Title: A historical sequence deletion in a commonly used *Bacillus subtilis* chromosome integration vector generates undetected loss-of-function mutations

Author list

K. Julia Dierksheide[a], Gene-Wei Li[a]*

[a] *Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, United States*

*Email: gwli@mit.edu

Keywords: *Bacillus subtilis,* integration vector, *amyE,* fermentation, pDG1661

Abstract

Since the 1980s, chromosome-integration vectors have been used as a core method of engineering *Bacillus subtilis.* One of the most frequently used vector backbones contains chromosomally derived regions that direct homologous recombination into the *amyE* locus. Here, we report a gap in the homology region inherited from the original *amyE* integration vector, leading to erroneous recombination in a subset of transformants and a loss-of-function mutation in the downstream gene. Internal to the homology arm that spans the 3′ portion of *amyE* and the downstream gene *ldh*, an unintentional 227-bp deletion generates two crossover events. The major event yields the intended genotype, but the minor event, occurring in ~10% of colonies, results in a truncation of *ldh*, which encodes lactate dehydrogenase. Although both types of colonies test positive for *amyE* disruption by starch plating, the potential defect in fermentative metabolism may be left undetected and confound the results of subsequent experiments.

Main text

The model Gram-positive bacterium *Bacillus subtilis* is widely used for strain engineering due to its natural competence and efficient homologous recombination system[1,2]. Synthetic DNA is commonly introduced into specific loci of the genome via homology-containing integration vectors that can be constructed and manipulated as plasmids in *Escherichia coli* (Figure 1A). One of the first genomic loci developed for integration vectors is at the gene *amyE,* which encodes α-amylase, a protein involved in starch degradation[3,4]. Successful integration leads to disruption of *amyE*, which can be easily screened for using an iodine stain that changes coloration upon binding to starch ("starch test")[2]. The original *amyE* double-crossover integration vector pBGtrp and its derivatives, such as pDR111 and pDG1661[5,6], have enabled studies on many aspects of microbiology, ranging from gene regulation to cell division[7–10]. They have also been central to the development of synthetic biology toolkits for *B. subtilis*[11–15].

However, we found that the homology regions in these commonly used *amyE* integration vectors are inconsistent with the genome sequence[16]. In the sequence of pDR111, the annotated *amyE*-back homology region is followed by an additional 153-bp sequence derived from a region of the genome 227 bp downstream of *amyE*-back (Figure 1B; Supplementary material). The resulting extended homology region includes a gap that belongs to the downstream *ldh* gene and its ribosome binding site. Due to this

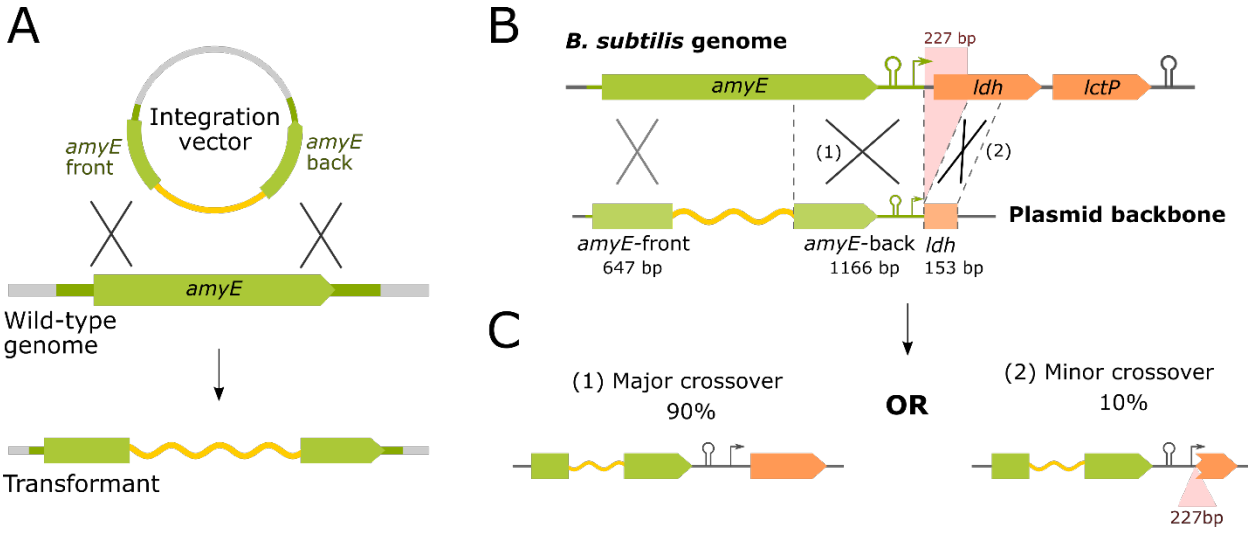

Figure 1: Double-crossover events at *amyE.* (A) Schematic of an *amyE* integration vector (top) designed to direct integration of the insert (yellow) into the genome as shown in the transformant genome (bottom). On the integration vector, the insert is flanked by two homology arms, *amyE*-front and *amyE*-back (green). (B) Schematic of the missing homology region. In the *B. subtilis* genome, *amyE* is followed by the *ldh-lctP* operon (top). In pBGtrp and its derivatives, the annotated *amyE*-back region is followed by a 153-bp fragment of *ldh*, while missing the intervening 227-bp sequence (bottom). (C) The two possible double-crossover events. In both cases, crossover occurs as expected at the upstream *amyE*-front region, but the missing genome sequence in the plasmid allows for two possible recombination events at the downstream *amyE*-back region. The minor event results in loss of 227 bp of genomic sequence containing the ribosome binding site and the first 215 nucleotides of *ldh.*

discontinuity in the homology region, in addition to the expected crossover at *amyE*-back, crossover can occur at the 153-bp region on the plasmid, disrupting *ldh,* a gene that codes for lactate dehydrogenase (LDH)[17] (Figure 1C). By colony PCR, we found that 4 of the 36 colonies tested after transformation with a derivative of pDR111 were missing the 227-bp region, indicating that the secondary crossover event occurs in a substantial proportion of transformants (Supplementary Table 1).

The discontinuous *amyE*-back homology region in pDR111 was inherited from pBGtrp, the original *amyE* double crossover integration vector developed in 1986[3,5,6,18,19]. The pBGtrp homology arms were generated from subclones of the *B. subtilis amyE* gene that were used to sequence the gene in 1983. We found that the corresponding sequence deposited in GenBank is missing the same 227 bp, indicating that this region was likely lost in the process of preparing *amyE* for sequencing in *E. coli*[4]. In addition to pDR111, many *amyE* double crossover integration vectors developed over the past 40 years, including pDG1661, likely have inherited the same discontinuous homology arms from pBGtrp and its derivative vectors.

To facilitate correction of this error in future work, we constructed modified plasmids of pDR110 and pDR111 where the 153-bp region downstream of *amyE*-back has been removed. The removal of the *ldh* homology region did not substantially impact transformation efficiency, and all colonies tested (18 of 18) integrated at *amyE* as expected for both plasmids. These plasmids are available on AddGene (www.addgene.org) as pGL003 (modified pDR110) and pGL004 (modified pDR111).

Historically, a single *B. subtilis* colony that tests positive by the starch test is carried forward after transformation for subsequent experiments. Our results suggest that, across all strains constructed with pBGtrp and its derivatives, ten percent of the strains may be missing the ribosome binding site and a major portion of LDH. Given LDH's role in fermentative metabolism and anaerobic growth[20], an undetected crossover in *ldh* may have influenced the results of previous experiments performed in these conditions. Furthermore, even in aerobic growth, LDH plays a role in re-utilizing lactate that is excreted as a by-product of overflow metabolism[21]. During aerobic growth in LB, addition of supplemental glucose induces *ldh* expression, indicating that loss of LDH function may also affect experiments performed in the presence of oxygen[22].

This discrepancy can also influence studies with large-scale libraries of strains – whether pooled or arrayed – at the *amyE* site. Libraries of *B. subtilis* cells with pooled CRISPRi, overexpression, or reporter variants are powerful tools for discovery when coupled to modern high-throughput assays. When generating a library of *B. subtilis* variants, all cells that carry the intended antibiotic resistance cassette are carried forward from one or multiple transformation reactions. If the current, discontinuous *amyE* homology region is used, each transformed variant will integrate at *amyE* through one of the two possible crossover events (Figure 1C). These distinct crossover events are challenging to distinguish in high-

throughput and introduce additional heterogeneity that could confound the results. Therefore, to ensure properly controlled experiments, especially in the context of fermentative *B. subtilis* studies, it will be important to correct the integration arms in future work.

Acknowledgements

We thank A. D. Grossman's laboratory for providing plasmids and members of the G.-W.L. and A.D.G. laboratories for discussions. This research was supported by NIH R35GM124732 and the NSF CAREER Award MCB-1844668.

Conflict of interest

The authors declare no conflicts of interest.

Funding information

NIH R35GM124732 and NSF CAREER Award MCB-1844668

References

(1) Harwood, C. R. *Bacillus subtilis* and Its Relatives: Molecular Biological and Industrial Workhorses. *Trends Biotechnol.* **1992**, *10*, 247–256. https://doi.org/10.1016/0167-7799(92)90233-L.

(2) Wozniak, K. J.; Simmons, L. A. Genome Editing Methods for *Bacillus subtilis*. *Methods Mol. Biol. Clifton NJ* **2022**, *2479*, 159–174. https://doi.org/10.1007/978-1-0716-2233-9_11.

(3) Hidenori, S.; Henner, D. J. Construction of a Single-Copy Integration Vector and Its Use in Analysis of Regulation of the Trp Operon of *Bacillus subtilis*. *Gene* **1986**, *43* (1), 85–94. https://doi.org/10.1016/0378-1119(86)90011-9.

(4) Yang, M.; Galizzi, A.; Henner, D. Nucleotide Sequence of the Amylase Gene from *Bacillus subtilis*. *Nucleic Acids Res.* **1983**, *11* (2), 237–249. https://doi.org/10.1093/nar/11.2.237.

(5) Britton, R. A.; Eichenberger, P.; Gonzalez-Pastor, J. E.; Fawcett, P.; Monson, R.; Losick, R.; Grossman, A. D. Genome-Wide Analysis of the Stationary-Phase Sigma Factor (Sigma-H) Regulon of *Bacillus subtilis*. *J. Bacteriol.* **2002**, *184* (17), 4881–4890. https://doi.org/10.1128/jb.184.17.4881-4890.2002.

(6) Guérout-Fleury, A. M.; Frandsen, N.; Stragier, P. Plasmids for Ectopic Integration in *Bacillus subtilis*. *Gene* **1996**, *180* (1–2), 57–61. https://doi.org/10.1016/s0378-1119(96)00404-0.

(7) Kuhlmann, N. J.; Chien, P. Selective Adaptor Dependent Protein Degradation in Bacteria. *Curr. Opin. Microbiol.* **2017**, *36*, 118–127. https://doi.org/10.1016/j.mib.2017.03.013.

(8) Kalamara, M.; Spacapan, M.; Mandic-Mulec, I.; Stanley-Wall, N. R. Social Behaviours by *Bacillus subtilis*: Quorum Sensing, Kin Discrimination and Beyond. *Mol. Microbiol.* **2018**, *110* (6), 863–878. https://doi.org/10.1111/mmi.14127.

(9) Cameron, T. A.; Margolin, W. Insights into the Assembly and Regulation of the Bacterial Divisome. *Nat. Rev. Microbiol.* **2024**, *22* (1), 33–45. https://doi.org/10.1038/s41579-023-00942-x.

(10) Kavita, K.; Breaker, R. R. Discovering Riboswitches: The Past and the Future. *Trends Biochem. Sci.* **2023**, *48* (2), 119–141. https://doi.org/10.1016/j.tibs.2022.08.009.

(11) Guiziou, S.; Sauveplane, V.; Chang, H.-J.; Clerté, C.; Declerck, N.; Jules, M.; Bonnet, J. A Part Toolbox to Tune Genetic Expression in *Bacillus subtilis*. *Nucleic Acids Res.* **2016**, *44* (15), 7495–7508. https://doi.org/10.1093/nar/gkw624.

(12) Radeck, J.; Kraft, K.; Bartels, J.; Cikovic, T.; Dürr, F.; Emenegger, J.; Kelterborn, S.; Sauer, C.; Fritz, G.; Gebhard, S.; Mascher, T. The Bacillus BioBrick Box: Generation and Evaluation of Essential Genetic Building Blocks for Standardized Work with *Bacillus subtilis*. *J. Biol. Eng.* **2013**, *7* (1), 29. https://doi.org/10.1186/1754-1611-7-29.

(13) Popp, P. F.; Dotzler, M.; Radeck, J.; Bartels, J.; Mascher, T. The Bacillus BioBrick Box 2.0: Expanding the Genetic Toolbox for the Standardized Work with *Bacillus subtilis*. *Sci. Rep.* **2017**, *7* (1), 15058. https://doi.org/10.1038/s41598-017-15107-z.

(14) Fu, G.; Yue, J.; Li, D.; Li, Y.; Lee, S. Y.; Zhang, D. An Operator-Based Expression Toolkit for *Bacillus subtilis* Enables Fine-Tuning of Gene Expression and Biosynthetic Pathway Regulation. *Proc. Natl. Acad. Sci.* **2022**, *119* (11), e2119980119. https://doi.org/10.1073/pnas.2119980119.

(15) Song, Y.; Nikoloff, J. M.; Fu, G.; Chen, J.; Li, Q.; Xie, N.; Zheng, P.; Sun, J.; Zhang, D. Promoter Screening from *Bacillus subtilis* in Various Conditions Hunting for Synthetic Biology and Industrial Applications. *PLOS ONE* **2016**, *11* (7), e0158447. https://doi.org/10.1371/journal.pone.0158447.

(16) Borriss, R.; Danchin, A.; Harwood, C. R.; Médigue, C.; Rocha, E. P. C.; Sekowska, A.; Vallenet, D. *Bacillus subtilis*, the Model Gram-Positive Bacterium: 20 Years of Annotation Refinement. *Microb. Biotechnol.* **2018**, *11* (1), 3–17. https://doi.org/10.1111/1751-7915.13043.

(17) Pedreira, T.; Elfmann, C.; Stülke, J. The Current State of SubtiWiki, the Database for the Model Organism *Bacillus subtilis*. *Nucleic Acids Res.* **2022**, *50* (D1), D875–D882. https://doi.org/10.1093/nar/gkab943.

(18) Perego, M. Integrational Vectors for Genetic Manipulation in *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-Positive Bacteria*; John Wiley & Sons, Ltd, 1993; pp 615–624. https://doi.org/10.1128/9781555818388.ch42.

(19) Antoniewski, C.; Savelli, B.; Stragier, P. The spoIIJ Gene, Which Regulates Early Developmental Steps in *Bacillus subtilis*, Belongs to a Class of Environmentally Responsive Genes. *J. Bacteriol.* **1990**, *172* (1), 86–93. https://doi.org/10.1128/jb.172.1.86-93.1990.

(20) Romero, S.; Merino, E.; Bolívar, F.; Gosset, G.; Martinez, A. Metabolic Engineering of *Bacillus subtilis* for Ethanol Production: Lactate Dehydrogenase Plays a Key Role in Fermentative Metabolism. *Appl. Environ. Microbiol.* **2007**, *73* (16), 5190–5198. https://doi.org/10.1128/AEM.00625-07.

(21) Sonenshein, A. L. Control of Key Metabolic Intersections in *Bacillus subtilis*. *Nat. Rev. Microbiol.* **2007**, *5* (12), 917–927. https://doi.org/10.1038/nrmicro1772.

(22) Lorca, G. L.; Chung, Y. J.; Barabote, R. D.; Weyler, W.; Schilling, C. H.; Saier, M. H. Catabolite Repression and Activation in *Bacillus subtilis*: Dependency on CcpA, HPr, and HprK. *J. Bacteriol.* **2005**, *187* (22), 7826–7839. https://doi.org/10.1128/jb.187.22.7826-7839.2005.