072 (1971).
14. Valtonen, V. V., Sarvas, M. & Mäkelä, P. H. The effect of T1 antigen on the virulence of *Salmonella* strains for mice. *J. Gen. Microbiol.* **69**, 99–106 (1971).
15. Nikaido, H. & Sarvas, M. Biosynthesis of T1 antigen in *Salmonella*: Biosynthesis in a cell-free system. *J. Bacteriol.* **105**, 1073–1082 (1971).
16. Kelly, S. D. et al. The biosynthetic origin of ribofuranose in bacterial polysaccharides. *Nat. Chem. Biol.* **18**, 530–537 (2022).
17. Liu, D. & Reeves, P. R. *Escherichia coli* K12 regains its O antigen. *Microbiology (N.Y)* **140**, 49–57 (1994).
18. Liston, S. D., Ovchinnikova, O. G. & Whitfield, C. Unique lipid anchor attaches VI antigen capsule to the surface of *Salmonella enterica* serovar Typhi. *Proc. Natl. Acad. Sci. USA* **113**, 6719–6724 (2016).
19. Bock, K. & Pedersen, C. Carbon-13 nuclear magnetic resonance spectroscopy of monnosaccharides. *Adv. Carbohydr. Chem. Biochem.* **41**, 27–66 (1983).
20. Imamura, A. & Lowary, T. Chemical synthesis of furanose glycosides. *Trends Glycosci. Glycotechnol.* **23**, 134–152 (2011).
21. Köplin, R., Brisson, J.-R. & Whitfield, C. UDP-galactofuranose precursor required for formation of the lipopolysaccharide O antigen of *Klebsiella pneumoniae* serotype O1 is synthesized by the product of the *rfbD* KPO1. *Gene. J. Biol. Chem.* **272**, 4121–4128 (1997).
22. Clarke, B. R. et al. A bifunctional O-antigen polymerase structure reveals a new glycosyltransferase family. *Nat. Chem. Biol.* **16**, 450–457 (2020).
23. Belánová, M. et al. Galactosyl transferases in mycobacterial cell wall synthesis. *J. Bacteriol.* **190**, 1141 (2008).
24. Taujale, R. et al. Deep evolutionary analysis reveals the design principles of fold a glycosyltransferases. *Elife* **9**, e54532 (2020).
25. Emond-Rheault, J. G. et al. A Syst-OMICS approach to ensuring food safety and reducing the economic burden of salmonellosis. *Front Microbiol* **8**, 269584 (2017).
26. Ferrari, R. G. et al. Worldwide epidemiology of *Salmonella* serovars in animal-based foods: a meta-analysis. *Appl. Environ. Microbiol.* **85**, e00591–19 (2019).
27. Boore, A. L. et al. *Salmonella enterica* infections in the United States and assessment of coefficients of variation: A novel approach to identify epidemiologic characteristics of individual serotypes, 1996–2011. *PLoS One* **10**, e0145416 (2015).
28. Government of Canada. *National Enteric Surveillance Program Annual Summary 2019: Public Health Agency of Canada*. (2020).
29. Schwan, W. R. Regulation of fim genes in uropathogenic *Escherichia coli*. *World J. Clin. Infect. Dis.* **1**, 17 (2011).
30. Whitfield, C., Richards, J. C., Perry, M. B., Clarke, B. R. & MacLean, L. L. Expression of two structurally distinct D-galactan O antigens in the lipopolysaccharide of *Klebsiella pneumoniae* serotype O1. *J. Bacteriol.* **173**, 1420–1431 (1991).
31. Vinogradov, E. et al. Structures of lipopolysaccharides from Klebsiella pnuemoniae: Elucidation of the structure of the linkage region between core and polysaccharide O chain and identification of the residues at the non-reducing termini of the O chains. *J. Biol. Chem.* **277**, 25070–25081 (2002).
32. Kelly, S. D. et al. Klebsiella pneumoniae O1 and O2ac antigens provide prototypes for an unusual strategy for polysaccharide antigen diversification. *J. Biol. Chem.* **294**, 10863–10876 (2019).
33. Wheatley, R. W., Zheng, R. B., Richards, M. R., Lowary, T. L. & Ng, K. K. S. Tetrameric structure of the GlfT2 galactofuranosyltransferase reveals a scaffold for the assembly of Mycobacterial arabinogalactan. *J. Biol. Chem.* **287**, 28132 (2012).
34. Poulin, M. B. & Lowary, T. L. Chemical insight into the mechanism and specificity of GlfT2, a bifunctional galactofuranosyltransferase from Mycobacteria. *J. Org. Chem.* **81**, 8123–8130 (2016).
35. Yamatsugu, K., Splain, R. A. & Kiessling, L. L. Fidelity and promiscuity of a Mycobacterial glycosyltransferase. *J. Am. Chem. Soc.* **138**, 9205–9211 (2016).
36. Janoš, P., Tvaroška, I., Dellago, C. & Koča, J. Catalytic mechanism of processive GlfT2: Transition path sampling investigation of substrate translocation. *ACS Omega* **5**, 21374–21384 (2020).
37. Oxley, D. & Wilkinson, S. G. Structures of neutral glycans isolated from the lipopolysaccharides of reference strains for *Serratia marcescens* serogroups O16 and O20. *Carbohydr. Res.* **193**, 241–248 (1989).
38. Black, I. M. et al. Structure of Lipopolysaccharide from *Liberibacter crescens* is low molecular weight and offers insight into *Candidatus* Liberibacter biology. *Int. J. Mol. Sci.* **22**, 11240 (2021).
39. Altman, E., Chandan, V., Li, J. & Vinogradov, E. A reinvestigation of the lipopolysaccharide structure of *Helicobacter pylori* strain Sydney (SS1). *FEBS J.* **278**, 3484–3493 (2011).
40. Kido, N. et al. Expression of the O9 polysaccharide of *Escherichia coli*: Sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J. Bacteriol.* **177**, 2178–2187 (1995).
41. Cota, I., Blanc-Potard, A. B. & Casadesús, J. *STM2209-STM2208* (*opvAB*): A phase variation locus of *Salmonella enterica* involved in control of O-antigen chain length. *PLoS One* **7**, e36863 (2012).
42. Broadbent, S. E., Davies, M. R. & Van Der Woude, M. W. Phase variation controls expression of *Salmonella* lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. *Mol. Microbiol* **77**, 337–353 (2010).
43. Krinos, C. M. et al. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414**, 555–558 (2001).
44. Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc. Natl. Acad. Sci. USA* **100**, 10446–10451 (2003).
45. Canals, R. I. et al. The fitness landscape of the African *Salmonella Typhimurium* ST313 strain D23580 reveals unique properties of the pBT1 plasmid. *PLoS Pathog.* **15**, e1007948 (2019).
46. Barquist, L. et al. A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res.* **41**, 4549–4564 (2017).
47. Mandal, R. K. & Kwon, Y. M. Global screening of *Salmonella enterica* serovar Typhimurium genes for desiccation survival. *Front Microbiol.* **8**, 278455 (2017).
48. Navarre, W. W. et al. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science (1979)* **313**, 236–238 (2006).
49. Kimura, S., Hubbard, T. P., Davis, B. M. & Waldor, M. K. The nucleoid binding protein H-NS biases genome-wide transposon insertion landscapes. *mBio* **7**, 10–1128 (2016).
50. Eriksson, S., Lucchini, S., Thompson, A., Rhen, M. & Hinton, J. C. D. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol Microbiol* **47**, 103–118 (2003).

51. Noster, J. et al. Proteomics of intracellular *Salmonella enterica* reveals roles of *Salmonella* pathogenicity island 2 in metabolism and antioxidant defense. *PLoS Pathog* **15**, e1007741 (2019).
52. Kröger, C. et al. An infection-relevant transcriptomic compendium for *Salmonella enterica* serovar Typhimurium. *Cell Host Microbe* **14**, 683–695 (2013).
53. Srikumar, S. et al. RNA-seq brings new insights to the intra-macrophage transcriptome of *Salmonella Typhimurium*. *PLoS Pathog.* **11**, e1005262 (2015).
54. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
55. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
56. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
57. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
58. Dereeper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
59. Westpahl, O. Bacterial lipopolysaccharide-extraction with phenol water and further application of procedure. *Methods Carbohydr. Chem.* **1**, 83–91 (1965).
60. Hitchcock, P. J. & Brown, T. M. Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. *J. Bacteriol.* **154**, 269–277 (1983).

Acknowledgements

The authors would like to thank Olga G. Ovchinnikova for assistance in solving the T1 antigen structure. We would like to give thanks to the support provided by the Advanced Analysis Center at the University of Guelph (NMR facility, mass spectrometry facility and genomics facility). We also thank the NMR facilities at Academia Sinica (Institute of Biological Chemistry, Institute of Biomedical Sciences and National Biotechnology Research Park), the Synthetic Core Facility in the Institute of Biological Chemistry for optical rotation data and the Mass Spectrometry facility in the Institute of Chemistry for mass spectrometric data to support the synthetic work described in the Supporting Information. This work was supported by funding from the Natural Sciences and Engineering Research Council (RGPIN-2020-03886) to C.W. and Academia Sinica (AS-IA-113-03) to T.L.; C.W. held a Canada Research Chair and S.D.K. received an NSERC Postgraduate Scholarship (PGS-D).

Author contributions

S.D.K. performed all of the bioinformatic and biochemical experiments. M.J.A. synthesized the synthetic acceptors under the supervision of T.L.L. C.W. and S.D.K. conceived the project and C.W oversaw the experimental work performed by S.D.K.; L.D.G. provided SalFoS dataset access. S.D.K., M.J.A., T.L.L. and C.W. prepared the initial draft of the paper and all authors made contributions to the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-50957-y.

Correspondence and requests for materials should be addressed to Todd L. Lowary or Chris Whitfield.

Peer review information *Nature Communications* thanks Zachary Dalebroux, Chris Bayliss, who co-reviewed with Anisha Thanki, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.