

Peer Review File

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



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reviewer #1 (Remarks to the Author):

The manuscript by Kelly et al. reports an intriguing and very complete investigation into the structure and biosynthesis of an elusive bacterial polysaccharide. As a given, such bacterial O-antigen polysaccharides are highly complex and therefore difficult to study. In this manuscript, several approaches and techniques are combined to prove its structure and biosynthesis machinery for the first time. I highly recommend publication, but I have points that can be clarified, and some suggestions to improve readability, especially for the wider audience of Nat Comm:

p.3 line 1 67 known OPS, does this mean identified, or also structurally characterized?

p.4 line 16 why did the authors select the Ribf-GT from Klebsiella? Are there any other strains that could have been chosen, how widely annotated are Ribf-GTs?

p.4 line 18 and 21 does the candidate Ribf-transferase gene have a name? 'encoded by a gene', it would be good to add any gene descriptors here, so increase the ease of reproduction.

p.4 line 22-23 'the locus', can you add more details? How many orfs, is it similar to the 11-orf locus that was selected in line 25?

p.6 line 24-28 'STM0719 gene product is...' this suggests it is a known fact, but instead, this is a result from the current study, isn't it? Same with the statement on STM0720 and STM0721, which are presented as established facts, while their homology would predict they are, right? Please make this more explicit.

p.7 line 8-9 the structure of the 'WecA-WbbN-WbbO product acceptor' is not clear. I would substitute this for the Und-PP-GlcNAc-Galf-Galp designation.

p.7 lines 19-21 this section on the polymerase is very dense in information, and the results that lead to the conclusion are difficult to follow. It would be good if the authors attempt to make a better distinction between known genes with known functions, and predicted genes with predicted functions, and remind the readers what results were anticipated with the antiserum.

p.7 line 30 mixed Galf-PS produced in vivo, was this batch also isolated/purified?

p.8 line 11 'only identifiable Ribf-transferase, I assume this was identified by gene homology? Please make this clear

p.9 line 6-7 'Galf and Ribf were the only sugars found...' add the reference here again for this statement.

p.9line 29 the authors conclude that het Galf residues are linked at the C5 of the non-reducing end of a disaccharide acceptor. Would it still be possible that, if a larger acceptor substrate was used, there also would be branching at C5 position of units in the chain?

p.11 line20 'all the required T1 biosynthetic genes', can you name them briefly here again, to also enforce the idea that they are all required?

p.13 line 3 add the bacterial strain where FimB is found.

In Figure 1 the glycans are not drawn in the regular right-to-left direction, but reversed. This is highly confusing, and I would strongly advise that the authors remake this figure to have the glycans in the same orientation as the glycans in Fig 2 and Fig 4. Also, at the arrow of WbbO/N the appropriate sugar-nucleotides should be added, if possible

p.9 line 20 should this reference to Figure 4c?

p.10 line 14 should this reference to Fig 4b?

Reviewer #2 (Remarks to the Author):

The manuscript "Structure, biosynthesis and regulation of the T1 antigen, a phase-variable surface polysaccharide conserved in many *Salmonella* serovars" aims at establishing the detailed structure, biosynthesis of T1 antigen, which is also here shown to be a phase variable antigen. The paper is technically sound, quite interdisciplinary, the investigation covers a range of established techniques and approaches, from isolation and characterization of T1 antigen to synthesis of acceptors, from bioinformatics to biochemistry and modern structural biology approaches, that the authors use in an appropriate and high level fashion.

The work is, in my opinion, comprehensive and of high quality, final results represent a detailed description of the structure and biosynthesis of T1 antigen. The manuscript is generally clear and well written, conclusions well consistent with the data presented. Besides, structure/function/phase variation was thoroughly evaluated in *E. coli*, so the authors should comment on this, and about the reason *Salmonella* was not used and further comment on what is new after having the information of *Salmonella* T1 O-antigen.

One fundamental general remark: Despite the interesting data and results, what is the reason of the low abundance of the T1 antigen, how about its physiological relevance and role in virulence?

Some questions and comment on figures and text:

Figure 1, the western blot and the SDS-PAGE should be presented side by side to better appreciate the bands of T1 antigen.

The authors found and demonstrated a linkage between the Galf and the Ribf domains; in this regards, have the authors attempted a HR MS approach, to further demonstrate the connection between the two domains? Any information regarding the (average) molecular weight of the T1 antigen?

Figure 1 and 2, please show the NMR spectrum with signals in scale; there are some, i.e., the anomeric proton of D1, out of scale; have the authors attempted an integration of the signals to get indication of the relative abundance of the glycan chains? I think this is an important information.

Pag. 9, lines 23-29, please show the NMR spectra with the linkage between the PS domains. The two Galf-PS and Ribf-PS should be connected via position 5 of Galf, and evidence from compositional, methylation analysis?

The quality of the 2D NMR spectra should be improved.

Pag. 20 please specify the yield of T1 OPS;

pag. 24 please detail the acquisition conditions of the NMR spectra. In addition, were NOESY spectra acquired?

Reviewer #3 (Remarks to the Author):

See the attached document.

Reviewer #3 Attachment on the following page

Comments to the authors

In this research article by *Kelly et. al* from the groups of Dr. Lowary and Whitfield, the authors conduct a chemical, biochemical, molecular, and genetic characterization, and assessment of T1 antigen biosynthesis natively produce by various *Salmonellae*. Remarkably, the exact structure of this unique repeating surface polysaccharide and the genes and enzymes involved in its biogenesis have eluded discovery. The authors provide us with these important facts through a comprehensive and highly rigorous series of analyses and experiments. The quality of the studies is excellent, and the potential impact is broad since the structure of the T1-antigen is unique and present in many highly pathogenic serovars, but suspiciously absent or mutated in others for which there is no biological explanation given.

Intriguingly, the authors use genomics and informatics to show that the T1 antigen genes are more frequently present in broad host-range serovars than human-restricted serovars. Further, in the serovars that carry the V1 antigen, the genes are differentially expressed (on versus off) by recombinatorial inversion, which suggests that T1 is produced and not produced under specific conditions. Additionally, the researchers show that the T1 antigen is covalently attached to a lipid A-core molecule by the WaaL ligase. As we know, lipid A-core provides the amphipathic anchor for O-polysaccharides or O-antigens of hallmark LPS molecules. Therefore, this work opens new doors of investigation for how the pathogens might substitute O for T1 antigens in specific environments. The involvement of *Salmonellae* glycolipids and capsular polysaccharides in disease pathogenesis is well understood and broadly appreciated so this work is impactful to a great audience in biology and medicine.

Although the work is rigorous, impactful, and significant there are some major and mostly minor weaknesses. For instance, the study is conducted heterologously in *E. coli* K12, a host microbe that doesn't carry O or T1 antigens. While this approach does have significant advantages, I am mildly confident this is not impacting the biochemistry, but without the controls like the biochemical characterization of T1 from Typhimurium LT2 or the Paratyphi B isolate, it is not certain.

In this same vein, it was confusing as to why the authors did not pursue the investigation in the LT2 or *S. Paratyphi* B for which the T1 antigen was originally discovered and previously chemically characterized, in part, by groups like Nikaido's. Perhaps on revision some mention or inclusion of this rationale could be included.

Its discovery as a target of antibodies indicates it is expressed in the mammalian host environment. Have the authors tested site-directed mutants in the genes encoding the enzymes in *S. Typhimurium* LT2 or another Typhimurium isolate that carries and express T1 for phenotypes in a mouse model? Some demonstration of biological significance or function would improve this already well-conducted investigation.

Major Comments

In general, the rigor of the structural and molecular genetic characterization is strong, and the article is well organized and experimentally conducted with a high level of expertise and attention to detail. The writing of the results section could be improved for clarity and readability.

1. Although the stated goal is to provide a biological platform for future investigation, some demonstration of biological function would go a long way increasing the significance of this study. I offered some suggestions in my narrative above. Without a function or demonstration of its presence and absence in *S. enterica* under certain conditions its hard to comprehend the exact relevance

2. It should be mentioned why the studies were not conducted in *S. Typhimurium* LT2 or *S. Paratyphi* B. It would be nice to provide a chemical characterization of the T1 antigen in a *S. enterica* genotype that has lost the ability to construct the O-antigen. This might perhaps inform unique structural features that are not produced in the heterologous host used here. In other words, are the authors certain the T1 that is produced in *E. coli* heterologously is identical to that produced by *S. enterica*? This suggestion is for providing further rigor to an already intensely rigorous study.

Minor Comments

In general, the results section is riddled with abbreviations and structural reporting that could be lightened with some wordsmithing for readability. The line numbers did not translate properly on my end so I am unable to provide these.

Would it be possible to include the predicted structure of one or two characteristic or abundant T1 antigen containing lipid A core structures. Like the full structure as you would for lipid A-core plus O-antigen? Do you understand exactly how the T1 antigen is connected? Is it identical to how O-antigen is attached?

1. Please clarify when you say, "*lost in Enteritidis*", are you meaning all species within this serovar? Is there a remnant? Is it on a mobile element? In other words, is there a genetic explanation for its absence?

2. "*The locus from the well-studied S. enterica serovar Typhimurium LT2 (strain SL3770) was used as a prototype for subsequent investigation.*"

Here is an opportunity to discuss why the parental organism (Typhimurium LT2) for which the DNA was provided was not used in the study, nor was the Paratyphi B isolate?

3. Please clarify, "*To determine the structure of the T1 antigen, it was purified from LPS produced in E. coli K-12*"

What do you mean it was purified from LPS, exactly? How does one purify a glycolipid from a glycolipid? This does not seem trivial.

4. In the discussion, "*Part of this is answered by the competition between the T1 antigen and the serotype-specific OPS for lipid A-core.*"

There is no reference to a figure or a citation. I do not understand how your data speaks to the competition between these two antigens. Have you done experiments in *S. enterica* to prove this, exactly?

Reviewer #4 (Remarks to the Author):

Joint review provided by Prof Chris Bayliss and Dr. Anisha Thanki (ECR)

This manuscript by Kelly et al. contains a comprehensive analysis of a novel phase-variable LPS-linked glycan that is widely distributed among *Salmonella* serovars. The authors provide an in depth, highly detailed analysis of both the structure of this glycan and of the genes that are responsible for specific steps of the biosynthesis and regulation.

The authors have determined the structure, biosynthesis and regulation of the T1 antigen, which for the first time the authors show is present in the genomes of multiple *Salmonella* serovars. The publication is in scope for *Nature* as the function of the T1 antigen has not been determined before and therefore is novel. The data presented is of high quality, written exceptionally well and the data will be of significance to the field.

What are the noteworthy results?

The results are noteworthy and the authors have filled an important knowledge gap on the biological function of T1 antigen. The authors took logical steps to identify the structure, biosynthesis and regulation of the T1 antigen and for the first time identified that the T1 antigen is widespread in *Salmonella* serotypes/strains. The results will be of relevance to others in the field. Key noteworthy results are that they determined the structure of the T1 antigen and identified the composition as consisting of two homopolymeric structural components. The authors identified six genes in the *Salmonella* locus (STM0719–STM0721 and STM0724–STM0726) and show that the enzymes encoded in these genes are required for specific steps in the biosynthesis of the T1 antigen. As mentioned above the authors have also shown that almost all *S. enterica* Typhimurium isolates (299/318) had complete T1 clusters. Lastly, they identified that the T1 antigen expression is regulated by a recombinational phase switch and OFF state was approximately 2–3 times more abundant than the ON state.

Many of the T1 glycan structural details and the biosynthetic genes have not been previously observed or are distant homologs of genes found in other organisms. As such, these analyses are highly novel and constituent a major advance in our understanding of glycan structures and biosynthetic operons in Gram -ve bacterial species. The authors were not solely satisfied with identifying this glycan but also addressed the reason for limited knowledge of this glycan despite a few decades of research on glycans and LPS structure in *Salmonella*. They elegantly demonstrate that this locus contains a putative phase-variable switch and then further demonstrate that this switch is indeed subject to phase variation. The phase variation of this structure is a partial explanation of the lack of previous detection of this glycan although there is a mystery as to why the presence of a FimB recombinase was not noted and explored previously. The authors note that there is further potential for phase variation of other *Salmonella* glycans. The wide distribution and novelty of this locus has enriched our knowledge of *Salmonella* LPS structures and provides a strong foundation for exploration of the biological functions of this novel glycan, for detection and characterisation of similar glycans in other *Salmonella* serovars (and indeed other Gram -ve bacterial species) and for phase-variable alterations in this and other LPS glycan structures in *Salmonella*.

Does the work support the conclusions and claims, or is additional evidence needed?

The work supports the conclusions and claims, additional evidence is not needed. What I do strongly suggest to the authors is a final figure showing all results (all genes and their identified functions) would be helpful in completing the story as there is a significant amount of information to adsorb.

By the end of manuscript it is clear that many *Salmonella*'s can produce a T1 antigen in addition to an O-antigen, ECA and exopolysaccharide and T1 antigen. It might be useful to have a diagram showing the relative sizes abundance and locations of these different glycans as visual summary of the findings of this work. Possibly, this figure could indicate switching between isolates with differing proportions of the T1 antigen.

Are there any flaws in the data analysis, interpretation and conclusions? - Do these prohibit

publication or require revision?

No

Is the methodology sound? Does the work meet the expected standards in your field?

Yes the methodology is sound and logically organised.

Is there enough detail provided in the methods for the work to be reproduced?

Almost. Page 19. Line 20. The methods do not indicate how the bacteria were grown prior to purification of the LPS. The authors should provide some details of the growth conditions for the 10L cultures.

Minor changes

The authors have referenced the Casadeus article on phase variation of chain length in Salmonella but should probably also refer to evidence of phase variation of other modifications of O antigen structures (e.g. Broadbent et al. PMID:20487280) to give a more rounded view of the field.

The manuscript reports several homology searches with the text providing % identity/homology scores. In some cases, these statements are supported by alignments that can be interrogated by interested readers to assess the nature of these matches. In other cases, these values are not supported and it is unclear if the scores represent conserved domains, regions of genes or full coverage of the gene sequence. This should be rectified by providing coverage data for unsupported data or providing the alignments in Supplementary Figures. One example is Page 4/Line 20 where identity scores are provided for the Ribf-transferase.

Page 4. Line 20. Would be useful to indicate which gene in Figure 1a encodes the Ribf transferase homolog to help orientate the reader.

Page 4. Line 25. The authors report cloning the whole 11 gene locus into pACYC184 and then imply that expression is now under control of the tetR promoter. However, the reader will immediately note that the 11 genes do not appear to be in a single operon and that part of this region contains one or more promoters leading to some potential confusion. The authors should clarify where the tetR promoter is located relative to the 11 genes (presumably upstream of STM0716) and state that expression could be driven by the tetR promoter or other native promoters contained within this region.

Minor textual changes

Page 4. Line 10. 'enzymes involved IN polysaccharide'

Page 8. Line 28. Should be Figure 4a not 3a. (may need to swap order of Figures 3 and 4 to reflect order in the text).

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

RESPONSE TO REVIEWERS

Reviewer #1 (Remarks to the Author):

The manuscript by Kelly et al. reports an intriguing and very complete investigation into the structure and biosynthesis of an elusive bacterial polysaccharide. As a given, such bacterial O-antigen polysaccharides are highly complex and therefore difficult to study. In this manuscript, several approaches and techniques are combined to prove its structure and biosynthesis machinery for the first time. I highly recommend publication, but I have points that can be clarified, and some suggestions to improve readability, especially for the wider audience of Nat Comm:

Thank you.

p.3 line 1 67 known OPS, does this mean identified, or also structurally characterized?

The number was actually misstated. This sentence has been updated and clarified to indicate that it refers to structurally characterized OPS serogroups. We also added an additional reference that provides broader insight and hints at further diversity.

p.4 line 16 why did the authors select the Ribf-GT from *Klebsiella*? Are there any other strains that could have been chosen, how widely annotated are Ribf-GTs?

This is the prototype Ribf-GT; when we began this work, there were only two biochemically defined Ribf-GTs of this type. Ribf-GTs were not annotated by BLAST because they represent a relatively new glycosyltransferase fold (see Kelly et al 2022). The discovery of homologs (including the *Salmonella* one) only came after we identified and assigned the *Klebsiella* enzyme. Wording was added in the opening section of results to make this clearer.

p.4 line 18 and 21 does the candidate Ribf-transferase gene have a name? 'encoded by a gene', it would be good to add any gene descriptors here, so increase the ease of reproduction.

There is currently no gene name for the gene, and we have opted to use the locus tags from LT2. We have included gene names where there is unequivocal evidence that a homolog has the same function as an established gene product (eg WbbN, Wbbo, Glf in this study). However, we are avoiding assigning new gene names going forward because there are too many bacterial GTs to accommodate unique names using conventional genetic nomenclature. The locus tag avoids conflicts in cases where a gene name is/has been applied in another system without recognition of the conflict; this has been a problem in this field in the past.

p.4 line 22-23 'the locus', can you add more details? How many orfs, is it similar to the 11-orf locus that was selected in line 25?

Yes, it has 11 orfs. This has been clarified in the text.

p.6 line 24-28 'STM0719 gene product is...' this suggests it is a known fact, but instead, this is a result from the current study, isn't it? Same with the statement on STM0720

and STM0721, which are presented as established facts, while their homology would predict they are, right? Please make this more explicit.

The reviewer is correct, it is predicted at this stage; the text has been changed to acknowledge this.

p.7 line 8-9 the structure of the 'WecA-WbbN-WbbO product acceptor' is not clear. I would substitute this for the Und-PP-GlcNAc-Galf-Galp designation.

Thank you, this has been changed.

p.7 lines 19-21 this section on the polymerase is very dense in information, and the results that lead to the conclusion are difficult to follow. It would be good if the authors attempt to make a better distinction between known genes with known functions, and predicted genes with predicted functions, and remind the readers what results were anticipated with the antiserum.

We reorganized the paragraph for better logical flow and have explained the interpretation more carefully.

p.7 line 30 mixed Galf-PS produced in vivo, was this batch also isolated/purified?

The Galf-PS was not independently isolated from the Ribf-PS; this would be a challenging purification and was not necessary for interpretation of the assembly system. We changed the wording slightly to be clear we are comparing a portion of the *in vivo* structure and removed the word "mixed" because it is confusing in this context.

p.8 line 11 'only identifiable Ribf-transferase, I assume this was identified by gene homology? Please make this clear'

Yes. This has been clarified.

p.9 line 6-7 'Galf and Ribf were the only sugars found...' add the reference here again for this statement.

While the original T1 paper identified the monosaccharide constituents, our results shows the same thing and goes further than that work by offering a definitive T1 structure. We clarified this with an "(as described above)".

p.9line 29 the authors conclude that the Galf residues are linked at the C5 of the non-reducing end of a disaccharide acceptor. Would it still be possible that, if a larger acceptor substrate was used, there also would be branching at C5 position of units in the chain?

We considered the possibility that branching could occur, but this was ruled out by the absence of any additional minor Galf signals in the T1 polysaccharide from the *E. coli* transformant. We have not raised branching here because it may be unnecessarily confusing for general readers.

p.11 line20 'all the required T1 biosynthetic genes', can you name them briefly here again, to also enforce the idea that they are all required?

We did not name each one but have revised the text to describe the exact block of genes being discussed.

p.13 line 3 add the bacterial strain where FimB is found.

The sources are now provided in the legend of Figure 5 and *E. coli* MG1655 is included in the text.

In Figure 1 the glycans are not drawn in the regular right-to-left direction, but reversed. This is highly confusing, and I would strongly advise that the authors remake this figure to have the glycans in the same orientation as the glycans in Fig 2 and Fig 4. Also, at the arrow of WbbO/N the appropriate sugar-nucleotides should be added, if possible

The depictions of the glycans have been reversed as requested. The panel of Fig 2 that is referred to here is now part of a new Fig. 1. In this panel (1B), sugar donors have been added for WbbN and WbbO and we have revised the pathway, so that the two steps are no longer combined; this makes the substrate for each clear.

p.9 line 20 should this reference to Figure 4c?

Corrected, thank you

p.10 line 14 should this reference to Fig 4b?

Corrected, thank you

Reviewer #2 (Remarks to the Author):

The manuscript "Structure, biosynthesis and regulation of the T1 antigen, a phase-variable surface polysaccharide conserved in many *Salmonella* serovars" aims at establishing the detailed structure, biosynthesis of T1 antigen, which is also here shown to be a phase variable antigen. The paper is technically sound, quite interdisciplinary, the investigation covers a range of established techniques and approaches, from isolation and characterization of T1 antigen to synthesis of acceptors, from bioinformatics to biochemistry and modern structural biology approaches, that the authors use in an appropriate and high level fashion.

The work is, in my opinion, comprehensive and of high quality, final results represent a detailed description of the structure and biosynthesis of T1 antigen. The manuscript is generally clear and well written, conclusions well consistent with the data presented. Besides, structure/function/phase variation was thoroughly evaluated in *E. coli*, so the authors should comment on this, and about the reason *Salmonella* was not used and further comment on what is new after having the information of *Salmonella* T1 O-antigen.

Thank you. When this work began, the transient nature of T1 production meant we were unable to advance the investigation in *Salmonella* itself. The T1-positive isolate from the original study could not be sourced, so we chose to move the locus into an *E. coli* K-12 background to remove these complications. This rationale has been added to the text beginning on p4. We are unsure how to respond to the final comment. We believe the advances and implications from this work for *Salmonella* are fully elaborated in the discussion.

One fundamental general remark: Despite the interesting data and results, what is the reason of the low abundance of the T1 antigen, how about its physiological relevance and role in virulence?

The low abundance of T1 in an isolate is likely due to a combination of the phase variation (established here in lab grown bacteria), and potential competition for undecaprenyl lipid carrier and LPS ligation sites/machinery. We have noted this in the discussion.

We cannot speculate on a role in virulence with any confidence. First, T1 antigen is not found in all *Salmonellae* and the reason for this distribution is not illuminated by our current data, or by any information in the literature. We do not know whether T1 is produced in the host, or if it is a component needed for survival outside the host. We added a sentence about this at the end of the discussion. The physiological role is an important and challenging question requiring an extensive additional investigation that is beyond the scope of our study. However, it is critical to note that the question could not be asked before the foundation provided by the current study.

Some questions and comment on figures and text:

Figure 1, the western blot and the SDS-PAGE should be presented side by side to better appreciate the bands of T1 antigen.

Understood. The blots and silver stain are now presented side-by-side (now in Fig 2 in the revision).

The authors found and demonstrated a linkage between the *Galf* and the *Ribf* domains; in this regards, have the authors attempted a HR MS approach, to further demonstrate the connection between the two domains?

HD-MS was not used. We consider NMR spectroscopy to be appropriate and sufficient, because all of the data on the LPS product was confirmed by more focused analysis of the critical assembly intermediates generate *in vitro* using synthetic acceptor substrates and defined enzymes, where analysis is more straightforward and unequivocal.

Any information regarding the (average) molecular weight of the T1 antigen?

The average molecular weight was not determined, but T1-LPS migration in SDS-PAGE suggests it is of comparable size to the major cluster in LPSs modified by other O-antigens. Because the migration in SDS-PAGE can be influenced by composition as well as chain length, we cannot comment with any certainty. We prefer to leave it to readers to make assumptions if they choose.

Figure 1 and 2, please show the NMR spectrum with signals in scale; there are some, i.e., the anomeric proton of D1, out of scale; have the authors attempted an integration of the signals to get indication of the relative abundance of the glycan chains? I think this is an important information.

Thank you. The scale has been adjusted. The integration of the *Ribf* anomeric peak compared to the combined anomeric peaks from the *Galf*-PS results in ~2:1 *Ribf* to *Galf*. This has now been added to the results section on p6.

Pag. 9, lines 23-29, please show the NMR spectra with the linkage between the PS

domains. The two Galf-PS and Ribf-PS should be connected via position 5 of Galf, and evidence from compositional, methylation analysis?

We have provided this region of the spectrum from the *in vivo* produced polysaccharide as a new SI figure. Space limitations precludes adding it to the main text.

Methylation analysis was performed on the original T1 polysaccharide and NMR data here were in agreement with the published work to the extent that it exists. We have not performed methylation analysis ourselves.

The NMR spectra of the polysaccharide did not fully resolve the linkage due to its complexity. That is why the linkage between the two domains was established in a defined product synthesized *in vitro* (the spectrum of this is in Fig 4). This was then used to look for any inconsistencies in the natural product spectra; none were found. The synthetic Galf-acceptors are defined mimics of the authentic product. As explained in the text, the key correlation between the Ribf C-1 and Galf H-5 in the HMBC spectrum allows us to be confident in our assignment. We believe the *in vitro* data gives an unequivocal conclusion. We apologize if we are missing the reviewer's intent here.

The quality of the 2D NMR spectra should be improved.

We are not clear what the reviewer is looking for here. Some of the overlapping peaks of the Galf-PS do contribute to less symmetrical signals, but the conclusions drawn from the data are not compromised by this.

Pag. 20 please specify the yield of T1 OPS;

Added as requested.

pag. 24 please detail the acquisition conditions of the NMR spectra. In addition, were NOESY spectra acquired?

The acquisition temperature has been added as requested. NOESY was acquired for the *in vivo* synthesized product, but the data was not used in the initial analysis. We have included it in the revision (in main text and SI figs) to provide additional support.

Reviewer #3 (Remarks to the Author):

Comments to the authors

In this research article by Kelly et. Al from the groups of Dr. Lowary and Whitfield, the authors conduct a chemical, biochemical, molecular, and genetic characterization, and assessment of T1 antigen biosynthesis natively produce by various *Salmonellae*. Remarkably, the exact structure of this unique repeating surface polysaccharide and the genes and enzymes involved in its biogenesis have eluded discovery. The authors provide us with these important facts through a comprehensive and highly rigorous series of analyses and experiments. The quality of the studies is excellent, and the potential impact is broad since the structure of the T1-antigen is unique and present in many highly pathogenic serovars, but suspiciously absent or mutated in others for which there is no biological explanation given.

Intriguingly, the authors use genomics and informatics to show that the T1 antigen genes are more frequently present in broad host-range serovars than human-restricted serovars. Further, in the serovars that carry the V1 antigen, the genes are differentially expressed (on versus off) by recombinatorial inversion, which suggests that T1 is produced and not produced under specific conditions. Additionally, the researchers show that the T1 antigen is covalently attached to a lipid A-core molecule by the WaaL ligase. As we know, lipid A-core provides the amphipathic anchor for O-polysaccharides or O-antigens of 6allmark LPS molecules. Therefore, this work opens new doors of investigation for how the pathogens might substitute O for T1 antigens in specific environments. The involvement of *Salmonellae* glycolipids and capsular polysaccharides in disease pathogenesis is well understood and broadly appreciated so this work is impactful to a great audience in biology and medicine.

Thank you.

Although the work is rigorous, impactful, and significant there are some major and mostly minor weaknesses. For instance, the study is conducted heterologously in *E. coli* K12, a host microbe that doesn't carry O or T1 antigens. While this approach does have significant advantages, I am mildly confident this is not impacting the biochemistry, but without the controls like the biochemical characterization of T1 from Typhimurium LT2 or the Paratyphi B isolate, it is not certain.

We added a comment on the rationale for doing the work in *E. coli* (see response to reviewer 2). At the beginning of this study, we had no *Salmonella* strains in the on-state, which prevented us from proceeding with that system. We would argue that there is no reason to expect that the biochemistry would be affected by changing in organism. We are simply expressing the enzymes in *E. coli* rather than the closely-related *Salmonella*. For the structure of T1 to change, one would have to invoke the action of another component in *Salmonella*, or the presence of additional (different) components in *E. coli*, and there is nothing obvious in the genomes to suggest this. Importantly, the solved T1 structure produced in *E. coli* is entirely consistent with the partial structure reported in *Salmonella*. Had the organism played a role, we would have identified any differences. We also back this up with antibody detection in *Salmonella*. That is, polyclonal antisera raised using the product produced in *E. coli*, cross-reacts with the product formed in *Salmonella*. While not conclusive, these observations do not identify any major differences. We consider that redoing the entire study in an engineered (on-state) *Salmonella* lies beyond the scope of an already extensive study (and we note the guidance of the editor's summary stating that no new experimental work is necessary). A comment has been added to the discussion to indicate the caveats, and we hope this addresses the reviewer's concern by alerting readers to the limits.

In this same vein, it was confusing as to why the authors did not pursue the investigation in the LT2 or S. Paratyphi B for which the T1 antigen was originally discovered and previously chemically characterized, in part, by groups like Nikaido's. Perhaps on revision some mention or inclusion of this rationale could be included.

Understood. We tried to acquire this specific Paratyphi isolate but were not successful. As shown here, most *Salmonella* make no appreciable amounts of T1 under lab growth conditions, so it was not obvious how to access the structure starting with *Salmonella*.

This is now stated in the rationale on p5. Given what we found later with the switching mechanism, we now know why most isolates have little T1 (and why T1 does not appear in the many O antigen structural studies with *Salmonella*). We believe the original T1-producing isolate was a fortuitous mutant of some kind, so it is impossible to replicate without a genome sequence of that specific isolate.

Its discovery as a target of antibodies indicates it is expressed in the mammalian host environment. Have the authors tested site-directed mutants in the genes encoding the enzymes in *S. Typhimurium* LT2 or another *Typhimurium* isolate that carries and express T1 for phenotypes in a mouse model? Some demonstration of biological significance or function would improve this already well-conducted investigation.

The antibody production was done with a phase-locked situation in *E. coli*, and it cannot be implied as an indication of T1 expression in WT *Salmonella* in the host.

We understand the reviewer's perspective, but we are not in a position to run mouse models in our groups, and we are unsure whether a single mouse model would provide conclusive results. As now indicated in the revised discussion (response to reviewer 1), we do not know whether T1 is required for fitness inside or outside the host, so a negative outcome would not be useful in terms of physiological roles. We believe this question merits a systematic investigation and lies beyond the scope of this study.

Major Comments

In general, the rigor of the structural and molecular genetic characterization is strong, and the article is well organized and experimentally conducted with a high level of expertise and attention to detail. The writing of the results section could be improved for clarity and readability.

1. Although the stated goal is to provide a biological platform for future investigation, some demonstration of biological function would go a long way increasing the significance of this study. I offered some suggestions in my narrative above. Without a function or demonstration of its presence and absence in *S. enterica* under certain conditions its hard to comprehend the exact relevance.

See the response to the same point in the reviewer's general comments above.

2. It should be mentioned why the studies were not conducted in *S. Typhimurium* LT2 or *S. Paratyphi* B. It would be nice to provide a chemical characterization of the T1 antigen in a *S. enterica* genotype that has lost the ability to construct the O-antigen. This might perhaps inform unique structural features that are not produced in the heterologous host used here. In other words, are the authors certain the T1 that is produced in *E. coli* heterologously is identical to that produced by *S. enterica*? This suggestion is for providing further rigor to an already intensely rigorous study.

The rationale for the use of the recombinant *E. coli* system and this point has been addressed in response to the reviewer's general points above. The question the reviewer asks requires engineering an on-state *Salmonella* isolate – an experiment only possible after the molecular foundation for switching was established. We firmly believe

that the existing data makes the possibility of an altered structure remote but have acknowledged the limitation.

Minor Comments

In general, the results section is riddled with abbreviations and structural reporting that could be lightened with some wordsmithing for readability. The line numbers did not translate properly on my end so I am unable to provide these.

We agree that this is a challenge, but it is not clear how we can avoid this. We have deliberately offered more description in the text than we might for a specialist to make the intent explicit for a broader audience.

Would it be possible to include the predicted structure of one or two characteristic or abundant T1 antigen containing lipid A core structures. Like the full structure as you would for lipid A-core plus O-antigen? Do you understand exactly how the T1 antigen is connected? Is it identical to how O-antigen is attached?

The observation that the O-antigen ligase is required (see fig S1) strongly suggests the site of attachment is the same for O antigen and T1. We have added a figure to the SI, as requested to show the predicted structure and we explain the caveat about linkage in the legend.

1. Please clarify when you say, “lost in Enteritidis”, are you meaning all species within this serovar? Is there a remnant? Is it on a mobile element? In other words, is there a genetic explanation for its absence?

We apologize that the text was unclear. It has been updated; there are no remnants of the cluster at all in Enteritidis; we cannot discriminate between lost versus never there.

2. “The locus from the well-studied *S. enterica* serovar Typhimurium LT2 (strain SL3770) was used as a prototype for subsequent investigation.”

Here is an opportunity to discuss why the parental organism (Typhimurium LT2) for which the DNA was provided was not used in the study, nor was the Paratyphi B isolate?

We have provided a brief rationale for the use of the recombinant system.

3. Please clarify, “To determine the structure of the T1 antigen, it was purified from LPS produced in *E. coli* K-12”

What do you mean it was purified from LPS, exactly? How does one purify a glycolipid from a glycolipid? This does not seem trivial.

The wording has been changed to be clearer.

4. In the discussion, “Part of this is answered by the competition between the T1 antigen and the serotype-specific OPS for lipid A–core.”

There is no reference to a figure or a citation. I do not understand how your data speaks to the competition between these two antigens. Have you done experiments in *S. enterica* to prove this, exactly?

This is implied by the shared use of the ligase enzyme in *Salmonella* and not from experimental data. The reviewer’s point is well taken, so the text has been changed to

acknowledge this is speculation, not fact.

Reviewer #4 (Remarks to the Author):

Joint review provided by Prof Chris Bayliss and Dr. Anisha Thanki (ECR)

This manuscript by Kelly et al. contains a comprehensive analysis of a novel phase-variable LPS-linked glycan that is widely distributed among *Salmonella* serovars. The authors provide an in depth, highly detailed analysis of both the structure of this glycan and of the genes that are responsible for specific steps of the biosynthesis and regulation.

The authors have determined the structure, biosynthesis and regulation of the T1 antigen, which for the first time the authors show is present in the genomes of multiple *Salmonella* serovars. The publication is in scope for *Nature* as the function of the T1 antigen has not been determined before and therefore is novel. The data presented is of high quality, written exceptionally well and the data will be of significance to the field.

Thank you.

What are the noteworthy results?

The results are noteworthy and the authors have filled an important knowledge gap on the biological function of T1 antigen. The authors took logical steps to identify the structure, biosynthesis and regulation of the T1 antigen and for the first time identified that the T1 antigen is widespread in *Salmonella* serotypes/strains. The results will be of relevance to others in the field. Key noteworthy results are that they determined the structure of the T1 antigen and identified the composition as consisting of two homopolymeric structural components. The authors identified six genes in the *Salmonella* locus (STM0719–STM0721 and STM0724–STM0726) and show that the enzymes encoded in these genes are required for specific steps in the biosynthesis of the T1 antigen. As mentioned above the authors have also shown that almost all *S. enterica* Typhimurium isolates (299/318) had complete T1 clusters. Lastly, they identified that the T1 antigen expression is regulated by a recombinational phase switch and OFF state was approximately 2–3 times more abundant than the ON state.

Many of the T1 glycan structural details and the biosynthetic genes have not been previously observed or are distant homologs of genes found in other organisms. As such, these analyses are highly novel and constitute a major advance in our understanding of glycan structures and biosynthetic operons in Gram -ve bacterial species. The authors were not solely satisfied with identifying this glycan but also addressed the reason for limited knowledge of this glycan despite a few decades of research on glycans and LPS structure in *Salmonella*. They elegantly demonstrate that this locus contains a putative phase-variable switch and then further demonstrate that this switch is indeed subject to phase variation. The phase variation of this structure is a partial explanation of the lack of previous detection of this glycan although there is a mystery as to why the presence of a FimB recombinase was not noted and explored

previously. The authors note that there is further potential for phase variation of other *Salmonella* glycans. The wide distribution and novelty of this locus has enriched our knowledge of *Salmonella* LPS structures and provides a strong foundation for exploration of the biological functions of this novel glycan, for detection and characterisation of similar glycans in other *Salmonella* serovars (and indeed other Gram -ve bacterial species) and for phase-variable alterations in this and other LPS glycan structures in *Salmonella*.

Does the work support the conclusions and claims, or is additional evidence needed? The work supports the conclusions and claims, additional evidence is not needed. What I do strongly suggest to the authors is a final figure showing all results (all genes and their identified functions) would be helpful in completing the story as there is a significant amount of information to adsorb.

We tried to provide this in Fig 1a in the original submission. The activities of all the GTs (with their products) are shown in the schematic, but we recognize that this could be improved. In the revision, we moved this panel to a new Fig 1, together with an additional model for context. We added this up front to better guide the description in the Introduction and throughout the text. We believe adding another figure at the end instead would create redundancy and would also push us over the space limit.

By the end of manuscript it is clear that many *Salmonella*'s can produce a T1 antigen in addition to an O-antigen, ECA and exopolysaccharide and T1 antigen. It might be useful to have a diagram showing the relative sizes abundance and locations of these different glycans as visual summary of the findings of this work. Possibly, this figure could indicate switching between isolates with differing proportions of the T1 antigen.

We understand the point being made by the reviewers, but this would require critical assumptions for which there is no experimental foundation (see reviewer 3's concern). For example, the amount of ECA-LPS depends on the isolate and is complicated by there being 3 different forms of ECA; only one is linked to LPS and the relative abundance of the forms is not known. We do not have such data for T1-producers. Respectfully, we argue that a figure like the one requested here poses the risk of the speculation being adopted as fact. We assume the molecules are distributed across the cell surface (typical of LPS) but we cannot predict whether there are localized domains for a particular antigen (current technical limits make this question challenging). We have no data on the abundance of the various LPS-based glycans on which to base such a diagram.

Are there any flaws in the data analysis, interpretation and conclusions? - Do these prohibit publication or require revision?

No

Is the methodology sound? Does the work meet the expected standards in your field? Yes the methodology is sound and logically organised.

Is there enough detail provided in the methods for the work to be reproduced?

Almost. Page 19. Line 20. The methods do not indicate how the bacteria were grown

prior to purification of the LPS. The authors should provide some details of the growth conditions for the 10L cultures.

Thank you, this has been added.

Minor changes

The authors have referenced the Casadeus article on phase variation of chain length in Salmonella but should probably also refer to evidence of phase variation of other modifications of O antigen structures (e.g. Broadbent et al. PMID:20487280) to give a more rounded view of the field.

This has been added.

The manuscript reports several homology searches with the text providing % identity/homology scores. In some cases, these statements are supported by alignments that can be interrogated by interested readers to assess the nature of these matches. In other cases, these values are not supported and it is unclear if the scores represent conserved domains, regions of genes or full coverage of the gene sequence. This should be rectified by providing coverage data for unsupported data or providing the alignments in Supplementary Figures. One example is Page 4/Line 20 where identity scores are provided for the Ribf-transferase.

We have now added coverage percentages.

Page 4. Line 20. Would be useful to indicate which gene in Figure 1a encodes the Ribf transferase homolog to help orientate the reader.

Good point. We have 'bolded' STM0726 in Fig 1 and addressed this in the figure legend

Page 4. Line 25. The authors report cloning the whole 11 gene locus into pACYC184 and then imply that expression is now under control of the tetR promoter. However, the reader will immediately note that the 11 genes do not appear to be in a single operon and that part of this region contains one or more promoters leading to some potential confusion. The authors should clarify where the tetR promoter is located relative to the 11 genes (presumably upstream of STM0716) and state that expression could be driven by the tetR promoter or other native promoters contained within this region.

The reviewer is correct. As described in the paper, we go on to show that there really is a promoter in this region that is active independent of the plasmid promoter. The text has been updated.

Minor textual changes

Page 4. Line 10. 'enzymes involved IN polysaccharide'

Corrected.

Page 8. Line 28. Should be Figure 4a not 3a. (may need to swap order of Figures 3 and 4 to reflect order in the text).

Corrected.

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed my comments, in my opinion the manuscript can be published.

Reviewer #3 (Remarks to the Author):

The authors have fully satisfied the minor and major concerns largely by revising the text and providing further detail, context, and explanation.

This work is broadly impactful on a biological level. The findings expand our understanding of how enterobacterales, in addition to LPS, regulate production of a diversity of antigenic glycolipids upon their OM outer leaflet, such as the T1 antigen of *Salmonella enterica*. Elegantly, the researchers establish regulation occurs through site-specific recombination and gene expression changes resulting polysaccharide production and ligation to the lipid A core oligosaccharide. In this manner, LPS and T1 share the lipid A-core macroamphile for their tethering to the cell surface. This work has important implications for infection biology and immune evasion of salmonellae.

It was a pleasure to read and review.

Thank you