Peer Review File

Manuscript Title: Intragenic DNA inversions expand bacterial coding capacity

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

This manuscript reports intragenic DNA inversions as a new class of invertible elements found entirely within genes, which allow a gene to encode two or more versions of a protein by flipping a DNA sequence within the coding region, thereby increasing coding capacity without increasing genome size. The authors use a combination of bioinformatic and experimental approaches to identify and study intragenic invertons in a subset of bacterial species. A long-read based inverton predictor (PhaVa) has been developed and applied to discover intragenic invertons. This bioinformatic approach identified intergenic, partial and intragenic invertons across multiple phyla in bacteria and archaea. A specific intragenic inverton in BT0650 (thiC) from Bacteroides thetaiotaomicron was analyzed, and two "locked" versions of thiC invertons were generated to prevent inversion from occurring within the gene. The locked forward and locked reverse strains were grown in varying concentrations of thiamine to assess their growth rate, and competitive growth experiments were performed to determine the relative fitness of locked forward and locked reverse strains compared to the wildtype strain. The results showed that the locked reverse strain had a fitness advantage over the locked forward strain at physiologically relevant thiamine concentrations. This suggests that the intragenic inversion within thiC gene may provide a competitive advantage in certain conditions.

Major comments:

1. The existence of potential variable proteins should be demonstrated at the protein level. If the intragenic inversions that don’t introduce stop codons nevertheless lead to loss of protein stability, for example, no protein will be present. In this case, both inversions that introduce stop codons and inversions that don’t create stop codons would be predicted to phenocopy gene deletion. Protein level confirmation would also help address the possibility that the reported novel sequences could be artifacts resulting from PCR chimeras or other steps in the sequencing approaches.

2. Many genetic rearrangements, transposon insertions, and indels lead to premature stop codons or other loss-of-function consequences. The novelty of this report would be increased by demonstrating that inversions that do NOT introduce stop codons don't simply cause loss of protein stability. No examples are provided of intragenic inversions that actually have distinct biochemical functions and phenotypes compared to both the non-inverted sequence and the deletion mutant.

3. The authors state that stop codons introduced by intragenic inversion are reversible, while other nonsense mutations are not. This should be clarified – mobile genetic elements can move twice, nonsense mutations can revert back (e.g. PMID 37745437), under the appropriate selective pressure. No information quantifying the rates of reversion through these different mutational mechanisms is provided.

4. Line 155: The statement that intragenic inversions in an invertase possibly changes the invertible repeats that it targets for flipping or its affinity for these repeats would require experimental support.

5. Line 240 and Figure 3D identifies both the forward and reverse orientations within a given 'clonal' sample. If this is the same genotype found in laboratory-grown BTh type strain (Figure 4B), what is the proportion of these forward and reverse gene copies in the wildtype culture?

6. It is unclear how (or whether) the inversions are catalyzed. Are there dedicated invertases, and what are the fitness and biochemical consequences of deleting these enzymes? This is a major aspect of the proposed biological mechanism that is not addressed.

7. In Figure 4C, the wildtype strain and the locked forward strain maintained similar growth densities at different concentrations of thiamine, while the density of the locked reverse strains gradually increased as thiamine concentration rose from 0.1-10 μM. The expression levels of the three versions of thiC should be assessed under these different thiamine concentrations to determine whether the expression of each version of thiC correlates with thiamine concentration.

8. In metagenomes and the isolates of Bacteroidetes dataset you have analyzed (Figure S6), what is the frequency of occurrence of the invertons in thiC, and what is the phylogenetic distribution?

Minor comments:

9. For Figures 3A and B, a table showing the number of genomes in the different phyla being investigated and the predicted number of intergenic, intragenic, and partial intergenic invertons, would support the result. In the middle panel of Figure 3B, the genomes of Deinococcus-Thermus and Gemmatimonadetes have the highest average mean number of invertons: which genes carry invertons and what consequences the invertons produce would help provide more information from this figure panel.

10. In Figure 4A, the complete nucleotide and AA sequences of forward and reverse orientations would provide a useful supplemental figure.

Referee #2 (Remarks to the Author):

Drs. Chanin, West, Bhatt and colleagues have produced a remarkable study and should be commended for their careful yet highly novel work. While they build on the findings of others, their achievements in this study are considerable and include: identifying a new subset of invertons that are intragenic and can produce early stop codons or new open reading frames, developing a new tool and using it to quantify invertons using long-read sequencing to better identify and characterize invertons, finding that invertons are much more abundant in metagenomic datasets obtained from patient fecal samples when compared to in vitro cultivated isolates, and finally experimentally demonstrating that the state of a particular intragenic inverton can modify the competitive fitness of a strain of bacteria. This study is impactful and will be of interest to a broad audience. I have some comments and questions that I think could help improve what is already an impressive body of work.

Fig 1b: In clinical samples where BTh intragenic inversions are broadly observed, is there anything notable about these samples? Do they possibly show domination by BTh? Are they particularly low in bacterial density?

It seems that when the authors in Fig 1B were evaluating for predicted intragenic invertons using PCR, they were evaluating a polyclonal mixture in liquid media, and confirmed 10 predicted intragenic invertons. What would they observe if they screened individual clones? Is it possible to perform a kinetic analysis of single clones over time, evaluating the status of these intragenic invertons? Or do these 10 intragenic invertons flip so rapidly that by the time a single clone expands sufficiently enough to produce enough genetic material to allow evaluation, there is already achievement of a steady-state mixture?

Relatedly, would it be possible to characterize the stability of invertons after introduction of a strain to mice? Introducing their well-characterized BTh strain to a germ-free mouse to monocolonize, followed by quantification of inverton status, would be very interesting, especially for the 49 predicted intragenic invertons which could not be confirmed in vitro. The kinetics of status changes in these invertons in vivo would also be interesting.

Can the authors try to search for genetic features that distinguish invertons that flip in vitro (and presumably also flip in vivo) from invertons that only flip in vivo? Or do they think that these invertons all flip similarly, but some permissive conditions allow for production of heterogeneity in the state of some invertons, while other conditions strongly select for a particular state?

The experimental strategy to examine for the consequences of the thiC intragenic inversion is very nicely and carefully done. Using the wild-type strain, is it possible for the authors to demonstrate over time enrichment of the two possible states, when BTh is cultivated in high or low thiamine concentrations? Can they additionally generate an invertase mutant strain that in turn fails to adapt to ambient thiamine conditions?

In Figure 2B, the authors should convey the SD or SEM in their analysis.

Referee #3 (Remarks to the Author):

Bacteria are known to use DNA inversions to create heterogeneity and express different surface sugars – a process long known as phase variation. Using long-read sequencing of a longitudinal microbiome study, Chanin et al. report for the first time the existence of intragenic DNA inversions in bacteria. They then develop a new method to identify thousands of such inversions in >35,000 complete bacterial genomes (and a limited number of MAGs), and experimentally characterize one specific intragenic variation leading to premature termination of thiC and of thiamine biosynthesis in Bacteroides thetaiotaomicron.

This is an exciting study that opens a new field of research in bacteria, providing also the tools to enable further discoveries. However, there are several points that need to be fixed in the analysis and presentation of the results and new tool, so the utility of this resource fulfills its great potential.

Main points

1. The authors make a major claim about the recoding potential of intragenic inversions (leading to expression of new functional proteins) throughout the manuscript, including the abstract. This is indeed exciting, and a clear depart from intergenic (or partial intergenic) inversions. Yet only a (small) fraction of intragenic inversions can do this – those with recoding potential (in-frame inversion) or when the premature stop codon leads to a split protein. Unfortunately these more interesting cases are not presented or followed up to any depth in the manuscript – e.g. the main experimental follow up focuses to a case of a premature stop codon, which leads to more mundane case of an ON/OFF state of the gene. Although I recognize that asking for a new experimental follow up on a recoding case is a big ask at this point, there are other ways the authors could showcase better the recoding cases. First the authors should clearly distinguish and report the numbers of recoding vs premature intragenic cases throughout the different datasets they use (transplantation patient microbiomes, complete genomes and MAGs) – now all reported together. Ideally they should further differentiate the premature cases to those that potentially lead to a split protein vs those that only seem to prematurely stop the protein. Second, some functional analysis on the recoded (and split) domains is warranted and whether it happens at specific, conserved, catalytic, or protein-protein/-ligand binding sites. Also, there must be some examples (out of the thousand cases) where there is good structural and/or functional understanding of the protein and its binding partners, so the authors can use structural predictions to illustrate possible ramifications of the recoding.

2. To identify intragenic inversions, the authors initially use a longitudinal microbiome study of patients going transplantations. It is unclear why they focus on B. thetaiotaomicron. I would imagine that more abundant and prevalent species (from Bacteroidota phylum that are equally genetic tractable) would be better cases to look into. Looking into several closely related species would also enable some evolutionary comparisons (and gain more confidence for conserved predicted inversions). I would find useful if the authors expanded their efforts here – especially since they make a claim on the necessity of using bacteria in their natural settings for identifying intragenic

inversions.

3. Pha Va is a new computational tool presented in this study that allows for identification of (intragenic) inversions in long-read data. Although the tool looks specific, I would expect more rigorous benchmarking. Sensitivity/false negatives are not addressed. Results are compared to Jiang study but not to Milman one, which identifies many more insertions using more or less the same set of complete genome sequences. It's important that authors benchmark better their tool and compare it exhaustively with other state-of-the-art tools.

4. Authors use Pha Va in 210 MAGs. First this analysis is important and IMO should be part of the main Figs, but has to be done and presented better (present species targeted, differentiate between recoding/premature inversions, perform functional analysis). Second, it's unclear why only 210 MAGs, when there are orders of magnitude more good quality MAGs (with high completeness) available – what is the limitation? This is important, especially if intragenic inversion are more common in bacteria growing natural environments. Third, it is unclear if Pha Va can be used only with MAGs or can be also used directly with reads from metagenomics data – authors should explain.

Minor points

1. Abstract – it would benefit from being more precise on results of this study. Currently half of abstract is background, and the other half is devoid of specific results/analysis/experiments of this study, having mainly novelty claims. E.g. "thousands" in line 19 is not particularly precise.

2. Intro – lines 45-54. Small proteins/microproteins are often encoded within coding frames of larger proteins, and this is clear mechanism of recoding. Viruses/phages are known to do this since decades. In the last years, people have started to realize that small proteins are abundant and important in all sorts of organisms.

3. The authors make a big point of intragenic inversions being (likely) more common when bacteria are part of more complex or natural settings. This is plausible, but I would not think that abiotic stressors (lines 114-116) is not the major driver here, but rather interactions with the host and the other members of the community.

4. Fig 1B and lines 143-144; it's very difficult to follow in which and how many genes you find intragenic invertons in BTh. This info as well as domains/functions they carry and type of inversion (recoding vs premature) should be all clearly shown.

5. lines 168-170; if ecological setting important to identify invertons, why not growing BTh as part of synthetic communities and sequencing? This should be fairly easy and help validate more.

6. lines 218-243: this part is rather descriptive and could be cut down. The interesting result is on intragenic inversions being less bimodal in complete genomes, which is a nice finding! It also begs the question if same holds true (less/more pronounced) in MAGs and metagenomics data.

7. thiC story: although data is clear that the inversion leads to a premature stop and a thiC null phenotype, it would be nice to check protein abundance of ThiC and other proteins in this operon in both locked versions (forward/reverse). It's unclear whether the SAM domain of ThiC (and other proteins of operon) is expressed in the reverse locked form, and if yes, why this would lead to loss-of-function, as this is the catalytic domain.

Knowing that E. coli has a riboswitch in this operon, makes me wonder if the inversion is regulated by substrate availability. Have the authors checked for this?

8. Discussion is rather long – e.g. part on advantages of phase variation (or heterogeneity more general) can be shortened and reader forwarded to tens and tens of reviews on the matter.

Referee #4 (Remarks to the Author):

In this paper, Chanin and colleagues perform an extensive examination of a particular aspect of prokaryotic genome variability, known as invertons. They detect invertons in a pre-existing dataset using an older short-read based tool, then create and apply a new tool for identifying invertons using long-read technology, across various isolated genomes and metagenomes.

I found the paper to be engaging and well-conducted. Additionally, I downloaded, tested, and used the publicly available tool, PhaVa, and found it to be intuitively designed and user-friendly. Nonetheless, as outlined below, I believe certain modifications to both the paper and the software are necessary prior to publication. These modifications include further validation experiments for the program in the paper, and enhancements to the public software to increase its user-friendliness. I anticipate these updates to be manageable and believe they will significantly enhance the understanding and uptake of the method. Detailed comments are provided below.

Major manuscript comments:

Initially, in the results section, the authors confirm some identified invertons using PCR, a robust real-world validation for PhaseFinder. However, I did not find a similar validation of PhaVa. The synthetic read tests to assess the false positive rate of PhaVa are appreciated, but real-world testing like this would strongly reinforce that PhaVa functions as expected.

I have several questions about the synthetic tests (discussed in the paragraph beginning on line 174, Figure 2B-D, and Figure S3). Given these tests are crucial for evaluating PhaVa's accuracy, they warrant careful consideration.

1) The description of the test is somewhat unclear. The methods mention obtaining a reference genome and long-read dataset from NCBI for each species, but Data S7 indicates the use of various long-read datasets per species. Please clarify the methods section.

2) The logic of the test is confusing to me. The use of random coverage values (e.g. “23.61962564”) instead of whole integers (25x, 50x, 100x, etc.) is confusing. Additionally, why 6 different replicates were performed at each coverage level is confusing. Furthermore, if synthetic reads are derived from actual reads instead of directly from the genome, the reasoning behind this choice needs clarification.

3) I don’t understand why Streptomyces eurocidicus and Enterococcus faecalis simulated reads had poor read mapping characteristics (line 844)? If the simulation process is consistent, why would these specific genomes result in lower quality simulated reads?

4) Figure S3 is challenging to comprehend. If possible, clarifying this figure would be beneficial.

5) The interpretation of the accuracy tests for PhaVa seems overly optimistic. In the main text it says “In general, the false positive rate was very low (mean false positive count per simulated sample of <0.1 in 9/10 species), with the exception of reads simulated from the Bordetella pertussis genome

(Fig. 2D).” The text indicates a low false positive rate, yet the presence of significant false positives in 10% of the genomes tested suggests a higher rate. Excluding this genome from the accuracy evaluation doesn't seem fair.

6) Extending this simulation to a larger number of genomes would be advantageous for assessing the commonality of this issue.. It seems that the authors already have ~4,000 genomes on hand, and it doesn’t seem hard to generate synthetic long-reads from each of them to run this test. Rather than using this “NanoSim” method, can you not just simulate long-reads based solely on the genome? I believe doing so would still reveal the problem shown in Figure S3, as intended. Running on far more genomes like this would be really ideal to see how common this problem is. Alternatively, if the authors feel this is too much trouble, they could report the 10% false positive rate more transparently.

The use of terms like “sample”, “genome”, “species”, “isolate”, and “dataset” in the text from lines 192 - 243 is confusing. For example, it seems sometimes “genome” refers to a set of reads, but sometimes it refers to the reference genome being mapped to. In the caption for Figure 3B, does genome mean “sequenced isolate”, or does genome mean “all read sets mapping a specific reference genome”? Precise definitions and consistent usage of these terms would aid comprehension.

In the discussion it’s stated that “PhaVa uses relatively strict mapping parameters and if the selected reference is distantly related to the sequenced strain, read mapping quality will decrease and reduce the discovery rate.” This seems important. I think it would be ideal to directly test and report how similar the reads have to be to the genome for PhaVa to work properly.

Minor manuscript comments:

The introduction section is highly informative and well-composed.

Is there any indication of how the requirement for long-reads to fully overlap the inverton affects false negatives?

Was PhaVa tested on the locked thiC strain to confirm no invertons at that location?

In Figure 2D, what constitutes a false positive? Is it a detected inverton or an inverted read mapping?

Please specify the version of PhaVa used throughout the manuscript in the methods section.

Comments on the PhaVa GitHub program:
Installation:

The installation instructions are insufficient, covering only the cloning of the repo without detailing program installation. Typically python programs have a “setup” script to install the program and put the “bin” scripts in the user’s path; I had to do that all myself.

The manual doesn't mention the requirement for python 3.9 (due to the version of pysam used)

Hosting the program on PyPi or conda would simplify installation for users.

In case it's helpful, a successful conda environment for the program can be created with the following command: “conda create -n phava emboss=6.5.7 minimap2=2.17 pysam=0.17.0 biopython=1.81”.

General software engineering:

The test suite failed post-installation. Unclear if this is due to unaccounted for differences in my system vs. yours, or if the test just doesn’t work anymore.

Versioning is good, but a changelog detailing version differences is needed.

User manual:
Output files lack headers, making interpretation difficult. Headers should be added.
Documentation on the “filter” command, as indicated by pytest and the manuscript, is missing.

More detailed information about each output file is needed. The program makes a lot of files in the output folder, and it’s unclear which are important and which aren’t.

It’s not essential, but a tutorial would greatly aid in making the program more accessible, especially for less experienced users.

Author Rebuttals to Initial Comments:

Referee #1 (Remarks to the Author):

This manuscript reports intragenic DNA inversions as a new class of invertible elements found entirely within genes, which allow a gene to encode two or more versions of a protein by flipping a DNA sequence within the coding region, thereby increasing coding capacity without increasing genome size. The authors use a combination of bioinformatic and experimental approaches to identify and study intragenic invertons in a subset of bacterial species. A long-read based inverton predictor (PhaVa) has been developed and applied to discover intragenic invertons. This bioinformatic approach identified intergenic, partial and intragenic invertons across multiple phyla in bacteria and archaea. A specific intragenic inverton in BT0650 (thiC) from Bacteroides thetaiotaomicron was analyzed, and two "locked" versions of thiC invertons were generated to prevent inversion from occurring within the gene. The locked forward and locked reverse strains were grown in varying concentrations of thiamine to assess their growth rate, and competitive growth experiments were performed to determine the relative fitness of locked forward and locked reverse strains compared to the wildtype strain. The results showed that the locked reverse strain had a fitness advantage over the locked forward strain at physiologically relevant thiamine concentrations. This suggests that the intragenic inversion within thiC gene may provide a competitive advantage in certain conditions.

We appreciate the reviewer's careful evaluation of our manuscript and the detailed comments. We have addressed each comment, point-by-point, below.

Major comments:

1. The existence of potential variable proteins should be demonstrated at the protein level. If the intragenic inversions that don't introduce stop codons nevertheless lead to loss of protein stability, for example, no protein will be present. In this case, both inversions that introduce stop codons and inversions that don't create stop codons would be predicted to phenocopy gene deletion. Protein level confirmation would also help address the possibility that the reported novel sequences could be artifacts resulting from PCR chimeras or other steps in the sequencing approaches.

The reviewer brings up a good point. Both intragenic invertons that lead to a premature stop codon and those that recode a gene may lead to loss of protein stability. As we have focused on experimentally characterizing consequences of inversion in *thiC* in our original report, we decided to focus on this protein for a new experiment and analysis to assess the question of protein stability. These data are now presented in a new panel in Fig. 5D and additional detailed proteomic information is presented in the new Fig. S10. Our new results are summarized below.

To determine the *thiC* intragenic inverton's effect on protein presence, we performed targeted proteomics on ThiC. We hypothesized the *thiC* intragenic inverton would introduce a premature stop codon and that it had the potential to create both a truncated N-terminal domain of the protein as well as a second open reading frame downstream of the premature stop codon (Fig. 5D, Fig. S10). Inversion in *thiC* recodes a small portion of ThiC between the inverted repeat and the stop codon. This creates a unique BTh peptide that we can target in our proteomic analyses. We found specific evidence of the truncated N-terminal domain of the *thiC* gene product in the unlocked reverse BTh mutant and locked reverse mutant. We did not detect this unique peptide in any of the *thiC* forward strains. The N-terminal domain contains the ThiC-associated domain. While this domain is found in ThiC proteins, it has no characterized function.

Furthermore, we did not identify any of the unique peptides associated with a second predicted ORF that begins downstream of the stop codon in the *thiC* reverse strain. By contrast, peptides associated with the downstream ORF were found in both the wild type and locked forward strains. This second ORF contains the catalytic ThiC radical SAM domain and its absence in the reverse strains likely explains why it is unable to grow without supplementation by exogenous thiamine.

Unexpectedly, our proteomic analysis also allowed us to detect an additional peptide that might originate from a new ORF oriented on the antisense strand of inverted *thiC* strains. This ORF falls almost entirely within the inverted region and, except for 3 amino acids at the end, its protein product is identical to the amino acid sequence found within the forward version of the ThiC inverton. While we currently do not know the mechanism underlying this ORF's transcription and translation, its presence indicates that inversion within coding regions can have both transcription and translational consequences. Analysis of the new RNAseq data that we generated suggests that there is an antisense transcript that is generated in the reverse mutants, which is likely translated into this new small peptide.

Taken together with our DNA and RNA analyses, we demonstrate that the *thiC* flipped intragenic inverton is transcribed and translated and therefore is not an artifact resulting from PCR chimeras or other steps in the sequencing approaches.

2. Many genetic rearrangements, transposon insertions, and indels lead to premature stop codons or other loss-of-function consequences. The novelty of this report would be increased by demonstrating that inversions that do NOT introduce stop codons don't simply cause loss of protein stability. No examples are provided of intragenic inversions that actually have distinct biochemical functions and phenotypes compared to both the non-inverted sequence and the deletion mutant.

While our report describes the phenomenon of intragenic invertons for the first time, we agree that adding an example of an intragenic inversion that leads to distinct biochemical functions and phenotypes would further increase the novelty of this report. To do so would require identification of a candidate gene that undergoes a recoding inversion, and then we would have to test for the function of the candidate gene and its inverted counterpart, ideally within its organism (i.e. not solely by purifying the protein and then assessing an in vitro function). We agree that this is important, and if done rigorously and done well it would be very exciting. Because this would require clear phenotypic assays to assess gene function, we sought to identify examples of inverton-recoded genes with well-described functions (and accompanying assays) that we could test. Unfortunately, the vast majority of the recoding invertons that we detected are in 'hypothetical proteins', and while some proteins do have named functions in organisms, the assays to test their functions are not straightforward.

Thus, to address the reviewer's important comment, we took into account a similar and thoughtful recommendation from Reviewer 3, who suggested addressing this topic with computational predictions. We carried out an in silico analysis focused on exciting intragenic invertons in genes of known function (or at least those that have bioinformatically annotated functions). We present three examples of interesting recoding inversions in the manuscript in the revised Fig. 3, and summarize our findings below.

First, we identified a recoding intragenic inverton in *Bordetella bronchiseptica slmA*. SlmA is a nucleoid occlusion factor important for timing cellular division. It contains a DNA binding domain and a FtsZ binding domain. The *slmA* inverton affects 128 out of its 191 amino acids and primarily affects the domain which interacts with FtsZ. When we AlphaFold both the

forward and the reverse versions of the protein, different structures are predicted (RMSD across all 157 pairs: 26.287 angstroms). As these structures only minimally overlap, we used the structural alignment software FoldSeek to determine which proteins these folded structures resemble. As expected the wild type SlmA structure is similar to other *slmA* genes. However, when we perform a FoldSeek analysis on the reverse version of the protein, we find that it is most closely related to DNA transcription factors that contain a tetracycline repressor domain. Based on the structural differences and predictions, we hypothesize that the inverted form of SlmA has a distinct biochemical function than the wild type form. We are keen to carry out follow up experiments to test this hypothesis, but feel that detailed characterization of the function of a protein of predicted, but unknown function in an organism like *B. bronchoseptica*, which is not a model organism with many readily available and robust genetic tools, will require a longer term, dedicated effort. We have reached out to experts who study *Bordetella* species with the hope of setting up a longer term collaboration to further characterize this intriguing intragenic inverton.

Second, we identified recoding intragenic invertons in *barA,* a two component system of *Aeromonas hydrophila*. While *barA* has not been characterized experimentally in *A. hydrophila*, two component systems are important signaling pathways in bacteria that allow them to sense and respond to different stimuli, cues or signals in their environment. The intragenic inverton we identified recodes 40 amino acids in the HPt domain which is responsible for passing on the signal from the sensing protein (BarA) to its cognate response regulator (generally a transcription factor that alters gene expression in response to the transmitted signal). Interestingly, the inverted form maintains the catalytic amino acids in this domain and we hypothesize that this could influence either the efficiency with which it transmits its signal or even alter the receptor that it interacts with.

Finally, we identified two recoding intragenic invertons in *hsdS* genes found in *Mycoplasma hominis*. HsdS is the specificity protein of the Type 1 restriction enzyme complex and determines which motifs are targeted for methylation or restriction. For the *hsdS* genes, the inverton recodes regions in the TRD domain which is important for interacting with DNA. Intriguingly, TRD domains in *hsdS* genes are known for their ability to phase vary as the capability to change interactions here is important for protection against phage or other foreign DNA and in methylation patterns which can impact transcriptional programming of the cell. Previously identified mechanisms of phase variation in these genes are via shufflon systems (where domains are swapped out between homologous genes located in close proximity) and via slipped strand mispairing which causes premature stop codons and alters expression of genes. Here, we hypothesize that intragenic invertons would expand on this and demonstrate an additional way of generating variation in these genes.

We present these findings in revised Fig. 3.

3. The authors state that stop codons introduced by intragenic inversion are reversible, while other nonsense mutations are not. This should be clarified – mobile genetic elements can move twice, nonsense mutations can revert back (e.g. PMID 37745437), under the appropriate selective pressure. No information quantifying the rates of reversion through these different mutational mechanisms is provided.

We have clarified our language in the discussion of the *thiC* intragenic inverton to remove any confusion. We agree with the reviewer that a reversion of a single nucleotide polymorphism or excision of a mobile genetic element could occur. Our intended argument here was that invertons may be more readily reversible as they flip through an enzyme mediated process (where the enzyme class is well described to carry out a reversible reaction) at targeted spots in the genome via motifs in inverted repeats that are recognized by an invertase. However, it is absolutely true that mobile genetic elements can also move twice. With regards to rates of frequency between MGE movement and inverton flipping, we are hesitant to discuss rates in absolute terms as the specific inverton/MGE and context of the bacterium can have a substantial impact on the rate of flipping/movement. For example, some invertons have been found as commonly as 1 in every 10 cells in certain conditions, whereas others are below the limit of detection in certain conditions. Exploration of the rates of inversion mediated by different invertases is an important area for further study, and one that we are excited to embark upon next.

4. Line 155: The statement that intragenic inversions in an invertase possibly changes the invertible repeats that it targets for flipping or its affinity for these repeats would require experimental support.

We agree with the reviewer that verifying the biochemical consequences of this inversion and the others described in this figure would require future experiments.

We have clarified the language of this section to make this more apparent.

5. Line 240 and Figure 3D identifies both the forward and reverse orientations within a given 'clonal' sample. If this is the same genotype found in laboratory-grown BTh type strain (Figure 4B), what is the proportion of these forward and reverse gene copies in the wildtype culture?

Thank you for the opportunity to clarify. Old Fig. 3D (now Fig. 2D) is not attempting to represent information from the laboratory-grown BTh strain. Rather, it displays the ratio of forward and reverse reads identified in our analysis of long read sequencing data from 30,000 prokaryotic isolates in NCBI's SRA (a large, public database). These values are taken from across all the isolate datasets PhaVa detected as containing invertons (having at least the minimum number of forward or reverse reads required by PhaVa's benchmarked sensitivity thresholds to call an inverton).

6. It is unclear how (or whether) the inversions are catalyzed. Are there dedicated invertases, and what are the fitness and biochemical consequences of deleting these enzymes? This is a major aspect of the proposed biological mechanism that is not addressed.

We agree that determining which invertases flip each inverton is important; in the revised manuscript, we have carried out both RNAseq and a 110-day experimental evolution experiment to try and address this important issue. Our results from these experiments (which are in the revised Fig. S11, S12, and S13) and a brief summary of how invertase/inverton pairs have been identified are provided below.

Invertons in intergenic regions of DNA have been well described in the literature - and detailed investigations (often over decades) have identified specific invertases that are responsible for catalyzing DNA inversion (Abraham et al. 1985; Howe 1980; Zieg and Simon 1980; Nakayama-Imaohji et al. 2009). These invertible regions are typically surrounded by prototypical inverted DNA repeats, which are recognized by their cognate 'invertase' enzyme - usually a tyrosine recombinase or small serine recombinase - and are flipped. An example of a classical example of this is the Type 1 fimbrial operon in *E. coli*. Because definitively matching an invertase to an inverton (the region that is flipped and the bounding inverted repeats) requires an important series of necessity and sufficiency experiments, only a handful of well studied intergenic invertons have known matched invertases. We hope that with our discovery of intragenic invertons, we and others will have

increased interest in developing methods to address this limitation and generate higher throughput methods to match invertases to invertons.

Despite the limitations noted above, we fully agree that attempting to identify the invertase that catalyzes an intragenic inversion would be very exciting. Unfortunately, in many organisms, there are many (>10) predicted invertases, which makes using genetic deletion approaches to study invertase specificity cumbersome. For example, in BTh there are 54 invertases; following on this, generating a mutant that is deficient in all but one invertase would be more than a year's worth of work (53 serial mutations), even if all of the invertases are non-essential.

Thus, we sought to define a series of detailed experiments, carried out over the course of several months, that could orthogonally address all aspects of the reviewer's comments.

Given that our detailed analyses focused on the *thiC* intragenic inverton, and that the phenotypic consequences of the locked forward vs. the locked reverse inverton are well characterized in our original manuscript, we decided to focus on this inverton.

FIRST, we tried to figure out which invertase was driving inversion and hypothesized that low thiamine might trigger expression of that invertase. We therefore performed RNAseq on BTh strains (WT, locked forward, locked reverse, and clean deletion) in different thiamine concentrations to see if an invertase was upregulated in response to thiamine availability. However, RNAseq did not reveal an invertase that is statistically significantly differentially expressed (Fig. S11, Data S9). Several Bth invertases are expressed constitutively in all conditions and all mutants; others are lowly expressed or not expressed at all. This suggested to us that the invertase driving *thiC* flipping is not transcriptionally regulated by thiamine availability. Consequently, we predicted that in a low thiamine concentration, the flipping would be slow, and likely a stochastic phenomenon mediated by a less active, poorly expressed or rarely expressed invertase.

SECOND, we carried out a 110-day experimental evolution of strains grown in low or high concentrations of thiamine. To detect inversion events, we created an unlocked reverse and unlocked forward mutant (Fig. S13). In contrast to the locked versions, these strains maintain their inverted repeats without any introduced nucleotide substitutions and can therefore be recognized by their cognate invertase. Additionally, these strains were generated from the *thiC* clean deletion strain, meaning that any flip event detected indicated an inversion event. Once generated, the unlocked reverse mutant was serially cultured in a low thiamine containing medium (0.001 µM), and the unlocked forward strain was grown in high concentrations of thiamine (10 µM). Cultures were passaged once they reached saturation (every 24-48 hrs). DNA was isolated from every 3-4 transfers was isolated from each transfer and the orientation of the *thiC* inverton was assessed with qPCR. This experiment was carried out for 110 days (100+ transfers).

Based on our competition assay (Fig. 5E), we hypothesized that the selected thiamine concentrations would provide a growth advantage for any strain that flips (FW to RV or RV to FW). As expected, we found that flipping did occur in both conditions, however it happened slowly and somewhat stochastically, consistent with our above model. Of note, one replicate of the unlocked reverse strain completely flipped to forward and then intriguingly the reverse orientation appeared at very low levels suggesting that there is an advantage to having a very low level of the opposite orientation, even when that orientation is at a significant growth disadvantage.

Taken together, we determined that based on the similarity of the *thiC* intragenic inverton to known intergenic invertons, this DNA flip is likely enzyme mediated. The invertase that flips this locus is expected to be encoded within BTh (not from a phage or other exogenous mobile genetic element), because in 110-day experimental evolution, we find evidence of flipping of this locus in all of our experimental arms. RNAseq suggests that the invertase that flips *thiC* is not transcriptionally induced by different thiamine concentrations, and again, based on the evolution experiment we postulate that flipping happens at a relatively slow rate. This is more consistent with a model where in these conditions the invertase that flips this locus is expressed at a very low level, and that it stochastically flips the *thiC* locus in rare instances. We cannot rule out that other conditions that we have not tested (including much more natural conditions such as presence in a complex microbial community or in a biofilm) induce expression of the invertase that flips this locus. We hope to design a dedicated future study to explore this interesting area.

7. In Figure 4C, the wildtype strain and the locked forward strain maintained similar growth densities at different concentrations of thiamine, while the density of the locked reverse strains gradually increased as thiamine concentration rose from 0.1-10 μM. The expression levels of the three versions of thiC should be assessed under these different thiamine concentrations to determine whether the expression of each version of thiC correlates with thiamine concentration.

To verify that there was no difference between expression levels of *thiC* in the different strains (WT, locked forward, and locked reverse), we performed RNAseq at various concentrations of thiamine 0.001, 0.1, and 10 μM (Fig. S12A). We found no significant difference in RPKM values of *thiC* between strains grown in similar conditions.

8. In metagenomes and the isolates of Bacteroidetes dataset you have analyzed (Figure S6), what is the frequency of occurrence of the invertons in thiC, and what is the phylogenetic distribution?

In the long read isolate datasets and metagenomic datasets we did not identify any additional Bacteroidetes strains that contained invertons with read evidence in the *thiC* gene. However, we did identify a recoding intragenic inverton in the thiamine pyrophosphate-binding protein gene of *Burkholderia mutivorans* in our long read isolate dataset. As we discuss in the paper, we think that the list of invertons we identify in the long read isolate dataset is a minimal set and that more may be discovered if/when organisms are exposed to different conditions.

Minor comments:

9. For Figures 3A and B, a table showing the number of genomes in the different phyla being investigated and the predicted number of intergenic, intragenic, and partial intergenic invertons, would support the result. In the middle panel of Figure 3B, the genomes of Deinococcus-Thermus and Gemmatimonadetes have the highest average mean number of invertons: which genes carry invertons and what consequences the invertons produce would help provide more information from this figure panel.

Thank you for your suggestion. We now include a new Fig. S7, which shows the number of genomes in the different phyla that were investigated. Each detected inversion and its category (intergenic, intragenic, partial) are included in the new Data S4. Species level information regarding the isolate data that were included in this study are also available in the new Data S5.

10. In Figure 4A, the complete nucleotide and AA sequences of forward and reverse orientations would provide a useful supplemental figure.

Thank you for your suggestion. We have included this data in Data S10.

Referee #2 (Remarks to the Author):

Drs. Chanin, West, Bhatt and colleagues have produced a remarkable study and should be commended for their careful yet highly novel work. While they build on the findings of others, their achievements in this study are considerable and include: identifying a new subset of invertons that are intragenic and can produce early stop codons or new open reading frames, developing a new tool and using it to quantify invertons using long-read sequencing to better identify and characterize invertons, finding that invertons are much more abundant in metagenomic datasets obtained from patient fecal samples when compared to in vitro cultivated isolates, and finally experimentally demonstrating that the state of a particular intragenic inverton can modify the competitive fitness of a strain of bacteria. This study is impactful and will be of interest to a broad audience. I have some comments and questions that I think could help improve what is already an impressive body of work.

We thank the reviewer for their careful review of our manuscript and their very kind comments. We are excited to build on many excellent foundational pieces in the inverton literature to help advance this exciting field. Below, we have addressed each of the reviewer's specific comments in detail.

Fig 1b: In clinical samples where BTh intragenic inversions are broadly observed, is there anything notable about these samples? Do they possibly show domination by BTh? Are they particularly low in bacterial density?

This is a good question. To address it, we now provide a new supplementary figure (Fig. S2 and related Data S2) with a histogram describing the taxonomic composition of each of the samples that are presented in the analysis. None of these samples are dominated by Bth. We also calculated Shannon diversity of these samples, and find that there is no obvious association between abundance of inversions measured and diversity. Taken together, we did not find any noticeable factor from any patient associated metadata that suggested a cause for the increased number of invertons detected in BTh. Rather, lack of invertons found in samples seemed to correspond to lack of BTh read coverage. Fig. S2 and Data S2 now show the taxonomic composition of short read samples for both samples where BTh intragenic inversions were detected, and samples where they were not. In addition, we show no significant difference in Shannon diversity between these two sets of samples.

However, we are actively investigating drivers of inverton flipping in vivo and are excited to explore this important aspect in future work.

It seems that when the authors in Fig 1B were evaluating for predicted intragenic invertons using PCR, they were evaluating a polyclonal mixture in liquid media, and confirmed 10 predicted intragenic invertons. What would they observe if they screened individual clones? Is it possible to perform a kinetic analysis of single clones over time, evaluating the status of these intragenic invertons? Or do these 10 intragenic invertons flip so rapidly that by the time a single clone expands sufficiently enough to produce enough genetic material to allow evaluation, there is already achievement of a steady-state mixture?

To clarify - our in vitro analysis of intragenic invertons in *B. thetaiotaomicron* was performed on the laboratory strain VPI-5482. All assays were performed on liquid cultures that derived from a single colony forming unit on agar plates. We would therefore not classify them as a classical polyclonal mixture, per se. To avoid confusion for the reader, we have clarified this in the methods section. However, based on our results detecting both orientations of these invertons, these cultures clearly contain subgroups of organisms with intragenic invertons flipped in different orientations. As we detected these flips across multiple cultures from different initial starting colonies, we suspect that a steady state is reached rapidly during growth in liquid media. It is possible that the rate of inversion is different in solid agar growth conditions - but obtaining enough DNA to prepare a high quality sequencing library from a single colony is more technically challenging, so this was not pursued at present. It could be an interesting experiment for future work. Future experiments could also leverage single cell bacterial DNA sequencing approaches, which while difficult, are now possible.

Relatedly, would it be possible to characterize the stability of invertons after introduction of a strain to mice? Introducing their well-characterized BTh strain to a germ-free mouse to monocolonize, followed by quantification of inverton status, would be very interesting, especially for the 49 predicted intragenic invertons which could not be confirmed in vitro. The kinetics of status changes in these invertons in vivo would also be interesting.

To start to answer this question, we analyzed all publicly available paired end short read data of the laboratory BTh strain (VPI-5482), both from mouse experiments and from laboratory conditions. We used the PhaseFinder tool to assess inverton status in these samples. We are limited to paired end data as PhaseFinder requires this type of input. Unfortunately the vast majority of paired end data are RNAseq datasets. While still helpful

in identifying intragenic invertons, as these would fall within the coding region, transcriptional differences between genes could greatly affect which intragenic invertons we would detect (intragenic inverton containing genes with lower levels of transcription may not be detected even if flips are present). Additionally, while there are a few DNA datasets taken from mouse samples, these are primarily from BTh strains that have been isolated from fecal samples and then propagated in laboratory conditions to generate enough DNA for sequencing. In most of these studies, authors were looking for evolved strains or SNPs that occurred over time. Subsequently, any intragenic invertons that may have been advantageous in vivo but not costly in vitro might have flipped back to the reference form or might not have been selected for during the extended growth in laboratory media.

With those caveats, it is challenging to make any definitive conclusions and we therefore have not added these data to the manuscript. However, in the datasets we analyzed we found 62 total intragenic invertons with sequence support. Five were only found in vitro only (standard laboratory conditions but experiments may include other variables), 19 found in vivo only (generally colonization experiments in mice), and 38 found in both. Of the intragenic invertons we confirmed in our laboratory strain (Table 1), two were found in these samples (BT3786 found in in vivo only samples and BT0629 in both in vitro and in vivo conditions). Additionally, we found sequence support for 4 invertons we did not confirm in our laboratory strain in vitro (BT1992, BT3708, BT0439, and BT4731).

We agree with the reviewer that our findings raise many more interesting questions and we are excited to pursue this further in other works. To achieve the necessary level of rigor for inclusion in a manuscript, we anticipate that carefully designing and executing future experiments will be critical. We envision experiments where both stool and cecal contents are evaluated - and ideally paired DNA and RNA sequencing are performed. We are in the process of working with collaborators to design such an experiment, and hope that we can also collect matched host phenotypic data (such as immune cell information).

Can the authors try to search for genetic features that distinguish invertons that flip in vitro (and presumably also flip in vivo) from invertons that only flip in vivo? Or do they think that these invertons all flip similarly, but some permissive conditions allow for production of heterogeneity in the state of some invertons, while other conditions strongly select for a particular state?

The reviewer raises a fascinating question by which we are also very intrigued. Our data indicate that certain flippable loci exist in a 'fixed' or near 100% forward or reverse orientation, whereas others have 'heterogeneity in the state' - i.e. the community has a mix of the forward and reverse orientation.

The reviewer offers two logical models that represent a 'reduction to extremes' of what might happen: in model 1, they suggest that certain invertons are flipped in some conditions vs. others - likely the result of, for example, induction of expression of specific invertases. In model 2, they suggest that invertons all flip similarly (all invertible loci are flipping all of the time) and that selection happens on this standing variation. While we do not have resolved enough data to come to a strong conclusion of which of these models are occurring, we have collected some additional data that give us some insight into this good question.

To address the question of whether invertases are active 'all of the time', we took our laboratory Bth strain under various growth conditions (in vitro) and performed RNAseq. Bth has 54 different annotated invertases. The expression of the different invertases seems to be pretty consistent between conditions, suggesting that the different in vitro conditions we tested don't dramatically affect invertase expression (new Fig. S11). One invertase seems to be rather highly expressed (BT2742). Other invertases seem to be pretty much "off" - with very low or undetectable RPKM values. Another subset seems to be expressed, but at a low level.

We then further assessed one of the RNAseq datasets in which we identified intragenic invertons for invertase expression. [REDACTED]

[REDACTED]

[REDACTED]

The experimental strategy to examine for the consequences of the thiC intragenic inversion is very nicely and carefully done. Using the wild-type strain, is it possible for the authors to demonstrate over time enrichment of the two possible states, when BTh is cultivated in high or low thiamine concentrations? Can they additionally generate an invertase mutant strain that in turn fails to adapt to ambient thiamine conditions?

We appreciate the reviewer's kind words on our experimental strategy. The reviewer has suggested an excellent experiment - which we have now done (with some slight modifications).

To address this comment, we created an unlocked reverse and unlocked forward mutant (new Fig. S12). In contrast to the locked versions, these strains maintain their inverted repeats without any introduced nucleotide substitutions and can therefore be recognized by their cognate invertase. Additionally, these strains were generated from the *thiC* clean deletion strain, meaning that any detected flip indicated an inversion had occurred. Once generated, the unlocked reverse mutant was serially cultured in a low thiamine containing medium (0.001 µM), and the unlocked forward strain was grown in high concentrations of thiamine (10 µM). Cultures were passaged once they reached saturation (every 24-48 hrs). We used qPCR to assess the orientation of the *thiC* inverton. This experiment was carried out for 110 days (100+ transfers). Based on our competition assay (Fig. 5E), we hypothesized that the selected thiamine concentrations would provide a growth advantage for any strain that flips (FW to RV or RV to FW).

While we detected flips in both conditions (reverse to forward and forward to reverse) the thiamine concentration did not appear to be a strong driver of flipping. This along with our RNAseq experiment (new Fig. S13, Data S9), which demonstrated that thiamine concentration did not change the expression of BTh invertases, suggests that in these conditions flipping is stochastic and low frequency.

Of note, one replicate of the unlocked reverse strain did completely flip to forward at 40 transfers and then intriguingly the reverse orientation appeared at very low levels after transfer 66. This may suggest that there is an incentive to have both populations in these conditions, even when one subpopulation is significantly less fit. It is also interesting that in both arms of the experiment (unlocked reverse in low thiamine and unlocked forward in high thiamine) that a ratio of one population to the other approached 1 in every 10^5 cells.

In Figure 2B, the authors should convey the SD or SEM in their analysis.

Thank you. We have added the appropriate statistical information for these data in a new associated supplementary file (Data S3); it contains the mean false positive count per sample standard deviation.

Referee #3 (Remarks to the Author):

Bacteria are known to use DNA inversions to create heterogeneity and express different surface sugars – a process long known as phase variation. Using long-read sequencing of a longitudinal microbiome study, Chanin et al. report for the first time the existence of intragenic DNA inversions in bacteria. They then develop a new method to identify thousands of such inversions in >35,000 complete bacterial genomes (and a limited number of MAGs), and experimentally characterize one specific intragenic variation leading to premature termination of thiC and of thiamine biosynthesis in Bacteroides thetaiotaomicron.

This is an exciting study that opens a new field of research in bacteria, providing also the tools to enable further discoveries. However, there are several points that need to be fixed in the analysis and presentation of the results and new tool, so the utility of this resource fulfills its great potential.

We thank the reviewer for their enthusiasm for our work and appreciate the comments and suggestions that were made to help our resource and tool achieve “its great potential”.

Main points

1. The authors make a major claim about the recoding potential of intragenic inversions (leading to expression of new functional proteins) throughout the manuscript, including the abstract. This is indeed exciting, and a clear depart from intergenic (or partial intergenic) inversions. Yet only a (small) fraction of intragenic inversions can do this – those with recoding potential (in-frame inversion) or when the premature stop codon leads to a split protein. Unfortunately these more interesting cases are not presented or followed up to any depth in the manuscript – e.g. the main experimental follow up focuses to a case of a premature stop codon, which leads to more mundane case of an ON/OFF state of the gene. Although I recognize that asking for a new experimental follow up on a recoding case is a big ask at this point, there are other ways the authors could showcase better the recoding cases. First the authors should clearly distinguish and report the numbers of recoding vs premature intragenic cases throughout the different datasets they use (transplantation patient microbiomes, complete genomes and MAGs) – now all reported together. Ideally they should further differentiate the premature cases to those that potentially lead to a split

protein vs those that only seem to prematurely stop the protein. Second, some functional analysis on the recoded (and split) domains is warranted and whether it happens at specific, conserved, catalytic, or protein-protein/-ligand binding sites. Also, there must be some examples (out of the thousand cases) where there is good structural and/or functional understanding of the protein and its binding partners, so the authors can use structural predictions to illustrate possible ramifications of the recoding.

Thank you for the suggestions. We agree that the recoding inversions are very intriguing and are excited for follow up work on this topic. We have addressed each of the reviewer's comments and appreciate their acknowledgment that detailed experimental characterization of the function of a protein and a recoded variant is a 'big ask'.

We have now differentiated cases of recoding vs. premature stop in our analyses of various sequencing datasets (Fig. 2A and 4A), as suggested. Additionally we have added the prediction of what the effect of each intragenic inverton (recoding or premature stop) to the PhaVa output.

We also thank the reviewer for their suggestions on how to increase the impact of our work through in silico predictions of intragenic inverton effects on well characterized proteins. We have now incorporated a new main figure that shows the results of this effort. Specifically, in addition to highlighting the impact intragenic invertons in BTh could have on the proteins highlighted in Fig. 1, we have added in additional analyses on 3 recoding intragenic invertons identified in our long read isolate analysis; these are highlighted in the new Fig. 3.

First, as discussed above in an earlier response to reviewer 1, we identified a recoding intragenic inverton in *slmA,* a nucleoid occlusion factor important for timing cellular division in *Bordetella bronchiseptica*. The inverton, which affects 128 out of its 191 amino acids, alters the structural prediction of the recoded gene. As the domain affected by this inversion is responsible for interacting with FtsZ, we hypothesize this flip would inhibit this interaction. Intriguingly, when put into structural alignment software FoldSeek, the inverted form of *slmA* is returned as now most closely resembling a Tet repressor. This likely results in a change of function of the protein.

Second, we identified recoding intragenic invertons in *barA,* a two component system of *Aeromonas hydrophila*. While *barA* has not been characterized experimentally in *A. hydrophila*, two component systems are important signaling pathways in bacteria that allow them to sense and respond to different stimuli, cues or signals in their environment. The intragenic inverton we identified recodes 40 amino acids in the HPt domain which is responsible for passing on the signal from the sensing protein (BarA) to its cognate response regulator (generally a transcription factor that alters gene expression in response to the transmitted signal). Interestingly, the inverted form maintains the catalytic amino acids in this domain and we hypothesize that this could influence either the efficiency with which it transmits its signal or even alter the receptor that it interacts with.

Finally, we identified two recoding intragenic invertons in *hsdS* genes found in *Mycoplasma hominis*. HsdS is the specificity protein of the Type 1 restriction enzyme complex and determines which motifs are targeted for methylation or restriction. For the *hsdS* genes, the inverton recodes regions in the TRD domain which is important for interacting with DNA. Intriguingly, TRD domains in hsdS genes are known for their ability to phase vary as the capability to change interactions here is important for protection against phage or other foreign DNA and in methylation patterns which can impact transcriptional programming of the cell. Previously identified mechanisms of phase variation in these genes are via shufflon systems (where domains are swapped out between homologous genes located in close proximity) and via slipped strand mispairing which causes premature stop codons and alters expression of genes. Here, we hypothesize that intragenic invertons would expand on this and demonstrate an additional way of generating variation in these genes.

For all three examples described above we have provided AlphaFold structures and included domain diagrams highlighting which regions are affected (Fig. 3). Additional experimental work would be needed to fully characterize the effects of these and other recoding intragenic invertons and we are excited to pursue that work in future studies.

2. To identify intragenic inversions, the authors initially use a longitudinal microbiome study of patients going transplantations. It is unclear why they focus on B. thetaiotaomicron. I would imagine that more abundant and prevalent species (from Bacteroidota phylum that are equally genetic tractable) would be better cases to look into. Looking into several closely related species would also enable some evolutionary comparisons (and gain more confidence for conserved predicted inversions). I would find useful if the authors expanded their efforts here – especially since they make a claim on the necessity of using bacteria in their natural settings for identifying intragenic inversions.

We appreciate the reviewer's suggestion and agree that organisms such as *B. fragilis, P. vulgatus*, and *P. dorei* would also be very interesting to study. To partially address this comment, we have expanded our analysis of the short read data to *B. fragilis*, and noted that we identify potential intragenic inversions in 45 genes (Data S1).

In terms of our choice to focus on BTh, we initially chose to focus our efforts on *B. thetaiotaomicron* because we have substantial experience in genetically manipulating this strain, and it has been the focus of many gut microbiome studies. Because the manuscript is fairly dense and lengthy (describing the discovery of intragenic invertons, characterizing these in patient samples, performing detailed microbiological experiments to deeply characterize a specific inverton in BTh, and then developing/benchmarking and applying a new computational tool), we felt that to give such an important topic its due, it would be better to embark on a new series of experiments in a subsequent project. To that end, we are now learning how to do genetics in *P. vulgatus* and *P. dorei* with the help of colleagues in other labs with expertise in this, and in the future hope to be able to further address the intragenic invertons in these organisms.

3. PhaVa is a new computational tool presented in this study that allows for identification of (intragenic) inversions in long-read data. Although the tool looks specific, I would expect more rigorous benchmarking. Sensitivity/false negatives are not addressed. Results are compared to Jiang study but not to Milman one, which identifies many more insertions using more or less the same set of complete genome sequences. It's important that authors benchmark better their tool and compare it exhaustively with other state-of-the-art tools.

Thank you for the suggestion, in response to this comment and the additional comments from reviewer 4, we have greatly expanded the benchmarking using simulated reads from pbsim2 with a pre-trained read simulation model. Our new benchmarking (Fig. S5A) tests the false positive rate and false negative rate in an additional 100 genomes. We also used this read simulation method in an attempt to address whether taxonomic distance affects the false negative rate (Fig. S5B), the requirement for reads to span invertons affects the false negative rate (Fig. S5D), and how coverage affects the false negative rate (Fig. S5C). We have also added a better description of both read simulation approaches to the methods. The Milman et al. paper released scripts to regenerate their analyses but did not release a generalized tool for us to compare to. They were able to identify more candidate invertons because their approach did not require read mapping based evidence for confirmation - thus, it would be fair to characterize their invertons as "possible invertons", whereas we

present validated invertons based on the inclusion of supportive read data. In our dataset, we find that a small minority of possible invertons actually have read data that support that they are inverted.

4. Authors use Pha Va in 210 MAGs. First this analysis is important and IMO should be part of the main Figs, but has to be done and presented better (present species targeted, differentiate between recoding/premature inversions, perform functional analysis). Second, it's unclear why only 210 MAGs, when there are orders of magnitude more good quality MAGs (with high completeness) available – what is the limitation? This is important, especially if intragenic inversion are more common in bacteria growing natural environments. Third, it is unclear if Pha Va can be used only with MAGs or can be also used directly with reads from metagenomics data – authors should explain.

Thank you for the comment - we agree that the analysis of long read metagenomic data is important - and we have revised the manuscript to improve the analysis and also make the results more prominent and in the main figures, as opposed to in the supplementary figures (new main Fig. 4).

To the reviewer's specific comments, we offer some clarification (which we have also clarified in the text). The 210 refers to long-read metagenomic sequencing datasets instead of individual MAGs. We have made this more explicit in the text. In the new Fig. 4A, we present the inverton count per phylum observed in the metagenomic data. In the new Fig. 4B, we present data demonstrating a statistically significantly higher count of invertons per genome in metagenomic data for all bacteria, and also phylum-specific higher counts of invertons per genome in Bacteroidetes and Firmicutes.

We used 210 long-read metagenomes because the number of metagenomes with high depth of long-reads is still limited. While there are many more hybrid metagenomes with low depth long-read sequencing and higher depth short read sequencing data, low-depth long read sequencing data are not appropriate to use with PhaVa because we require moderate to higher read depth to call invertons with high precision.

Minor points

1. Abstract – it would benefit from being more precise on results of this study. Currently half of abstract is background, and the other half is devoid of specific results/analysis/experiments of this study, having mainly novelty claims. E.g. “thousands” in line 19 is not particularly precise.

Thank you for your feedback. We have adjusted the abstract and inserted more specific results of analyses and experiments including the number of intragenic invertons we identified with PhaVa in our long read isolate analysis, the number of BTh intragenic invertons we confirmed, and detailing our experimentations on the *thiC* intragenic inverton.

2. Intro – lines 45-54. Small proteins/microproteins are often encoded within coding frames of larger proteins, and this is clear mechanism of recoding. Viruses/phages are known to do this since decades. In the last years, people have started to realize that small proteins are abundant and important in all sorts of organisms.

We have updated our introduction to include this information. Thank you for your feedback.

3. The authors make a big point of intragenic inversions being (likely) more common when bacteria are part of more complex or natural settings. This is plausible, but I would not think that abiotic stressors (lines 114-116) is not the major driver here, but rather interactions with the host and the other members of the community.

Thank you for your insight and feedback. Based on our additional experiments, where we exposed BTh to abiotic stressors, we suspect that the reviewer is correct and that invertons might be more likely to be driven by interactions with the host or other members of the community. We have updated the language of this section to include that interactions with the host and other members of the microbial community could also be a driver for inverton flipping and selection.

4. Fig 1B and lines 143-144; it’s very difficult to follow in which and how many genes you find intragenic invertons in BTh. This info as well as domains/functions they carry and type of inversion (recoding vs premature) should be all clearly shown.

Thank you. We have clarified this in various places in the revised manuscript.

For BTh predicted intragenic invertons we chose to highlight the ones that we had experimental evidence of in Fig1B. Table 1 also contains this information for BTh. However, the other predicted intragenic inventions of which we could not confirm in our BTh strain grown in laboratory conditions can be found in Data S1.

More globally, we have updated the PhaVa output (new Fig. S4) so that the effects of intragenic invertons are predicted (recoding vs premature stop). Additionally we have supplied this information for both the long read isolate dataset sets and the long read metagenomic datasets (Fig. 2A and 4A).

5. lines 168-170; if ecological setting important to identify invertons, why not growing BTh as part of synthetic communities and sequencing? This should be fairly easy and help validate more.

We currently have an ongoing collaboration with another research group in which we are actively pursuing this. In this ongoing project, in addition to assessing *B. thetaiotomicron,* we are expanding our in vitro analyses of invertons to other members of the gut microbial community as well. We are excited to pursue this avenue in future work as we agree that we may be able to validate additional invertons (intragenic, intergenic, or partial) in more complex, in vivo-like conditions.

6. lines 218-243: this part is rather descriptive and could be cut down. The interesting result is on intragenic inversions being less bimodal in complete genomes, which is a nice finding! It also begs the question if same holds true (less/more pronounced) in MAGs and metagenomics data.

We are glad that the reviewer found our result - that intragenic inversions are less bimodal in complete genomes - interesting. Our long-read metagenomic data was de novo assembled and we mapped the same reads used for assembly back to the assemblies. Therefore, we largely expect inversion results from PhaVa to report inversion proportions <50%, similar to what would be observed for allele frequencies when calling SNPs with a de novo assembly and its reads. However, we examined inversion proportion in our

metagenomic samples and found, rather than being concentrated at very low inversion proportion, a wide range of inversion proportions for all three categories of inversions, suggesting the presence of populations with a more mixed inversion state in our metagenomes.

We also agree that the section of our results describing our findings from the PhaVa analysis of the long read isolate datasets is descriptive and somewhat dry (previous lines 218-243). We have now removed a few sentences in which we discussed the findings in Verrucomicrobia and speculated on the role the high number of invertible elements in this phylum might play in its lifestyle and interactions with the host. This streamlines the section and, we hope, improves its readability. We do think it important to include the specific findings from PhaVa with context and comparisons to previous studies that have used publicly available data to identify additional invertible elements (Milman et al, and Jiang et al). We have therefore opted to keep some of these details in the revised manuscript.

7. thiC story: although data is clear that the inversion leads to a premature stop and a thiC null phenotype, it would be nice to check protein abundance of ThiC and other proteins in this operon in both locked versions (forward/reverse). It's unclear whether the SAM domain of ThiC (and other proteins of operon) is expressed in the reverse locked form, and if yes, why this would lead to loss-of-function, as this is the catalytic domain.

Thank you for this suggestion. To determine the *thiC* intragenic inverton's effect on protein presence, we performed targeted proteomics on ThiC. We hypothesized the *thiC* intragenic inverton would introduce a premature stop codon and that it had the potential to create both a truncated N-terminal domain of the protein as well as a second open reading frame downstream of the premature stop codon (new Fig. 5D and Fig. S10). Inversion in *thiC* recodes a small portion of ThiC between the inverted repeat and the stop codon. This creates a unique BTh peptide that we can target in our proteomic analyses. We found specific evidence of the truncated N-terminal domain of the *thiC* gene product in the unlocked reverse BTh mutant and locked reverse mutant. We did not detect this unique peptide in the wild type or locked forward strain. The N-terminal domain contains the ThiC-associated domain. While this domain is found in ThiC proteins, it has no characterized function.

Furthermore, we did not identify any of the unique peptides associated with a second predicted ORF that begins downstream of the stop codon in the *thiC* reverse strain. By

contrast, peptides associated with the downstream ORF were found in both the wild type and locked forward strains. This second ORF contains the catalytic ThiC radical SAM domain and its absence in the reverse strains might explain why it is unable to grow without supplementation by exogenous thiamine.

We also evaluated whether *thiH*, the gene encoded downstream of *thiC* is translated in the locked reverse mutant, as it is possible that a premature stop codon in *thiC*, such as that which is in the locked reverse mutant, might prevent *thiH* from being translated. We found clear evidence of ThiH protein being produced in the locked reverse mutant, demonstrating that this is not the case. In conclusion, we are able to detect both the N-terminally truncated version of ThiC and ThiH protein in the locked reverse mutant. The data on ThiH is included in new Fig. S12.

Knowing that E. coli has a riboswitch in this operon, makes me wonder if the inversion is regulated by substrate availability. Have the authors checked for this?

Thanks for this suggestion. We have now checked to see whether inversion is regulated by substrate availability. The details of this experiment are outlined below and now included in the new Fig. S13 and are detailed in the text of the revised manuscript.

To test whether thiamine availability regulated the *thiC* intragenic inverton, we first created an unlocked reverse and unlocked forward mutant (Fig. S13A). In contrast to the locked versions, these strains maintain their inverted repeats without any introduced nucleotide substitutions and can therefore be recognized by their cognate invertase. Additionally, these strains were generated from the *thiC* clean deletion strain, meaning that any flip event detected indicated an inversion event. Once generated, the unlocked reverse mutant was serially cultured in a low thiamine containing medium (0.001 µM), and the unlocked forward strain was grown in high concentrations of thiamine (10 µM). Cultures were passaged once they reached saturation (every 24-48 hrs). We used qPCR to assess the orientation of the *thiC* inverton. This experiment was carried out for 110 days (100+ transfers). Based on our competition assay (Fig. 5E), we hypothesized that the selected thiamine concentrations would provide a growth advantage for any strain that flips (FW to RV or RV to FW).

While we detected flips in both conditions (reverse to forward and forward to reverse) the thiamine concentration did not appear to be a strong driver of flipping. This along with our

RNAseq experiment, which demonstrated that thiamine concentration did not change the expression of BTh invertases, suggests that in these conditions flipping is stochastic and low frequency.

Of note, one replicate of the unlocked reverse strain did completely flip to forward at 40 transfers and then intriguingly the reverse orientation appeared at very low levels after transfer 66. This may suggest that there is an incentive to have both populations in these conditions, even when one subpopulation is significantly less fit. It is also interesting that in both arms of the experiment (unlocked reverse in low thiamine and unlocked forward in high thiamine) that a ratio of one population to the other approached 1 in every 10^5 cells.

8. Discussion is rather long – e.g. part on advantages of phase variation (or heterogeneity more general) can be shortened and reader forwarded to tens and tens of reviews on the matter.

Thank you for your feedback. We removed the paragraph discussing the role of phase variation in bacterial lifestyle and community dynamics as we agree that there are many wonderful reviews on this very fascinating subject. Additionally, we have removed the paragraph highlighting the potential impact of recoding intragenic invertons in BTh as we have significantly expanded discussion of this type of intragenic inverton in the results section. We provide 3 examples (*slmA* from *B. bronchiseptica*, *barA* from *A. hydrophila*, and *hsdS* from *M. hominis*), generate AlphaFold structures, and highlight their potential consequences not only on the protein but on microbial lifestyle (Fig. 3).

Referee #4 (Remarks to the Author):

In this paper, Chanin and colleagues perform an extensive examination of a particular aspect of prokaryotic genome variability, known as invertons. They detect invertons in a pre-existing dataset using an older short-read based tool, then create and apply a new tool for identifying invertons using long-read technology, across various isolated genomes and metagenomes.

I found the paper to be engaging and well-conducted. Additionally, I downloaded, tested, and used the publicly available tool, PhaVa, and found it to be intuitively designed and user-friendly. Nonetheless, as outlined below, I believe certain modifications to both the paper and the software are necessary prior to publication. These modifications include further validation experiments for the program in the paper, and enhancements to the public software to increase its user-friendliness. I anticipate these updates to be manageable and believe they will significantly enhance the understanding and uptake of the method. Detailed comments are provided below.

We thank the reviewer for these positive comments - we are glad to hear that they found the paper to be “engaging and well-conducted”, and are also glad that they found PhaVa to be “intuitively designed and user-friendly”. We greatly appreciate their detailed comments, which we have used to substantially improve and enhance both the benchmarking and the usability of our tool.

Major manuscript comments:

Initially, in the results section, the authors confirm some identified invertons using PCR, a robust real-world validation for PhaseFinder. However, I did not find a similar validation of PhaVa. The synthetic read tests to assess the false positive rate of PhaVa are appreciated, but real-world testing like this would strongly reinforce that PhaVa functions as expected.

I have several questions about the synthetic tests (discussed in the paragraph beginning on line 174, Figure 2B-D, and Figure S3). Given these tests are crucial for evaluating PhaVa's accuracy, they warrant careful consideration.

1) The description of the test is somewhat unclear. The methods mention obtaining a reference genome and long-read dataset from NCBI for each species, but Data S7 indicates the use of various long-read datasets per species. Please clarify the methods section.

There was only one readset used per species. The column in Data S7 was incorrect, which we have now fixed (now Data S12).

2) The logic of the test is confusing to me. The use of random coverage values (e.g. “23.61962564”) instead of whole integers (25x, 50x, 100x, etc.) is confusing. Additionally, why 6 different replicates were performed at each coverage level is confusing. Furthermore, if synthetic reads are derived from actual reads instead of directly from the genome, the reasoning behind this choice needs clarification.

The coverage values were a best approximation as NanoSim doesn’t allow for a specified coverage value as an input. Six replicates were used so we could easily differentiate between stochastic false positives and recurrent false positives. The reads and their mapping characteristics were used to generate a new read simulation model, which we then used to simulate reads directly from the genome. We have clarified these points in the text.

3) I don’t understand why Streptomyces eurocidicus and Enterococcus faecalis simulated reads had poor read mapping characteristics (line 844)? If the simulation process is consistent, why would these specific genomes result in lower quality simulated reads?

We used the reads and their mapping characteristics to generate a new read simulation model, so if the reads mapping characteristics were poor these characteristics would be inherited in the read simulation model. We have clarified this point in the text.

4) Figure S3 is challenging to comprehend. If possible, clarifying this figure would be beneficial.

Thank you for pointing this out, we have tried to make this figure more clear by modifying the figure titles and making the figure legend more descriptive.

5) The interpretation of the accuracy tests for PhaVa seems overly optimistic. In the main text it says “In general, the false positive rate was very low (mean false positive count per simulated sample of <0.1 in 9/10 species), with the exception of reads simulated from the Bordetella pertussis genome (Fig. 2D).” The text indicates a low false positive rate, yet the presence of significant false positives in 10% of the genomes tested suggests a higher rate. Excluding this genome from the accuracy evaluation doesn't seem fair.

In *B. pertussis* simulated genomes, we had a single recurrent false positive across replicates which increased their discovery rate. In addition to significantly expanding our benchmarking, we have modified the text significantly to make this point more clear. (Fig. S5, see response below)

6) Extending this simulation to a larger number of genomes would be advantageous for assessing the commonality of this issue.. It seems that the authors already have ~4,000 genomes on hand, and it doesn’t seem hard to generate synthetic long-reads from each of them to run this test. Rather than using this “NanoSim” method, can you not just simulate long-reads based solely on the genome? I believe doing so would still reveal the problem shown in Figure S3, as intended. Running on far more genomes like this would be really ideal to see how common this problem is. Alternatively, if the authors feel this is too much trouble, they could report the 10% false positive rate more transparently.

Thank you for the suggestion, we have greatly expanded the benchmarking using simulated reads from pbsim2 with a pre-trained read simulation model. Our new benchmarking (Fig. S5) tests the false positive rate and false negative rate in an additional 100 genomes (the original manuscript used only 10 genomes to benchmark). We also used this read simulation method in an attempt to address whether taxonomic distance affects the false negative rate (Fig. S5B), the requirement for reads to span invertons affects the false negative rate (Fig. S5D), and how coverage affects the false negative rate (Fig. S5C). We have also added a better description of both read simulation approaches to the methods.

The use of terms like “sample”, “genome”, “species”, “isolate”, and “dataset” in the text from lines 192 - 243 is confusing. For example, it seems sometimes “genome” refers to a set of reads, but sometimes it refers to the reference genome being mapped to. In the caption for Figure 3B, does genome mean “sequenced isolate”, or does genome mean “all read sets mapping a specific reference genome”? Precise definitions and consistent usage of these terms would aid comprehension.

We have made the language more consistent in the referenced text section, and have made clear in the figure (now Fig. 2B) and figure legend that 'genome' refers to all read sets mapping to a specific reference genome.

In the discussion it's stated that "PhaVa uses relatively strict mapping parameters and if the selected reference is distantly related to the sequenced strain, read mapping quality will decrease and reduce the discovery rate." This seems important. I think it would be ideal to directly test and report how similar the reads have to be to the genome for PhaVa to work properly.

Although there doesn't appear to be a strict cutoff on when a readset is too dissimilar, we have added a figure showcasing how decreasing the ANI between a simulated readset and a reference decreases the average coverage across potential invertons (Fig. S5B) which will decrease the sensitivity for detecting low frequency inversions with PhaVa.

Minor manuscript comments:

The introduction section is highly informative and well-composed.

Thank you for your feedback.

Is there any indication of how the requirement for long-reads to fully overlap the inverton affects false negatives?

Thank you for the suggestion, we have added Fig. S5D showing that when you lower the mean read length of a simulated long-read dataset, it does increase the false negative rate (particularly when your library mean read length <200bp), highlighting the importance of read length and quality long-read datasets when using PhaVa

Was PhaVa tested on the locked thiC strain to confirm no invertons at that location?

Thank you for this comment. To address this, we generated new long read DNA sequencing data on the locked *thiC* strains and then applied PhaVa to this dataset to detect invertons. In the locked forward strain, PhaVa did not detect any reads aligning in the reverse orientation. Whereas in the locked reverse strain, only reads aligning in the reverse orientation were detected. We have uploaded the long read data that supports this finding to SRA and will make those data publicly available with the rest of the newly generated data from this manuscript upon publication.

In Figure 2D, what constitutes a false positive? Is it a detected inverton or an inverted read mapping?

A false positive refers to a detected inverton. We have clarified this in the text.

Please specify the version of PhaVa used throughout the manuscript in the methods section.

We have clarified in the text that version 0.1.0 was used for analyses within the manuscript.

Comments on the PhaVa GitHub program:

Installation:

The installation instructions are insufficient, covering only the cloning of the repo without detailing program installation. Typically python programs have a “setup” script to install the program and put the “bin” scripts in the user’s path; I had to do that all myself.

We have added the package to bioconda and users can now install much easier with a simple ‘conda install phava -c bioconda -c conda-forge’

The manual doesn't mention the requirement for python 3.9 (due to the version of pysam used)

We have now listed this as a dependency in the github README.

Hosting the program on PyPi or conda would simplify installation for users.

In case it's helpful, a successful conda environment for the program can be created with the following command: “conda create -n phava emboss=6.5.7 minimap2=2.17 pysam=0.17.0 biopython=1.81”.

Thank you for the suggestion. We have added the package to bioconda and users can now install much easier with a simple ‘conda install -c bioconda phava’

General software engineering:

The test suite failed post-installation. Unclear if this is due to unaccounted for differences in my system vs. yours, or if the test just doesn’t work anymore.

Thank you for pointing this out. The test needed to be updated to reflect changes in the PhaVa output and we have updated it to account for these changes. In addition, we moved the test from a separate pytest script to an incorporated subcommand ‘test’ that you can run from the main PhaVa entry point

Versioning is good, but a changelog detailing version differences is needed.

We have added a changelog text file to the github repo with the most recent changes listed.

User manual:

Output files lack headers, making interpretation difficult. Headers should be added.

Thank you, we have added headers to the primary output files.

Documentation on the “filter” command, as indicated by pytest and the manuscript, is missing.

The filter command in the pytest is redundant with a newly added PhaVa --minRC option. We have added clearer description to the text as well as a description in the PhaVa help of what this option does. As mentioned above, we moved the installation test from a separate pytest script to an incorporated subcommand ‘test’ that you can run from the main PhaVa entry point

More detailed information about each output file is needed. The program makes a lot of files in the output folder, and it’s unclear which are important and which aren’t.

Thank you for your suggestion. To clarify the importance of different output files, we reorganized the output folder, adding a new subfolder (data) that contains the non-essential, but sometimes useful, outputs. In addition we added headers to help understand what is present in each output.

It’s not essential, but a tutorial would greatly aid in making the program more accessible, especially for less experienced users.

Thank you for your suggestion. We have added a short tutorial to the GitHub readme, showing a use case of PhaVa with three *B. theta* long-read datasets from different conditions.

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

Three of my prior comments have not been addressed. My other comments have been addressed.

1. The existence of potential variable proteins should be demonstrated at the protein level. If the intragenic inversions that don't introduce stop codons nevertheless lead to loss of protein stability, for example, no protein will be present. In this case, both inversions that introduce stop codons and inversions that don't create stop codons would be predicted to phenocopy gene deletion. Protein level confirmation would also help address the possibility that the reported novel sequences could be artifacts resulting from PCR chimeras or other steps in the sequencing approaches.

The question was about intragenic inversions that don't introduce stop codons. In the authors response, they focused on an intragenic inversion that does introduce a stop codon. There still remains no examples of intragenic inversions that don't introduce stop codons and result in a stable protein.

2. Many genetic rearrangements, transposon insertions, and indels lead to premature stop codons or other loss-of-function consequences. The novelty of this report would be increased by demonstrating that inversions that do NOT introduce stop codons don't simply cause loss of protein stability. No examples are provided of intragenic inversions that actually have distinct biochemical functions and phenotypes compared to both the non-inverted sequence and the deletion mutant.

This comment has not been addressed. The authors have not provided any examples of intragenic inversions that have distinct biochemical functions and phenotypes compared to both the non-inverted sequence and the deletion mutant. Instead, the authors provide additional computational predictions and hypotheses about potential functions of intragenic inversions. While these hypotheses are neither uniquely supported nor ruled out by the experiments, many alternate hypotheses are also reasonable.

6. It is unclear how (or whether) the inversions are catalyzed. Are there dedicated invertases, and what are the fitness and biochemical consequences of deleting these enzymes? This is a major aspect of the proposed biological mechanism that is not addressed.

This comment has not been addressed. It remains unclear how the reported inversions are mediated. The authors report experiments with negative results, which could have many explanations.

Referee #2 (Remarks to the Author):

The authors have been very responsive to reviewer comments, and the manuscript is substantially improved with higher breadth and quality of evidence. I have no further concerns.

Referee #3 (Remarks to the Author):

I appreciate the extensive work that has gone into improving the manuscript. The authors have substantially strengthened their manuscript with new experiments (to verify the alternative coding ThiC and to check whether inversion is affected by substrate availability), analysis (highlighting the potential for new functional domains expressed by inversions and the analysis of the long-read metagenomics data) and improved presentation throughout (new Figs and text). All my comments have been addressed, and the manuscript, in my opinion, is ready to be published.

Referee #4 (Remarks to the Author):

All my comments have been adequately addressed. In my view the manuscript is now ready for publication.

Referee #4 (Remarks on code availability):

Code is in useful and reproducible state.

Author Rebuttals to First Revision:

Referee #1 (Remarks to the Author):

Three of my prior comments have not been addressed. My other comments have been addressed.

We thank the reviewer for their comments and suggestions throughout this revision process. We will address their remaining concerns below.

1. The existence of potential variable proteins should be demonstrated at the protein level. If the intragenic inversions that don't introduce stop codons nevertheless lead to loss of protein stability, for example, no protein will be present. In this case, both inversions that introduce stop codons and inversions that don't create stop codons would be predicted to phenocopy gene deletion. Protein level confirmation would also help address the possibility that the reported novel sequences could be artifacts resulting from PCR chimeras or other steps in the sequencing approaches.

The question was about intragenic inversions that don't introduce stop codons. In the authors response, they focused on an intragenic inversion that does introduce a stop codon. There still remains no examples of intragenic inversions that don't introduce stop codons and result in a stable protein.

We had a slightly different understanding of the emphasis of the reviewer's comment from the original review. Therefore, instead of focusing solely on recoding intragenic invertons, we interpreted the comment to be enquiring about protein evidence of intragenic invertons, in general. As our report had focused on the *thiC* intragenic inverton, we had decided to focus our proteomic effort here. Therefore, as the reviewer points out, we have not molecularly characterized the proteomic consequences of a recoding intragenic inverton in this report. The reviewer's clarification of the focus of their question is helpful here, and we agree that the lack of molecular characterization of the proteomic consequences of a recoding intragenic inverton is a limitation of our current study. We have added additional

language to the discussion detailing this. We agree that this is an exciting avenue for future research and we are actively pursuing the characterization of additional intragenic invertons, including recoding intragenic invertons. We are excited to present that work in future studies.

2. Many genetic rearrangements, transposon insertions, and indels lead to premature stop codons or other loss-of-function consequences. The novelty of this report would be increased by demonstrating that inversions that do NOT introduce stop codons don't simply cause loss of protein stability. No examples are provided of intragenic inversions that actually have distinct biochemical functions and phenotypes compared to both the non-inverted sequence and the deletion mutant.

This comment has not been addressed. The authors have not provided any examples of intragenic inversions that have distinct biochemical functions and phenotypes compared to both the non-inverted sequence and the deletion mutant. Instead, the authors provide additional computational predictions and hypotheses about potential functions of intragenic inversions. While these hypotheses are neither uniquely supported nor ruled out by the experiments, many alternate hypotheses are also reasonable.

We agree with the reviewer that it would be very novel to provide biological characterization of a recoding intragenic inverton or an intragenic inverton that creates a distinct biochemical function or phenotype from a deletion mutant. We note that the current manuscript is already fairly extensive - and goes from identifying the phenomenon of intragenic invertons to deeply characterizing one, informatically characterizing several, and the developing, benchmarking and applying a new software tool to determine the prevalence of this new type of within-gene inverton across the prokaryotic tree of life.

The lack of an additional section that provides biological characterization of a recoding inverton is a current limitation of our work and we have emphasized this in the discussion section. Rigorous molecular characterization of an additional intragenic inverton would require substantial effort and thoughtful experimental design. We are currently undertaking this work and are excited to present them in future studies.

Additionally, our proteomic analysis of the ThiC intragenic inverton does demonstrate that inversion can create novel proteins (the truncated N terminal domain and an additional small protein - ED10). The detection of these protein products illustrate that inversion does not simply cause proteomic degradation. To be clear, we have not characterized these small proteins and their functional significance remains to be elucidated. We emphasize this in both the results and the discussion.

6. It is unclear how (or whether) the inversions are catalyzed. Are there dedicated invertases, and what are the fitness and biochemical consequences of deleting these enzymes? This is a major aspect of the proposed biological mechanism that is not addressed.

This comment has not been addressed. It remains unclear how the reported inversions are mediated. The authors report experiments with negative results, which could have many explanations.

We agree with the reviewer that identifying the mechanism (enzymatic or otherwise) that drives the inversion of these loci is important. Based on the characteristics of these loci (ie surrounded by short palindromic DNA repeat regions - ED9a,b, the ability to flip both forward and reverse - ED12) we suspect that they are indeed invertons and that flipping is performed by an enzyme (invertase). As we were unable to identify which invertase is responsible for flipping the *thiC* inverton in this study (Bth encodes 54 invertases), this is a limitation to our current study. We identify this limitation in the conclusion section of our paper. However, we are very excited to continue working on this aspect of this project and are currently developing and validating in vitro experimental assays to address this.

Referee #2 (Remarks to the Author):

The authors have been very responsive to reviewer comments, and the manuscript is substantially improved with higher breadth and quality of evidence. I have no further concerns.

We thank the reviewer for their comments and experimental suggestions as they have helped us substantially improve our manuscript.

Referee #3 (Remarks to the Author):

I appreciate the extensive work that has gone into improving the manuscript. The authors have substantially strengthened their manuscript with new experiments (to verify the alternative coding ThiC and to check whether inversion is affected by substrate availability), analysis (highlighting the potential for new functional domains expressed by inversions and the analysis of the long-read metagenomics data) and improved presentation throughout (new Figs and text). All my comments have been addressed, and the manuscript, in my opinion, is ready to be published.

We thank the reviewer for their input in this review process and their acknowledgement of the extensive work we have done highlighting the potential impact intragenic invertons may have on prokaryotic biology.

Referee #4 (Remarks to the Author):

All my comments have been adequately addressed. In my view the manuscript is now ready for publication.

We thank the reviewer for their comments on our manuscript.

Referee #4 (Remarks on code availability):

Code is in useful and reproducible state.

We also thank the reviewer for their feedback and suggestions on our tool. Their input substantially strengthened PhaVa's benchmarking and accessibility.