Trends in Genetics

CellPress

Spotlight

# Bacterial intragenic inversions: a new layer of diversity

Brantley Hall[1,2] and Xiaofang Jiang[3,*]

**DNA inversions in bacteria were known to create diversity through intergenic or partial intergenic changes. Now, Chanin, West, *et al*. reveal intragenic inversions, enabling single genes to encode multiple protein variants via sequence recoding or truncation – an unexpected mechanism for expanding protein diversity without increasing genome size.**

Bacteria have long been known to employ phase variation mechanisms to generate phenotypic heterogeneity within clonal populations. One such strategy is DNA inversion, a reversible process mediated by enzymes called invertases that flip regions of DNA [1]. These inversions often target regulatory elements or gene segments, allowing bacteria to switch between different states and equipping populations with diverse phenotypes to withstand environmental changes.

Early studies in the 1970s and 1980s identified DNA inversions in promoters controlling features such as flagellar antigens and fimbrial proteins [1]. Over subsequent decades, researchers uncovered numerous examples of phase-variable systems mediated by DNA inversions across bacterial pathogens and commensals. These systems regulate critical determinants of bacterial fitness and host interaction, such as virulence factors, capsular polysaccharides, and outer membrane proteins. The advent of computational tools has since accelerated our understanding of these genomic phenomena. PhaseFinder, introduced in 2019, identified 4686 invertible promoters across bacterial genomes that they called invertons [2]. Porter *et al*. demonstrated that phase-variable capsular polysaccharides in *Bacteroides thetaiotaomicron* mediate interactions with bacteriophages, underscoring the ecological significance of DNA inversions [3]. Milman *et al*. systematically identified over 11 000 gene-altering programmed inversions across 35 000 bacterial species, revealing their diverse targeting of genes beyond regulatory regions [4].

For decades, intragenic inversions were considered rare or even improbable due to the evolutionary constraints imposed by their flanking inverted repeats, which must preserve open reading frames. Disrupting a coding sequence through inversion could easily disrupt the open reading frame rendering the gene nonfunctional, likely explaining the scarcity of these events. However, the discovery of intragenic invertons by Chanin, West, *et al*. has reshaped this view, revealing unexpected genomic flexibility in bacterial genomes [5].

This breakthrough arose from a systematic analysis of metagenomic data from hematopoietic cell transplantation patients, whose gut microbes face selective pressures from chemotherapy, antibiotics, and nutrient fluctuations. Using PhaseFinder, the researchers not only confirmed previously known intergenic invertons but also identified novel intragenic invertons in *B. thetaiotaomicron*. Of 63 predicted intragenic invertons, ten were validated through PCR experiments, demonstrating their presence in both reference and inverted orientations within laboratory-grown populations. At the protein level, intragenic invertons can recode sequences, potentially altering protein functions or interactions, as seen in a 57-bp inversion in BT0375 that changes its amino acid sequence and likely affects DNA targeting or enzymatic kinetics. Alternatively, intragenic invertons may introduce premature stop codons, leading to truncated proteins with modified functionalities, such as the hybrid two-component system BT3786, where an inversion separates sensing and response domains, potentially disrupting signaling pathways.

Recognizing the limitations of short-read sequencing for resolving genomic structural variations, Chanin, West, *et al*. developed PhaVa, a computational tool optimized for long-read sequencing data. By leveraging the extended read lengths of long-read platforms, PhaVa maps reads to both forward and reverse orientations of potential invertons. Applying this tool to datasets from the NCBI Sequence Read Archive, the researchers identified ~4600 unique invertons, including intergenic, partial intergenic, and intragenic invertons. Of 372 intragenic invertons, 169 were predicted to recode proteins, enabling single genes to produce multiple protein variants. Examples include a recoding inverton named *slmA* from *Bordetella bronchiseptica*, altering its structure and function, and another in *barA* from *Aeromonas hydrophila*, which affects signal transmission. Notably, intragenic invertons were also found in *hsdS* genes of *Mycoplasma hominis*, demonstrating a novel mechanism for genetic variation.

To investigate the biological consequences of an intragenic inversion, Chanin, West, *et al*. focused on an inverton within the *B. thetaiotaomicron* BT0650 gene, which encodes the thiamine biosynthesis protein ThiC, essential for numerous cellular processes. Disruptions in thiamine biosynthesis can significantly affect microbial community dynamics, making this gene an ideal candidate for study. The team engineered 'locked' versions of *thiC*, fixing the inverton in either the reference or inverted orientation to assess its effects. Locked-reverse strains, mimicking a ThiC null mutant, required higher thiamine concentrations to achieve growth comparable

with wild-type strains, confirming the inversion's disruptive impact on protein function. These findings were further supported by quantitative proteomics, which revealed distinct protein expression patterns in the locked-reverse strains. Competitive growth experiments also demonstrated that under specific thiamine conditions, locked-reverse strains outcompeted locked-forward strains, indicating a potential adaptive advantage in certain environments. However, the study found no evidence that thiamine availability directly regulates *thiC* inversion, suggesting that the flipping occurs stochastically at low frequencies.

Intragenic invertons reveal a previously unrecognized mechanism of bacterial adaptability, enabling the generation of protein diversity and providing significant environmental advantages. These genetic elements showcase the remarkable plasticity of bacterial genomes, dynamically modifying protein function to support bacterial survival and evolution. The increasing adoption of long-read sequencing technologies has been pivotal in uncovering these features, with tools like PhaVa leveraging the extended read lengths to map complex structural variations with unprecedented accuracy. By facilitating the identification of intragenic invertons, these tools open new avenues for exploring the regulatory networks shaping microbial physiology, ecology, and pathogenesis.

## Acknowledgments

This research was supported in part by the Intramural Research Program of the National Library of Medicine at the National Institutes of Health.

## Declaration of interests

The authors declare no competing interests.

[1]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, College Park, MD, USA
[2]Center for Bioinformatics and Computational Biology, University of Maryland, College Park, College Park, MD, USA
[3]National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

*Correspondence:
xiaofang.jiang@nih.gov (X. Jiang).

https://doi.org/10.1016/j.tig.2024.12.002



## References

1. Van Der Woude, M.W. and Bäumler, A.J. (2004) Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* 17, 581–611
2. Jiang, X. *et al.* (2019) Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. *Science* 363, 181–187
3. Porter, N.T. *et al.* (2020) Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in *Bacteroides thetaiotaomicron*. *Nat. Microbiol.* 5, 1170–1181
4. Milman, O. *et al.* (2023) Systematic identification of gene-altering programmed inversions across the bacterial domain. *Nucleic Acids Res.* 51, 553–573
5. Chanin, R.B. *et al.* (2024) Intragenic DNA inversions expand bacterial coding capacity. *Nature* 634, 234–242