Marderstein *et al.* is an invaluable resource that can be used to examine the evolution of epigenetic, transcriptomic and cellular features during the progression of leukaemia in children with Down's syndrome. It will also help to further dissect the cell–cell communication between HSCs and their environment, and to broadly expand scientists' understanding of predisposition to leukaemia and other cancers.

Sébastien Malinge is at The Kids Research Institute Australia (Cancer Centre), Perth Children's Hospital, Nedlands, Western Australia 6009, Australia.
e-mail: sebastien.malinge@thekids.org.au

1. Marderstein, A. R. *et al. Nature* **634**, 104–112 (2024).
2. Antonarakis, S. E. *et al. Nature Rev. Dis. Primers* **6**, 9 (2020).
3. Chou, S. T. *et al. Blood* **112**, 4503–4506 (2008).
4. Roy, A. *et al. Proc. Natl Acad. Sci. USA* **109**, 17579–17584 (2012).
5. Jardine, L. *et al. Nature* **598**, 327–331 (2021).
6. Wechsler, J. *et al. Nature Genet.* **32**, 148–152 (2002).
7. Arkoun, B. *et al. J. Clin. Invest.* **132**, e156290 (2022).
8. Wagenblast, E. *et al. Science* **373**, eabf6202 (2021).
9. Vuckovic, D. *et al. Cell* **182**, 1214–1231 (2020).

The author declares no competing interests.
This article was published online on 25 September 2024.

Microbiology

Bacteria invert gene sequences to flip the script

Chia-Chi Chang & Robert R. Jenq

If a cell can diversify its own genome, this can be advantageous in the face of changing environmental conditions. Bacteria have been found that can alter encoded proteins by using the trick of sequence inversion inside genes. **See p.234**

Random mutation is a powerful approach used by many organisms, including bacteria, to generate genomic changes that can lead to improved success (fitness) when encountering altered conditions in their environment. However, drawbacks to relying on random mutation include the relatively slow rate of genetic change in rare organisms, as well as a lack of genetic specificity. An alternative strategy of rapidly producing changes at specific genomic sites is called phase variation and this can give rise to variation in the characteristics of cells (phenotypic heterogeneity). DNA-sequence inversion is one form of phase variation, in which DNA-binding proteins and enzymes called recombinases can reversibly flip the orientation of a region of the genome, called an inverton, that is flanked by protein-binding sites[1].

On page 234, Chanin *et al.*[2] reveal that bacteria frequently invert DNA sequences inside genes, a process termed intragenic inversion. This type of inversion had not been observed previously in bacteria. The only prior report of intragenic inversion to our knowledge is a much shorter inversion (just seven nucleotides) described in human DNA from mitochondrial organelles[3].

Examples of previously described invertons include sequences that drive gene-expression, called promoters, being flipped so that the promoter drives expression of two differing proteins[4]. Alternatively, invertons can be used to swap between two downstream sequences of the promoter while keeping a 'constant' initial sequence that encodes a fixed portion of the gene product.

So far, the large-scale identification of invertons has been hampered by the limitations of a type of DNA sequencing called short-read sequencing, which requires computational analysis (imputation) to generate full-length bacterial genome maps. Chanin and colleagues introduce an analysis tool called PhaVa that takes advantage of improved sequencing technologies (long-read sequencing) to identify previously unknown potential phase variants. Because the long stretches of continuous DNA obtained in long-read sequencing can encompass an entire inverted region of DNA, identification of invertons is made with high accuracy, as the authors demonstrate. Interestingly, long-read-based identification was also found to be more sensitive than short-read methods, detecting around tenfold more invertons per genome. Bacterial subsets rich in invertons include those from the groupings Bacteroidetes, Fusobacteria, Gammaproteobacteria and Verrucomicrobia.

Chanin and colleagues also identify a previously unknown class of invertons — intragenic invertons. Part of the 'forward' sequence (corresponding to the reference sequence for the encoded protein) is flipped to swap in the 'reverse' inverted sequence (Fig. 1). Some of these invertons enable the production of two versions of a protein of identical length; others can generate a shorter version of the protein by introducing a sequence that ends protein production (a stop codon). On occasion, these types of inverton can disconnect the sensing and response functions of certain proteins, such as the inverton that affects the protein BarA of the bacterium *Aeromonas hydrophila*, which recodes part of a protein domain and might alter signalling by the protein.

Intragenic invertons might lead to changes in protein function, structure and downstream signalling, but more experiments will be required to fully understand such effects, especially in the context of varying growth environments. The authors note that unlike intergenic invertons, which occur in non-coding DNA sequences between genes, and tend to be universally in one of the two orientations in a sample, intragenic inversions are commonly present in both forward and inverted (reverse) orientations in a given sample.

The authors also compared invertons identified in genomic sequences from mixtures of bacterial species (metagenomes) in samples from human faeces with those derived from *in vitro* cultivated isolates of the bacterium strain *Bacteroides thetaiotaomicron* VPI-5482. They observed that invertons in bacteria from human samples were more numerous and more likely to be heterogeneous than those from the *in vitro* sample — indicating flipping activity. Invertons in bacteria from human samples were more often than not in the reverse orientation relative to the reference genome sequences, which are generated from *in vitro* isolates. This suggests that reference genomes might be a poor representation of *in vivo* bacterial genomes, at least from the perspective of inverton status.

The authors took a deep dive into a particular intragenic inverton that they identified in a gene from *B. thetaiotaomicron* that encodes the protein ThiC, which is needed to produce the vitamin thiamine. The inverton results in the introduction of an early stop codon in the flipped, non-reference orientation of the sequence. Thiamine is a cofactor for several aspects of cellular metabolism. The biosynthesis of thiamine requires many enzymes, including ThiC, encoded by the gene *thiC*.

The authors investigated the biological effects of the inversion of *thiC* sequences, speculating that the flipped non-reference orientation would produce a mutant type of bacterium that was dependent on acquiring thiamine. The authors generated strains with 'locked' forward or inverted versions of the *thiC* inverton, as well as a *thiC* deletion mutant, and evaluated growth when exposed to a variety of thiamine concentrations in the growth medium. As predicted, the authors found that the locked-forward strain (corresponding to the reference genome) had

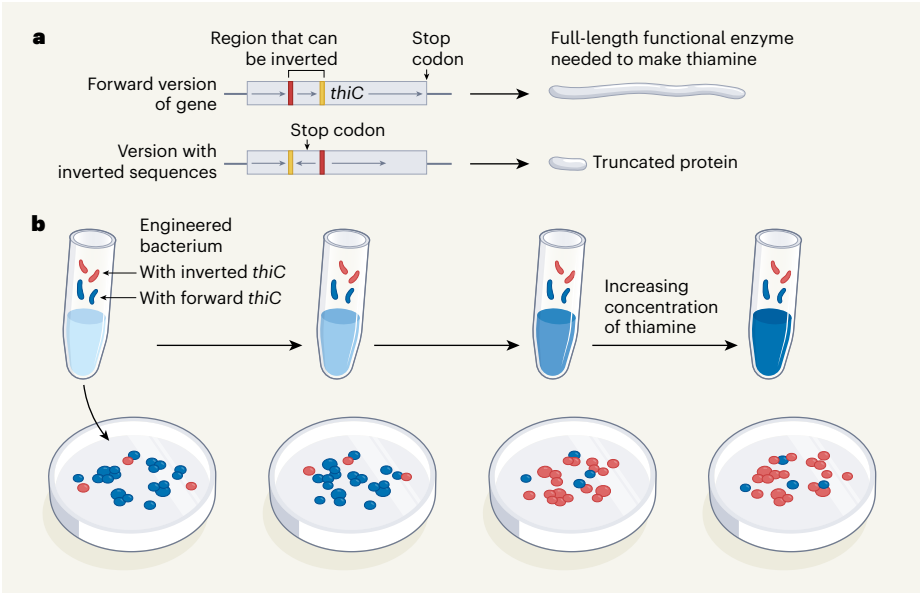

Figure 1 | Bacteria can invert sequences inside a gene. a, Chanin *et al.*[2] reveal that bacteria have the gene-editing capacity to flip sequences back to front, which can generate two different proteins from a gene. For example, the authors report that the bacterium *Bacteroides thetaiotaomicron* has a gene *thiC* with sequences that can invert (indicated here as the sequences between the red and yellow lines). This gene is needed to make the vitamin thiamine. If the gene has inverted sequences, this results in the formation of an abnormally shortened version of the protein — this truncation occurs because the inverted region of the gene contains a sequence (stop codon) that terminates the sequence being used for protein coding. **b**, The authors engineered bacteria that were 'locked' to either express the forward or the inverted version of *thiC* and examined the survival of the two populations of bacteria when grown at different concentrations of thiamine. Under conditions of low thiamine, bacteria that expressed the forward version of *thiC* dominated; when thiamine was high, the bacteria that expressed the inverted version of *thiC* had a survival advantage.

thiamine-independent growth, similar to the wild-type strain. By contrast, the locked-inverted strain had similar characteristics to the *thiC* deletion mutant and had thiamine-dependent growth.

The authors were able to demonstrate that production of the shorter version of ThiC and ThiC-inverton-derived peptides were unique to the locked-inverted bacterial strain. To evaluate the relative fitness of locked-forward and -inverted strains, the authors performed *in vitro* competitive growth experiments in varying thiamine concentrations (Fig. 1).

The locked-inverted strain had a substantial advantage over the locked-forward strain in conditions of between 0.01 micromolar and 10 µM concentrations of thiamine. The human intestine has concentrations of 0.02–2 µM of thiamine, indicating that inversion might enable the bacterium to quickly adapt to the varying thiamine concentrations it encounters *in vivo*, similar to a phenomenon described previously[5]. Thiamine concentrations, however, do not seem to affect expression of the invertase enzymes that are needed for inversion or the rates of *thiC* inversion. Future studies might help to identify mechanisms by which invertases are regulated.

Phase variation by DNA inversion is a fascinating mechanism that might offer options for adaptation to new environments. By developing PhaVa to capitalize on long-read sequencing technologies and apply them to this phenomenon, Chanin and colleagues have provided scientists with an important tool to better understand the ways in which bacteria and other organisms might adjust their genomic programming to quickly adapt to and survive in varying conditions.

Chia-Chi Chang is in the Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. **Robert R. Jenq** is in the Department of Hematology and Hematopoietic Cell Transplantation, City of Hope National Medical Center, Duarte, California 91010, USA.
e-mails: cchang2@mdanderson.org; rjenq@coh.org

1. Jiang, X. *et al. Science* **363**, 181–187 (2019).
2. Chanin, R. B. *et al. Nature* **634**, 234–242 (2024).
3. Musumeci, O. *et al. Am. J. Hum. Genet.* **66**, 1900–1904 (2000).
4. Zieg, J., Silverman, M., Hilmen, M. & Simon, M. *Science* **196**, 170–172 (1977).
5. D'Souza, G. *et al. Evolution* **68**, 2559–2570 (2014).

R.R.J. declares competing interests; see go.nature.com/4e2k52p for details.
This article was published online on 25 September 2024.