Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A compact engineered TnpB editor enables efficient genome engineering in non-model industrial microorganisms

Nan Pen

nanp@mail.hzau.edu.cn

Huazhong Agricultural University https://orcid.org/0000-0002-1637-6729

Qiujin Guo

Huazhong agricultural university

Youyou Wang

Huazhong agricultural university

Binyang Xiong

Huazhong agricultural university

Boya Xu

Huazhong agricultural university

Jie Yang

Huazhong agricultural university

Yu Zhang

Huazhong agricultural university

Ziyan Fan

Huazhong agricultural university

Weike Ren

Huazhong agricultural university

Guanhua Ling

Huazhong agricultural university

Jilin Liu

Huazhong Agricultural University

Qing Ye

Huazhong Agricultural University

Biological Sciences - Article

Keywords: TnpB, DNA unwinding, Genome editing, Saccharomyces cerevisiae, Non-model industrial microorganisms

Posted Date: June 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9485642/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. All authors declare that they have patents pending to this material.

Abstract

Transposon-associated TnpB nucleases are promising compact alternatives to CRISPR-Cas nucleases, but their application to non-model industrial microorganisms remains limited. Here, we identified ISLde6 from Lactobacillaceae as a compact TnpB scaffold with cross-host activity and flexible TAM recognition capacity. Structure-guided engineering of this scaffold generated a G9F/E350R mutant with improved editing efficiency. Biochemical analyses suggested that, under low Mg^{2+} conditions, these mutations enhance cleavage activity by alleviating a pre-catalytic barrier associated with target-DNA unwinding. The combination of these mutations with functional domain fusion yielded FIRE-TnpB, a compact editor that enabled one-step integration of a 10.5-kb biosynthetic pathway into industrial yeast and efficient genome editing across multiple non-model industrial microorganisms. Collectively, these findings establish FIRE-TnpB as a compact and efficient genome-editing platform, while also offering mechanistic insight for the engineering of compact RNA-guided nucleases.

Introduction

The CRISPR-Cas system has revolutionized genome engineering by enabling precise DNA manipulation across diverse organisms [1]. Widely used effectors such as SpCas9 and AsCas12a exhibit robust activity and broad programmability, but their large size (~ 1,000–1,400 amino acids) restricts delivery and application [2], particularly to genetically intractable non-model microorganisms related to synthetic biology and biotechnology [3–5].

In these microorganisms, restrictive host defense systems (such as restriction–modification systems) and interference from endogenous plasmids lower exogenous plasmid transformation efficiency, thereby limiting genome editing [6, 7]. In addition, the application of endogenous CRISPR systems to genome engineering is limited in many industrially relevant species due to the lack of these functional systems or the difficulty in repurposing [8, 9]. Compact genome-editing tools can reduce vector loading burden and relieve delivery constraints in these hosts, hence contributing to strain engineering. This highlights the necessity for developing compact and broadly applicable genome-editing nucleases.

These technical barriers create a pressing need for compact alternatives, motivating exploration of transposon-encoded TnpB proteins within the OMEGA systems, which are only about one-third the size of Cas9 or Cas12 nucleases [10, 11]. Phylogenetic analyses have revealed extensive diversity of the TnpB family and clarified their evolutionary relationship to Cas12 effectors [12, 13]. Importantly, functional characterization of individual TnpB orthologs has demonstrated that this TnpB family exhibits various desirable properties, such as enhanced thermostability [14] and flexible target-adjacent motif (TAM) recognition capacity [15]. Together, these findings suggest that unexplored microbial lineages may harbor novel TnpB with useful properties that can be used to develop efficient and versatile genome-editing tools.

However, despite the identification of various naturally occurring TnpB, their performance in eukaryotic contexts remains variable, prompting substantial efforts to engineer TnpB nucleases for improved functionality. For example, structure-guided engineering has expanded the target-adjacent motif (TAM) recognition of ISDra2 TnpB from 5'-TTGAT to 5'-TYKAT [16]. In addition, ωRNA engineering [17, 18] and mutational scanning–guided TnpB engineering [19] have been employed to enhance the editing efficiency of ISDra2 TnpB. These studies jointly highlight the engineerability of TnpB system. However, compact editors with broad applicability in non-model industrial microorganisms remain underexplored [20]. Identifying and engineering naturally favorable new TnpB scaffolds may accelerate the development of robust genome-editing platforms for genetically intractable microorganism.

In this context, phylogeny-guided mining of Lactobacillaceae, a lineage enriched in TnpB diversity, identified ISLde6 TnpB as a promising editing scaffold with relatively flexible TAM recognition ability and cross-host genome editing activity. Through structure-guided engineering, we obtained mutants with improved editing efficiency. Mechanistic analysis further suggested that these mutations may enhance editing activity by alleviating a pre-catalytic barrier related to target-DNA unwinding under intracellular low-Mg^{2+} conditions. We further fused the optimized mutant with functional domains to generate FIRE-TnpB, which enables one-step integration of a large biosynthetic pathway in industrial yeast and supports efficient genome editing across multiple non-model industrial microorganisms.

Result

ISLde6 obtained in phylogeny-guided mining as a cross-host active TnpB scaffold

To systematically explore TnpB diversity, we obtained representative bacterial genomes from the NCBI database and constructed a phylogenetic framework based on 16S rRNA sequences (Fig. 1a). Mining of TnpB proteins revealed that Enterobacteriaceae, Bacillaceae, Lactobacillaceae, Peptostreptococcaceae, and Burkholderiaceae are top 5 harboring the largest number of TnpB-encoding genomes (Fig. 1a; Supplementary Fig. 1a). Among them, Bacillaceae and Lactobacillaceae exhibited the highest TnpB sequence diversity, as revealed by clustering at 70% sequence identity and coverage (Fig. 1a; Supplementary Fig. 1b).

Given the high sequence diversity and the high abundance of multicopy TnpB-encoding genomes in Lactobacillaceae, this family was selected for further mining. We obtained 65 non-redundant TnpB candidates from 721 genomes with ≥ 3 TnpB copies by clustering (Fig. 1b). Further, twelve representative candidates were selected from these 65 candidates based on phylogenetic distinctness and/or conserved catalytic residues for subsequent functional screening (Fig. 1c; Supplementary Fig. 1c, 2a).

To evaluate activity across hosts, we established a dual-host screening pipeline in both prokaryotic and eukaryotic systems. In *Limosilactobacillus reuteri* 03, 7 out of 12 TnpB candidates displayed significantly lower transformation efficiency in targeting groups than in non-targeting controls, suggesting that these

TnpB candidates possess DNA cleavage activity (Fig. 2a and Fig. 2b). In *Saccharomyces cerevisiae* S288c, 4 of these 7 candidates exhibited detectable *ADE2* deletion activity, with ISLde6 exhibiting the highest editing efficiency and the broadest target compatibility (Fig. 2c).

Among all tested candidates, ISLde6 TnpB consistently exhibited the strongest and most transferable activity in both prokaryotic and eukaryotic systems, establishing it as a promising scaffold for further engineering.

Identified TAM recognition flexibility of ISLde6 TnpB

To investigate the properties distinguishing ISLde6 from other TnpBs in the dual-host screening, we next characterized its cleavage properties, target requirements, and targeting scope, with ISLam TnpB analyzed in parallel as an active TnpB with distinct TAM specificity and protein size. In vitro assays showed that both ISLde6 and ISLam TnpBs required guide-target complementarity and appropriate TAM for efficient DNA cleavage (Fig. 3a; Supplementary Fig. 3a, b). Sanger sequencing further showed that both TnpBs cleaved target DNA region distal to the TAM, generating 5’ overhangs through asymmetric cleavage (Fig. 3b), consistent with the cleavage pattern reported for AsCas12f1 [21]. ISLde6 TnpB exhibited 1.2-fold to 88-fold higher cleavage efficiency on non-target strand (NTS) than on target strand (TS) at different reaction time points (Supplementary Fig. 3a).

Mismatch tolerance assays of ISLde6 and ISLam TnpBs localized their seed regions to positions 1–14 proximal to the TAM, where mismatches caused 87–100% loss of activity in the seed region, compared with ≤ 27% loss at distal positions (17–20 nt) (Fig. 3c) Further characterization of both TnpBs showed that their optimal cleavage activity occurred at 37°C within the tested temperature range (20–42°C) (Fig. 3d). At this temperature, both Mg^{2+} and Ca^{2+} supported cleavage activity (Supplementary Fig. 3c and d). Genome interference assays identified optimal guide RNA length ranges of 18–26 nt for ISLde6-mediated cleavage and 18–30 nt for ISLam-mediated cleavage (Fig. 3e). Based on these results, a guide length of 20 nt was selected for subsequent experiments.

Bioinformatic analysis predicted that ISLde6 and ISLam TnpBs recognized 5'-TTGAT-3' and 5'-TTTAA-3' TAM sequences, respectively (Supplementary Fig. 2b). TAM mutagenesis revealed that the two TnpB systems exhibited markedly different recognition profiles. ISLde6 TnpB could flexibly recognize 5'-NNGVB-3' TAM (V = A, C or G; B = T, C or G), whereas ISLam TnpB merely recognized 5'-TTTAA-3' TAM (Fig. 3f). To further dissect TAM recognition capacity of ISLde6 TnpB, we constructed a dsDNA library containing all 64 possible 5'-NNN-3’ TAM sequences to determine its TAM preference, and found that ISLde6 TnpB retained cleavage activity across multiple alternative TAM (such as 5’-GAT/GTT/GCT/GAG/GGG/TAT-3’) (Fig. 3g).

Taken together, these results indicate that ISLde6 TnpB has a well-defined cleavage mechanism with flexible TAM recognition, making it a strong candidate for being engineered into a broadly applicable editor.

Improvement of eukaryotic activity of ISLde6 TnpB by rational engineering Despite its favorable properties, ISLde6 TnpB showed limited editing efficiency in eukaryotic cells, prompting further engineering. We next applied structure-guided rational engineering to improve its editing performance. The ISLde6 structure was predicted by AlphaFold3 and then aligned with the resolved ISDra2 structure, revealing high structural similarity (TM-score [22] = 0.785). This alignment enabled domain assignment of ISLde6 based on the resolved ISDra2 structure [23] (Fig. 4a; Supplementary Fig. 4a, b). We then generated AlphaFold3-predicted ternary complex structure of ISLde6 - ωRNA - dsDNA and analyzed the ISLde6-DNA and ISLde6-RNA interfaces to identify candidate residues for further optimization (Fig. 4b). Based on this analysis, we identified regions to be optimized and designed 62 mutants to improve ISLde6 - RNA and ISLde6 - DNA interactions, steric complementarity, and ISLde6 conformational stability required for catalysis.

We performed high-throughput screening in *S. cerevisiae* S288C and identified G9F in the REC domain and E350R in the RuvC domain as the most favorable substitutions, which increased *ADE2* deletion efficiency at the *ADE2*-4 site by 18.2-fold and 4.3-fold, respectively (Fig. 4c). The combined mutant, TnpB-G9F/E350R, showed additionally increased deletion efficiency at several sites in the ADE2, most notably at ADE2-5 site where editing efficiency was 33-fold higher than that of the wild-type, 30.31 folds higher than that of TnpB-G9F mutant, and 27.27 folds higher than that of TnpB-E350R mutant (Fig. 4d). In contrast, neither the single nor the double mutants exhibited improved deletion efficiency at *ADE2*-2, which was consistent with commonly observed target-dependent preferences of nucleases such as Cas proteins and TnpB orthologs [16, 24, 25].

To explore the potential structural basis for the improved editing activity of the two mutants, we analyzed the predicted TnpB - ωRNA - DNA complex structures in ChimeraX. We then evaluated residue - DNA contacts based on interatomic distances using non-bonded contact criteria (OVDW overlap ≥ -0.40). This analysis indicated that the substituted residues in both mutants form additional contacts with DNA that are not present in the wild-type. In the G9F mutant, the introduced phenylalanine side chain showed close contact with the phosphate group of dA (− 4) on the target-strand backbone (Fig. 4e). In the E350R mutant, new interactions were also observed between the arginine side chain showed multiple close contacts with the phosphate groups of dC (11)–dC (14) on the non-target strand (Fig. 4e).

These results demonstrate that the intrinsic activity of ISLde6 TnpB can be substantially enhanced through minimal rational engineering, highlighting its suitability as a scaffold for optimization of compact editor.

Alleviation of a low-Mg-sensitive DNA unwinding barrier by G9F and E350R

To investigate molecular basis for the improvement of eukaryotic editing of TnpB-G9F and TnpB-E350R, we examined key steps in target cleavage, including DNA binding, unwinding, R-loop formation, and

cleavage [26] (Fig. 5a). We first explored whether the improved editing activity was derived from increased DNA binding or cleavage efficiency of these mutants. Purified RNP complexes (wild-type TnpB, TnpB-G9F, and TnpB-E350R) and their catalytically inactive variants generated by introducing D251A/E344A/D432A substitutions into each construct (yielding dTnpB, dTnpB-G9F, and dTnpB-E350R) were subjected to in vitro cleavage and electrophoretic mobility shift assays (EMSAs). Under standard in vitro conditions (10 mM Mg^{2+}), both TnpB-G9F and TnpB-E350R complexes exhibited cleavage activities comparable to wild-type TnpB (Fig. 5b), suggesting that the improved editing activity is unlikely to be attributed to increased cleavage activity under high-Mg^{2+} conditions. Notably, EMSA revealed that dTnpB-G9F displayed a stronger DNA-binding activity than dTnpB under standard in vitro conditions (10 mM Mg^{2+}), which was consistent with the structural predictions (Supplementary Fig. 5a).

Since intracellular Mg^{2+} concentrations in eukaryotic cells (0.1–1 mM) are significantly lower than those in typical biochemical assays [27, 28], we next examined whether the concentration of Mg^{2+} affects the cleavage activity of ISLde6 TnpB (Fig. 4d). Cleavage and EMSA assays were performed across an Mg^{2+} gradients (0.01–25 mM). Under low Mg^{2+} conditions (0.1 mM), the cleavage activity of wild-type TnpB decreased markedly to 12.7%, whereas TnpB-G9F and TnpB-E350R retained substantially higher activities (47.7% and 38.3%, respectively) (Fig. 5c and Supplementary Fig. 5b). In contrast, EMSA showed that DNA-binding activity of the catalytically inactive mutants was unaffected by Mg^{2+} concentration, with dTnpB-G9F consistently displaying stronger DNA-binding activity under both low and high Mg^{2+} conditions (Fig. 5d). These results collectively suggest that the reduced cleavage activity of TnpB at low Mg^{2+} concentrations is unlikely to arise from altered DNA-binding activity.

Further, we examined whether differences in DNA unwinding could account for the improved editing activity of the mutants. To this end, we performed cleavage assays under substrate-limiting conditions using two substrate types: fully matched and pre-unwound substrates. Under 0.1 mM Mg^{2+}, both mutants exhibited substantially higher observed rate constants (k_{obs}) (k_{obs}=0.22 min^{-1} for TnpB-G9F and k_{obs}=0.51 min^{-1} for TnpB-E350R) than wild-type TnpB (k_{obs}=0.02 min^{-1}) with fully matched substrates used (Fig. 5e and Supplementary Fig. 5c). When pre-unwound substrates (2mm-NTS-2 and 2mm-NTS-3 respectively containing 2-bp mismatches in the middle or at the 3’ end of the NTS target region) were used, wild-type TnpB showed a dramatic increase in cleavage activity (~ 140-fold increase in k_{obs}), whereas the mutants exhibited only modest improvements (≤ 6-fold increase in k_{obs}) (Fig. 5f, g and Supplementary Fig. 5e, f). These results suggest that the cleavage activity of wild-type TnpB is strongly limited by an unwinding-related step under low physiological Mg^{2+} conditions, and that the G9F and E350R substitutions partially alleviate this limitation.

Collectively, our results support that the G9F and E350R substitutions alleviate an unwinding-related kinetic constraint under low physiological Mg^{2+} conditions, thereby improving genome-editing activity in eukaryotic cells.

Improved genome editing in non-model industrial microorganisms by FIRE-TnpB

Next, we investigated the performance of TnpB-G9F/E350R by integrating a 2.8-kb glucoamylase expression cassette (GlcA) into the diploid *S. cerevisiae* strain M, an industrial yeast. The results showed that TnpB-G9F/E350R achieved integration efficiencies comparable to those of SpCas9 (Supplementary Fig. 6a), and that the resulting GlcA-integrated strain displayed the expected starch-degradation phenotype (Supplementary Fig. 6b, c, d).

Although TnpB-G9F/E350R supported efficient integration of a moderate-sized cassette, its efficiency dropped to 2–5% for 10.5-kb astaxanthin-synthesis pathway integration, highlighting the necessity for further optimization of large-fragment integration. Therefore, we examined the effects of domain fusion on TnpB editor performance. To sensitively evaluate fusion-dependent performance increase across target sites, we performed the initial fusion screening on the wild-type ISLde6 scaffold. Specifically, we fused T5 exonuclease (T5E) [29], Exonuclease III (ExoIII) [30], or the small non-specific DNA-binding protein Sso7d [31] to either the N or C terminus of TnpB, generating seven fusion constructs. C-terminal fusions generally outperformed their N-terminal counterparts, with TnpB-T5E and TnpB-Sso7d showing the largest editing activity gains, increasing editing efficiency by 22.4 folds and 5.4 folds, respectively, relative to that of wild-type TnpB (Fig. 6b, c). Then, we generated two tandem fusion mutants TnpB-Sso7d-T5E and TnpB-T5E-Sso7d by inserting Sso7d within TnpB-T5E or fusing it to the C-terminus. Compared with TnpB-T5E, TnpB-Sso7d-T5E showed the greatest enhancement in *ADE2* deletion efficiencies, with 5.9-fold and 6.4-fold increases at the *ADE2*-1 and *ADE2*-5 sites, respectively, whereas TnpB-T5E-Sso7d exhibited only approximately two-fold improvement at ADE2-5 (Fig. 6d).

Incorporation of the G9F and E350R substitutions into TnpB-Sso7d-T5E yielded FIRE-TnpB (functional domain-integrated and rationally engineered TnpB). We next evaluated the performance of FIRE-TnpB in non-model industrial microorganisms. The results showed that FIRE-TnpB mediated efficient one-step integration of the complete 10.5-kb astaxanthin biosynthetic pathway into multicopy non-transcribed spacer (NTS) regions of the rDNA, with integration efficiencies of 14.05%, 50%, and 21.91% at three target sites, exceeding those achieved by SpCas9 at the same sites (Fig. 6e). These results highlight the utility of FIRE-TnpB for pathway engineering in complex host background and demonstrate that Sso7d-T5 fusion improves large-fragment integration efficiency in TnpB-G9F/E350R mutant.

To further test whether FIRE-TnpB functioned across diverse genetically intractable microbial hosts, we next evaluated its gene deletion efficiency in representative non-model industrial microorganisms, especially non-model lactic acid bacteria. In *S. cerevisiae*, FIRE-TnpB achieved high gene deletion efficiencies in both industrial diploid strains M and MP (81.48–95.38%) and model haploid strains CEN.PK2-1c and S288C (87.96–98.41%) (Fig. 6f). This engineered editor also exhibited high gene deletion efficiency in multiple non-model lactic acid bacteria strains (86.58–99.19%) and in *Escherichia coli* MG1655 (95.56–99.17%) (Fig. 6f).

These results indicate that fusion of DNA-binding and exonuclease modules to TnpB enhances its editing performance, providing an effective engineering strategy for further improving TnpB scaffold performance. Moreover, these results also demonstrate that FIRE-TnpB is a compact, efficient, and broadly applicable genome-editing platform capable of supporting complex genetic engineering across diverse microbial hosts.

Discussion

In this study, we identified ISLde6 TnpB as a compact RNA-guided nuclease with engineering potential. We then improved its performance through rational engineering and found that, under low Mg^{2+} conditions, these mutations alleviate a pre-catalytic barrier associated with target-DNA unwinding. Through further domain fusion, we ultimately developed FIRE-TnpB, which enables one-step integration of a large metabolic pathway in industrial yeast and efficient genome editing across multiple non-model industrial microorganisms.

Continued discovery of TnpB and related editors has expanded the repertoire of compact effectors available for tool development [10, 12, 32]. Emerging evidence suggests that these systems exhibit advantages in target range [15], host compatibility [14], and engineering potential [25, 33, 34]. ISLde6 TnpB represents a useful addition to the current TnpB toolbox due to its compact size and favorable editing properties, including cross-host activity (Fig. 2b, c), and relatively flexible TAM recognition (Fig. 3f, g), all of which are desirable properties for subsequent editor development.

Notably, wild-type TnpB showed a lower cleavage rate than its mutants under low Mg^{2+} conditions (Fig. 5e), while binding assays demonstrate that this difference is not caused by impaired DNA binding ability (Fig. 5d). Moreover, the cleavage rate of wild-type TnpB was comparable to that of its variants when the pre-unwound DNA substrates were used (Fig. 5f and g). Collectively, these results suggest that the limited activity of wild-type TnpB under low Mg^{2+} conditions is mainly due to inefficiency in the DNA unwinding process. This result is in line with previous report that DNA unwinding and R-loop formation of iGeoCas9 become rate-limiting steps under low Mg^{2+} conditions [28]. In Cas12, the TNB domain is thought to contribute to DNA unwinding by engaging both the guide RNA–target DNA hybrid and the re-hybridized DNA duplex in the PAM-distal region [35–38]. In contrast, structural studies of ISDra2 TnpB suggest that its TNB domain may not exert corresponding function, indicating that the protein relies more on the spontaneous unwinding of substrate DNA [38]. Notably, the effect of pre-unwound DNA on the cleavage rate of wild-type TnpB was position dependent, with mismatches in the middle or TAM-distal NTS increasing this rate under low Mg^{2+} conditions (Fig. 5f, g and Supplementary Fig. 5e, f), whereas TAM-proximal mismatches showed no such effect (Supplementary Fig. 5d). This result suggests that the TAM-proximal region may have an important role in TnpB-mediated DNA cleavage. However, this proposed unwinding-related limiting step is inferred mainly from indirect kinetic evidence and should be further examined by direct mechanistic studies.

This study further optimized the ISLde6 scaffold through functional domain fusion. Based on the current data (Fig. 6c), we propose that T5 exonuclease may facilitate the processing of DNA ends generated after cleavage [39], while Sso7d may enhance the binding of TnpB to target DNA [31]. With these improvements, the resulting FIRE-TnpB system enables one-step integration of a complete 10.5-kb biosynthetic pathway into multicopy rDNA sites in industrial yeast (Fig. 6e) and supports efficient genome editing in several non-model lactic acid bacteria (Fig. 6f). These findings suggest that FIRE-TnpB has considerable application potential in industrially relevant microorganisms that remain challenging for genetic manipulation and DNA delivery, especially in cases where Cas-based systems are constrained by vector size and cargo capacity [40, 41]. At the same time, we observed that the editing efficiency of FIRE-TnpB still varies with host and target-site. Therefore, future improvements in target-site accessibility [42, 43] and post-cleavage DNA repair efficiency [44, 45] may further broaden the applicability of FIRE-TnpB and other compact RNA-guided nucleases.

Overall, this study establishes an integrated route for developing compact RNA-guided nucleases, spanning natural TnpB scaffold identification, mechanistic characterization, and engineering optimization. It further establishes FIRE-TnpB as a compact and engineerable genome-editing platform for non-model industrial microorganisms, and provides a mechanistic basis for the future optimization of TnpB and other compact RNA-guided nucleases.

Declarations

Acknowledgements

We acknowledge the National Key Laboratory of Agricultural Microbiology and the Core Facility and Experimental Teaching Center of Bioengineering at Huazhong Agricultural University for providing access to the protein purification and bioinformatics computing platforms.

Funding

The National Key Research and Development Program of China (2022YFA0912200); The Hubei Key R&D Program in Agricultural Core Technologies (HBNYHXGG2023-7); The Hubei Special Project for Science Development (No. 2024CSA060); The Hubei Key Technology R&D Program (2025BCA006); The Major Special Project for the Development of Agricultural Microbial Industry (NYWSWZX2025-2027-10); The Hubei Natural Science Foundation (2024AFD125); The Hubei Special Project for Science Development (No. 2024CSA060); The Fundamental Research Funds for the Central Universities (No.2662025SKPY010)

Author contributions

Conceptualization: N.P. and Q.J.G.; Methodology: Z.Y.F., Y.Z., J.Y., W.K.R. Boya Xu and G.H.L.; Investigation: J.L.L. and Q.Y.; Visualization: N.P. and Q.J.G.; Funding acquisition: N.P.; Project

administration: Q.J.G., Y.Y.W., Binyang Xiong.; Supervision: N.P. and Q.J.G.; Writing – original draft: Q.J.G.; Writing – review & editing: N.P.

Competing interests: All authors declare that they have patents pending to this material.

References

1. Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096
2. Madigan V, Zhang F, Dahlman JE (2023) Drug delivery systems for CRISPR-based genome editors. Nat Rev Drug Discov 22:875–894
3. Yang JH, Song JY, Feng ZY, Ma YQ (2025) Application of CRISPR-Cas9 in microbial cell factories. Biotechnol Lett 47
4. Collias D et al (2023) Systematically attenuating DNA targeting enables CRISPR-driven editing in bacteria. Nat Commun 14
5. Lu LY et al (2022) CRISPR-based metabolic engineering in non-model microorganisms. Curr Opin Biotechnol 75
6. Geng BN et al (2024) Efficient genome-editing tools to engineer the recalcitrant non-model industrial microorganism Zymomonas mobilis. Trends Biotechnol 42
7. Riley LA, Guss AM (2021) Approaches to genetic tool development for rapid domestication of non-model microorganisms. Biotechnol Biofuels 14
8. Guo Q et al (2023) Developing the Limosilactobacillus reuteri Chassis through an Endogenous Programmable Endonuclease-Based Genome Editing Tool. ACS Synth Biol 12:3487–3496
9. Liu ZL et al (2023) Endogenous CRISPR-Cas mediated genome editing: State-of-the-art and the road ahead for engineering prokaryotes. Biotechnol Adv 68
10. Altae-Tran H et al (2021) The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. Science 374:57–65
11. Karvelis T et al (2021) Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. Nature 599:692–696
12. Xiang G et al (2023) Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. Nat Biotechnol 42:745–757
13. Altae-Tran H et al (2023) Diversity, evolution, and classification of the RNA-guided nucleases TnpB and Cas12. Proc Natl Acad Sci U S A 120:e2308224120
14. Xu Y et al (2023) Reprogramming an RNA-guided archaeal TnpB endonuclease for genome editing. Cell Discovery 9
15. Feng X et al (2024) Flexible TAM requirement of TnpB enables efficient single-nucleotide editing with expanded targeting scope. Nat Commun 15:3464

16. Marquart KF et al (2024) Effective genome editing with an enhanced ISDra2 TnpB system and deep learning-predicted ωRNAs. Nat Methods. 10.1038/s41592-024-02418-z
17. Li Z et al (2024) Engineering a transposon-associated TnpB-ωRNA system for efficient gene editing and phenotypic correction of a tyrosinaemia mouse model. Nat Commun 15
18. Lu J et al (2026) Miniature and versatile genome regulation TnpB-ωRNA toolkits facilitate cancer immunotherapy. Nat Commun. 10.1038/s41467-026-71327-w
19. Thornton BW et al (2026) Engineered TnpB genome editors for plants and human cells identified by ribonucleoprotein mutational scanning. Nat Biotechnol. 10.1038/s41587-026-03059-7
20. Volke DC, Orsi E, Nikel PI (2023) Emergent CRISPR–Cas-based technologies for engineering non-model bacteria. Curr Opin Microbiol 75
21. Song X et al (2024) CRISPR-AsCas12f1 couples out-of-protospacer DNA unwinding with exonuclease activity in the sequential target cleavage. Nucleic Acids Res 52:14030–14042
22. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33:2302–2309
23. Sasnauskas G et al (2023) TnpB structure reveals minimal functional core of Cas12 nuclease family. Nature. 10.1038/s41586-023-05826-x
24. Kulcsár PI et al (2023) A cleavage rule for selection of increased-fidelity SpCas9 variants with high efficiency and no detectable off-targets. Nat Commun 14:5746
25. Kannan S et al (2025) Evolution-guided protein design of IscB for persistent epigenome editing in vivo. Nat Biotechnol. 10.1038/s41587-025-02655-3
26. Sasnauskas G et al (2023) TnpB structure reveals minimal functional core of Cas12 nuclease family. Nature 616:384–389
27. Beeler T, Bruce K, Dunn T (1997) Regulation of cellular Mg2 + by Saccharomyces cerevisiae. Bba-Biomembranes 1323:310–318
28. Eggers AR et al (2024) Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. Cell 187:3249–3261e3214
29. Zhang Q et al (2020) Fusing T5 exonuclease with Cas9 and Cas12a increases the frequency and size of deletion at target sites. Sci China Life Sci 63:1918–1927
30. Zhang K et al (2022) Fusing an exonuclease with Cas9 enhances homologous recombination in Pichia pastoris. Microb Cell Fact 21
31. Han L et al (2024) Engineering miniature IscB nickase for robust base editing with broad targeting range. Nat Chem Biol 20:1629–1639
32. Saito M et al (2023) Fanzor is a eukaryotic programmable RNA-guided endonuclease. Nature. 10.1038/s41586-023-06356-2
33. Zhang F et al (2025) Engineering a CRISPR-associated IscB system for developing miniature genome-editing tools in human cells and mouse embryos. Nat Commun 16

34. Xiao QQ et al (2025) Engineered IscB-ωRNA system with expanded target range for base editing. Nat Chem Biol 21
35. Takeda SN et al (2021) Structure of the miniature type V-F CRISPR-Cas effector enzyme. Mol Cell 81
36. Swarts DC, Jinek M (2019) Mechanistic Insights into the cis- and trans-acting DNase activities of Cas12a. Mol Cell 73:589–
37. Liu JJ et al (2019) CasX enzymes comprise a distinct family of RNA-guided genome editors. Nature 566:218–
38. Nakagawa R et al (2023) Cryo-EM structure of the transposon-associated TnpB enzyme. Nature 616:390–397
39. Han D et al (2023) Development of miniature base editors using engineered IscB nickase. Nat Methods 20:1029–1036
40. Rozanov AS, Shaposhnikov LA, Bondarenko KD, Sazonov AE (2025) Advances in Genetic Transformation of Lactic Acid Bacteria: Overcoming Barriers and Enhancing Plasmid Tools. Int J Mol Sci 26
41. Song X et al (2020) CRISPR-Cas-mediated gene editing in lactic acid bacteria. Mol Biol Rep 47:8133–8144
42. Jensen KT et al (2017) Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. FEBS Lett 591:1892–1901
43. Offerhaus HSJ, Jones I, Depken SK (2026) Martin, A mechanistic free-energy model explains and predicts sequence- and context-dependent CRISPR–Cas9 activity. 10.64898/2025.12.12.691775
44. Reginato G et al (2024) HLTF disrupts Cas9-DNA post-cleavage complexes to allow DNA break processing. Nat Commun 15
45. Nambiar TS, Baudrier L, Billon P, Ciccia A (2022) CRISPR-based genome editing through the lens of DNA repair. Mol Cell 82:348–388

Methods

Genome-wide identification of multicopy TnpB

A total of 353,677 bacterial genomes were obtained from the RefSeq database (till May 2024) [1]. Protein-coding sequences were predicted using Prodigal v2.6.3 (with parameters:-p single) [2]. To identify TnpB homologs, a customized reference database was constructed based on all annotated TnpB proteins in the IS200/IS605 and IS607 families obtained from in ISfinder (till March 2024) [3]. Homology analyses were conducted using DIAMOND BLASTp with the parameters “--identity 20 --query-cover 30 --evalue 1e-5” [4]. Only proteins whose lengths were between 200 and 600 amino acids were retained [5]. To calculate the copy number, candidate TnpB proteins in each genome were clustered using CD-HIT (with parameters: -c 0.9 -n 5 -s 0.9). Subsequently, the last 500-bp nucleotide sequences corresponding to the C-terminal regions were extracted, with sequences lacking sufficient length excluded. These sequences

were clustered with CD-HIT-EST (with parameters: -c 0.9 -n 8) to remove duplicates. Only genomes containing at least two non-redundant TnpB homologs were selected for subsequent analysis [6].

To identify their TAM motifs and reRNAs, multicopy TnpBs with at least three copies from the Lactobacillaceae family were further analyzed. For each cluster, a 1,500-bp region upstream TnpB was extracted, including 1,300 bp upstream the start codon and first 200-bp TnpB-coding sequence [5]. Multiple sequence alignment was performed using MAFFT to detect conserved TAM motifs [7]. In addition, regions comprising the last 200-bp TnpB-coding sequence and 300-bp sequence downstream TnpB were aligned to identify TnpB-associated reRNAs. All multicopy TnpB proteins were further clustered using CD-HIT (with parameters: -c 0.7 -n 5 -s 0.9), and 65 representative non-redundant sequences were selected based on copy number and conservativeness.

To identify conserved residues, these 65 sequences were aligned against representative TnpB proteins including ISDra2, SisTnpB1, ISDge10, ISAba30, and ISTfu1 using MAFFT.

Phylogenetic analysis

To examine the distribution of TnpB proteins across bacteria, a domain-level phylogenetic tree was constructed using 16S rRNA gene sequences from the SILVA SSU database (release V138) [8]. Sequences were clustered at the family level using MMseqs2 with 90% identity and 90% coverage (with parameters: cov-mode 1) [9]. The longest sequence from each cluster was selected as a representative, yielding a total of 753 representative sequences. Multiple sequence alignment was performed using MAFFT v7.525 (with parameters: --genafpair --maxiterate 1,000) The obtained alignment results were subsequently trimmed using trimAl with the -gappyout method [7, 10]. The phylogenetic relationships were reconstructed using the maximum-likelihood method in IQ-TREE, with the optimal substitution model automatically selected by ModelFinder. Branch support was evaluated using 1,000 SH-aLRT and 1,000 ultrafast bootstrap replicates [11-13].

To determine the phylogenetic relationships of these TnpBs, a second phylogenetic tree was constructed using 65 multicopy TnpBs together with 27 previously characterized TnpBs [11, 12, 14]. Alignment and trimming were performed using MAFFT (with parameters: --auto) and trimAl (with parameters: -automated1), respectively [7,10]. Phylogenetic inference was conducted using IQ-TREE with model selection (with parameters: -m MFP) and 1,000 ultrafast bootstrap replicates [11-13].

In addition, a third phylogenetic analysis was conducted using the same 65 sequences and 5 reference sequences (ISDra2, SisTnpB1, ISDge10, ISAba30, and ISTfu1 TnpBs). The same alignment, trimming, and tree-building procedures were implemented. All trees were visualized using iTOL [14].

Plasmid construction

DNA fragments were amplified by PCR using Phanta Max Super-Fidelity DNA Polymerase (Vazyme, Nanjing, China). Primers were designed with 15-20 bp homologous sequences to facilitate seamless

assembly. PCR products were digested with Dpn I to remove template plasmid DNA and subsequently assembled into the desired vectors via T5 exonuclease-dependent assembly (TEDA). The resulting plasmids were transformed into *E. coli* DH5α, and correct colonies were selected for plasmid extraction using the Omega Plasmid Mini Kit I (D6943-02, Omega Bio-Tek). All vectors were verified by Sanger sequencing (Quintara, Wuhan), and primers were synthesized by Sangon (Shanghai, China).

For lactic acid bacteria, TnpB-encoding sequences (controlled by the P32 promoter) and ωRNAs (driven by PG00694, a high-efficiency promoter cloned from the genomic locus upstream phosphopyruvate hydratase gene in *L. reuteri* 03) were assembled into pMG36e and pMG36c [15] via TEDA, yielding interference plasmid pMG36e-TnpB-ωRNA and pMG36c-TnpB-ωRNA. Additionally, 1-kb homology arms flanking the deletion region were PCR-amplified and cloned into interference plasmids via TEDA to generate corresponding gene deletion plasmids.

For bacterium *B. subtilis*, the TnpB-encoding genes (under the control of Bacitracin-inducible PliaI promoter) and ωRNAs (under the control of PglyA promoter) were cloned into pHY [16] via TEDA, producing interference plasmid pHY-TnpB-ωRNA. Furthermore, 500-bp homology arms flanking the deletion region were PCR-amplified and cloned into interference plasmids (e.g., pHY-TnpB-ωRNA) via TEDA to generate the corresponding gene deletion plasmids.

For yeast *S. cerevisiae*, the TnpB-encoding genes (under the control of TEF1 promoter) and ωRNAs (under the control of SNR52 promoter) were cloned into pSCm via TEDA, producing interference plasmid pSCm-TnpB-ωRNA. To enhance its editing efficiency, TnpB was fused with T5E-, Sso7d-, or ExoIII-encoding genes at its N- or C-terminus via TEDA, respectively, yielding TnpB fusion plasmids. For gene deletion, 500-bp homology arms flanking the deletion region were amplified and fused by SOE PCR (gene splicing by overlap extension PCR) to generate linear donor DNA. The TnpB-encoding gene sequences were codon optimized using GenSmart (GenScript, nanjing, China).

For bacterium *E. coli*, the TnpB-encoding genes (under the control of J23119 promoter) and ωRNAs (under the control of J23119 promoter) were cloned into pBAD33 via TEDA, producing interference plasmid pBAD33-TnpB-ωRNA. For gene deletion, 500 bp homology arms flanking the deletion region were PCR-amplified and cloned into interference plasmid pBAD33-TnpB-ωRNA via TEDA to generate the corresponding gene deletion plasmids. For protein expression in *E. coli*, TnpB-His (under the control of T7 operator-inducing IPTG) and its ωRNA (under the control of constitutive T7 promoter) were cloned into pET30a via TEDA, yielding protein-expressing plasmid pET30a-TnpB-His-ωRNA.

Genome editing in lactic acid bacteria

L. reuteri 03 [8] and H1-2 were cultured in MRS medium supplemented with 2% (w/v) glycine at 37°C for 4 h until optical density at 600 nm (OD_{600}) reached 0.8. Cells were harvested, washed for three times in buffer 1 (containing 205.37 g/L sucrose, 0.20 g/L $MgCl_2 \cdot 6H_2O$, 1.86 g/L $K_3PO_4 \cdot 3H_2O$, pH7.5), and resuspended in buffer 2 (171.15 g/L sucrose, 10% v/v glycerol) to obtain competent cells. Then, 800 ng

of the editing plasmid was electroporated into *L. reuteri* cells (1 kV, 25 μF, 200 Ω) using a Gene Pulser Xcell (Bio-Rad). Cells were recovered in MRS-Glu medium at 37°C for 6 h, plated on MRS agar with 5 μg/mL chloramphenicol, and incubated at 37°C for 48 h. Twenty-four colonies were randomly selected and verified by colony PCR and sequencing.

L. lactis LL09 was cultured in SGM17 medium with 1.1% (w/v) glycine at 30°C for 5 h until OD_{600} reached 0.8. Cells were washed for three times in buffer 3 (171.14 g/L sucrose, 126.10 g/L glycerol) and resuspended in buffer 3 to obtain competent cells. Then, 800 ng of the editing plasmid was electroporated into *L. lactis* cells (1.34 kV, 25 μF, 400 Ω). Cells were recovered in modified SGM17 medium at 30°C for 6 h, plated on GM17 medium containing 5 μg/mL erythromycin, and incubated at 30°C for 48 h. Twenty-four colonies were randomly selected and verified by colony PCR and sequencing.

P. acidilactici LA412 [17] was cultured in MRS medium added with 0.3% (w/v) D/L-threonine at 37°C with shaking (180 rpm) till an OD_{600} of 1.5. Cells were harvested by centrifugation (6,000 × g, 4°C, 5 min), washed twice with 20 mL buffer I (205.37 g/L sucrose, 0.20 g/L $MgCl_2 \cdot 6H_2O$, 1.86 g/L $K_2HPO_4 \cdot 3H_2O$, pH 7.5), and resuspended in 1 mL buffer I containing 0.1 mg/mL lysozyme. After 30-min incubation at 37°C, cells were washed twice with 20 mL buffer I and once with 1 mL buffer II (171.15 g/L sucrose, 10% glycerol). Then, 500 ng of the editing plasmid was electroporated into competent cells (2.5 kV, 25 μF, 200 Ω). Cells were recovered in modified MRS medium at 37°C for 6 h, plated on MRS medium added with 5 μg/mL erythromycin, and incubated at 37°C for 48 h. Twenty-four colonies were randomly selected and verified by colony PCR

Genome editing in *Saccharomyces cerevisiae*

Cells of haploid *S. cerevisiae* S288C, CEN.PK2-1c and diploid industrial *S. cerevisiae* strain M, MP (Angel Yeast Co., Ltd, Yichang, China) were cultured in YPD medium (1% yeast extract, 2% peptone, and 2% glucose, w/v) with a 4% (v/v) inoculum at 30°C and with shaking at 180 rpm until reaching an OD600) of ~0.8. Cells were harvested and washed three times with 10% (v/v) glycerol, followed by transformation. For single-gene editing, 500 ng of the editing plasmid was mixed with 2 μg of linear donor DNA and electroporated into competent cells using a Gene Pulser Xcell (Bio-Rad, Hercules, California) at 1.8 kV with a 5 ms pulse. Donor DNA fragments were PCR-amplified, gel-purified, mixed in equimolar amounts to obtain mixed linear donor DNA, and 6 μg of this mixed DNA fragments, including homologous arms (0.5 kb each for the left and right arms) and 10.5-kb fragments encoding the complete pathway, were co-electroporated with 500 ng of editing plasmids into competent cells for integration of the astaxanthin biosynthetic pathway containing five genes. Following electroporation, cells were recovered in liquid YPD medium at 30°C for 6 h and subsequently plated onto YPD medium supplemented with 500 μg/mL hygromycin B. Plates were incubated at 30°C for 5 days to allow transformant growth.

Transformants displayed a red phenotype, indicating successful ADE2 deletion or integration of astaxanthin synthetic pathway. Genome editing efficiency was calculated in the following formula: Editing Efficiency (%) = (Number of Pigmented Colonies/Total Colony Count) ×100.

A total of 24 transformants without visible phenotypes were randomly selected from each plate and subjected to genomic DNA release through standard alkali lysis. Subsequently, the target region of transformants were PCR amplified using primers ~100 bp upstream and downstream the editing site. PCR products underwent agarose gel electrophoresis and verification by Sanger sequencing when necessary. The number of randomly selected transformants were defined as X1, and the number of transformants verified to carry the correct edited band was defined as X2. Each electroporation was performed in triplicate, and editing efficiency was calculated according to the following formula: Editing Efficiency (%) =X2/X1×100.

Genome editing in *Bacillus subtilis*

Bacillus subtilis SCK6 was cultured overnight in YN medium at 37°C with shaking (220 rpm). The overnight culture was then subcultured into fresh YN medium containing 1.5% xylose to reach an initial OD_{600} of 1.0 and incubated at 37°C with shaking (200 rpm) for an additional 2 h to generate competent cells. The 100 µL competent cells were mixed and electrophorated with 1 µL of plasmid DNA, followed by incubation at 37°C with shaking (200 rpm) for 90 min. The resulting cells were serially diluted, plated onto LB medium containing 20 µg/mL tetracycline, and incubated overnight at 37°C. Twenty-four colonies were randomly selected and verified by PCR.

Protein expression and purification

To express and purify TnpB protein, *Escherichia coli* Rosetta (DE3) competent cells were transformed with pET30a-based plasmid coexpressing ωRNA and C-terminal 10His-tagged TnpB, and cultured in LB medium containing 30 µg/mL kanamycin at 37°C. When the OD_{600} reached 0.6–0.8, protein expression was induced with 4 mM IPTG at 16°C for 16 h. Protein-expressing cells were harvested and lysed in buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 5% v/v glycerol) by sonication. Cell lyate was centrifuged to remove cell debris, and the supernatant was loaded onto a Ni2+-NTA affinity chromatography column (HisTrap HP, GE Healthcare). The protein was eluted in the buffers containing 20 mM HEPES (pH7.5 at 25°C), 500 mM NaCl, 5% (v/v) glycerol, and a gradient of imidazole from 20 mM to 500 mM. The target protein was collected, concentrated through ultrafiltration, and separated through a Superdex 200 Increase 10/300 GL column (Cytiva, Marlborough, MA, USA) using the buffer (20 mM HEPES, 500 mM NaCl, and 5% (v/v) glycerol, pH = 7.5). The peak target proteins were analysed by SDS-PAGE. The samples containing TnpB RNP complexes were dialysed in the buffer (20 mM HEPES, 250 mM NaCl, and 5% (v/v) glycerol, pH =7.5) and stored at -80°C for subsequent use. TnpB RNP complex concentration was quantified by SDS-PAGE and xxxx (BCA) assay.

DNA substrates

The pUC19-derived plasmids harboring TAM and target sequences were PCR amplified using primer pairs F/R-19 or F/R-19-1. PCR products were purified with the Cycle Pure Kit (Omega Bio-Tek, Georgia, America) to obtain linear double-stranded DNA (dsDNA) substrates (sequences listed in Supplementary Table2).

The 59 nt 5’-end FAM-labeled ssDNA oligonucleotides were synthesized in Sango (Shanghai, China). The 5’-end FAM-labeled and -unlabeled complementary strands (1:1 molar ratio) were mixed, heated to 95℃, slowly cooled to room temperature, and purified by recovering the corresponding bands from native polyacrylamide gel, finally yielding the 59nt dsDNA substrates.

In vitro DNA cleavage assays

Plasmid and dsDNA cleavage assays were conducted in 20 μl reaction system (containing 1×cleavage buffer (20 mM HEPES, 0.2 mM DTT, 20 mM NaCl, and 10 mM $MgCl_2$, pH= 7.5), 200 nM TnpB RNP complex, and 10 nM dsDNA) at 37°C for 1 h. The reaction was quenched with 5×loading dye (0.05% bromophenol blue, 25 mM EDTA, 4% w/v SDS) and 2 μL 20 mM proteinase K at 37°C for 1 h. The reaction products were subjected to 2% agarose gel electrophoresis, ethidium bromide staining, and visualization with a Gel Doc XR+ system (Bio-Rad). To determine the optimal temperature with maximum TnpB cleavage activity, reactions were performed at 20, 30, 37 and 42°C. To investigate metal ion preference, $MgCl_2$ was replaced with alternative metal chlorides (final 10 mM). The plasmid cleavage products were recovered from the agarose gel using Gel Extraction Kit D2500 (Omega Bio-Tek) and subjected to Sanger sequencing to determine cleavage patterns of TnpB.

The 5’-end FAM-labeled 59-bp DNA substrates were cleaved in 20 μl 20 μl reaction system (containing 1×cleavage buffer, 200 nM TnpB RNP complex, and 10 nM dsDNA) at 37°C for 0, 0.5, 1, 2, 5, 10, 30, 60 and 120 min. Then, the cleavage reaction was quenched by 2×loading dye solution (95% (v/v) formamide, 0.01% bromophenol blue, 25 mM EDTA, and 4% SDS), and subjected to electrophoresis with denaturing gel (20% polyacrylamide and 8.5 M urea in 0.5× TBE buffer). DNA fragments in denaturing gel were detected using a Typhoon scanner. Band intensity was determined by densitometry using the software ImageJ.

AlphaFold3 prediction of TnpB structure

Three-dimensional structures of ISLde6 TnpB and its mutants were predicted using AlphaFold3 with default parameters [18]. Uisng ChimeraX [19], the substrate recognition and ωRNA-binding capacities of top-ranked model were analyzed, and its catalytic sites and hydrophobic core regions were identified. Residues were substituted to optimize charge complementarity or reduce steric hindrance. Residues in ISLde6 TnpB were optimized as the candidate residues, generating a mutant pool containing 62 mutants.

Electrophoretic mobility shift assay (EMSA)

EMSA was used to investigate the 59-bp substrate dsDNA-binding capacities of dTnpB, dTnpB-G9F, and dTnpB-E350R, with substrate listed in Supplementary Table 2. Purified 59-bp dsDNA (2 nM) was incubated with different concentrations of dTnpB RNP complexes (50, 100, 150 and 200 nM) in 1×cleavage buffer at 37°C for 10 min, respectively. Reactions were quenched by 6×loading dye (50% glycerol, 0.1% bromophenol blue, and 0.1% xylene cyanol), and reaction products were then loaded on 6% native polyacrylamide gels, followed by gel electrophoresis in Tris-glycine buffer (25 mM Tris, 192 mM

glycine, 5 mM $MgCl_2$) at 100 V for 1 h. Fluorescent signals were visualized using an Amersham Typhoon 5 scanner (Cytiva).

Starch utilization assay

Glucoamylase-G5 expression cassette-integrated diploid industrial Saccharomyces cerevisiae strain was pre-cultured in liquid YPD medium at 30 °C with shaking at 120 rpm for 12 h. Aliquot of 10 µL culture was spotted onto agar plates containing 5 g/L soluble starch, 5 g/L peptone, 5 g/L yeast extract, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 0.01 g/L $FeSO_4 \cdot 7H_2O$, and 0.01 g/L NaCl. Plates were incubated statically at 30 °C for 48 h. After incubation, the plates were stained with iodine solution (10 g/L KI and 1 g/L I_2) to visualize starch hydrolysis. Glucoamylase activity of integrated strain was quantified by measuring the diameter of the clear halo (reflecting starch utilization capacity) around colonies.

Determination of cumulative mass loss and reducing sugar residual

Fermentation efficiency of wild-type and glucoamylase-integrated strains was assessed by measuring cumulative mass loss and residual reducing sugar concentration. Yeast strains were pre-cultured in YPD medium for 16 h and subsequently inoculated at 1.2% (v/v) into 100 mL YPD medium. The cultures were incubated at 30 °C with shaking at 200 rpm for 24 h. Cells were then harvested by centrifugation, washed, and resuspended in sterile water at a 1:1 ratio (wet biomass: volume). A 680 µL aliquot of the resulting suspension was inoculated into 300 mL of corn mash fermentation medium and cultured at 30 °C with shaking 200 rpm. The corn mash medium contained 330 g/L corn flour, 0.002 g/L $MgSO_4 \cdot 7H_2O$, 0.003 g/L YE-FM888 (a yeast extract rich in amino acids and micronutrients), and 0.00001% (v/v) H_2SO_4. The gelatinization of corn mash medium was initiated by adding 0.15 mL of α-amylase per 300 mL medium was conducted at 101°C for 75 min with continuous stirring. After cooling, 2.7 g/L urea and 0.45 mL of glucoamylase per 300 mL medium were added to initiate saccharification. Fermentation of glucoamylase-integrated strains was performed for 72 h. The total mass of each flask containing saccharified corn mash medium was measured at 0, 24, 48, and 72 h. Cumulative mass loss (ΔM, g) was calculated as follows: $\Delta M = M_0 - M_t$, where M_0 represents the total mass (g) of the fermentation system at 0 h, and M_t is the corresponding mass at time t (24, 48, or 72 h).

A glucose standard curve (0–1 mg/ml) was prepared. Aliquots of glucose (0, 0.2, 0.4, 0.6, 0.8, and 1.0 mL) were transferred to 15 mL tubes, adjusted to 1.0 mL with deionized water, added with 2 mL of DNS reagent, vortexed, heated in a boiling water bath for 2 min, cooled, and diluted to 15 mL with deionized water. Absorbance at 540 nm was measured. Residual reducing sugars in corn mash fermentation samples of wild type strain and glucoamylase-integrated strain were quantified using the DNS method. Samples were diluted 50 folds in acetate–sodium acetate buffer, mixed with 2 mL of DNS reagent, boiled for 10 min, cooled, and absorbance of samples at 540 nm was recorded.

Statistical analysis

Experiments were performed with 3 technical replicates. The mean genome editing efficiency at each target site is presented as mean± SD generated by Prism software (version 10.1.2). Cleaved curves of TnpB were fit according to the following formula: $Y = Y_{max} \times (1 \exp(-k \times t))$, where Y_{max} is the pre-exponential factor; k is the rate constant k_{obs}; and t is the reaction time in minutes [20].

References

1. A. O'Leary et al., Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44, D733-745 (2016).
2. Hyatt et al., Prodigal: prokaryotic gene recognition and translation initiation site identification. Bmc Bioinformatics 11 (2010).
3. Siguier, J. Perochon, L. Lestrade, J. Mahillon, M. Chandler, ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34, D32-36 (2006).
4. Buchfink, K. Reuter, H. G. Drost, Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods 18, 366 (2021).
5. Shen, B. Sipos, L. Zhao, SeqKit2: A Swiss army knife for sequence and alignment processing. Imeta 3, e191 (2024).
6. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658-1659 (2006).
7. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30, 772-780 (2013).
8. Quast et al., The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41, D590-D596 (2013).
9. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nature Biotechnology 35, 1026-1028 (2017).
10. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972-1973 (2009).
11. Q. Minh et al., IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37, 1530-1534 (2020).
12. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14, 587-+ (2017).
13. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35, 518-522 (2018).
14. Letunic, P. Bork, Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res 52, W78-w82 (2024).
15. Gútiez et al., Genetic and Biochemical Evidence That Recombinant Enterococcus spp. Strains Expressing Gelatinase (GelE) Produce Bovine Milk-Derived Hydrolysates with High Angiotensin

Converting Enzyme-Inhibitory Activity (ACE-IA). Journal of Agricultural and Food Chemistry 62, 5555-5564 (2014).

16. Zou et al., Development and application of a rapid all-in-one plasmid CRISPR-Cas9 system for iterative genome editing in. Microbial Cell Factories 21 (2022).
17. Liu et al., High-Efficiency Genome Editing Based on Endogenous CRISPR-Cas System Enhances Cell Growth and Lactic Acid Production in *Pediococcus acidilactici*. Appl Environ Microbiol 87, e0094821 (2021).
18. Abramson et al., Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493-500 (2024).
19. F. Pettersen et al., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci 30, 70-82 (2021).
20. R. Eggers et al., Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. Cell 187, 3249-3261.e3214 (2024).

Figures

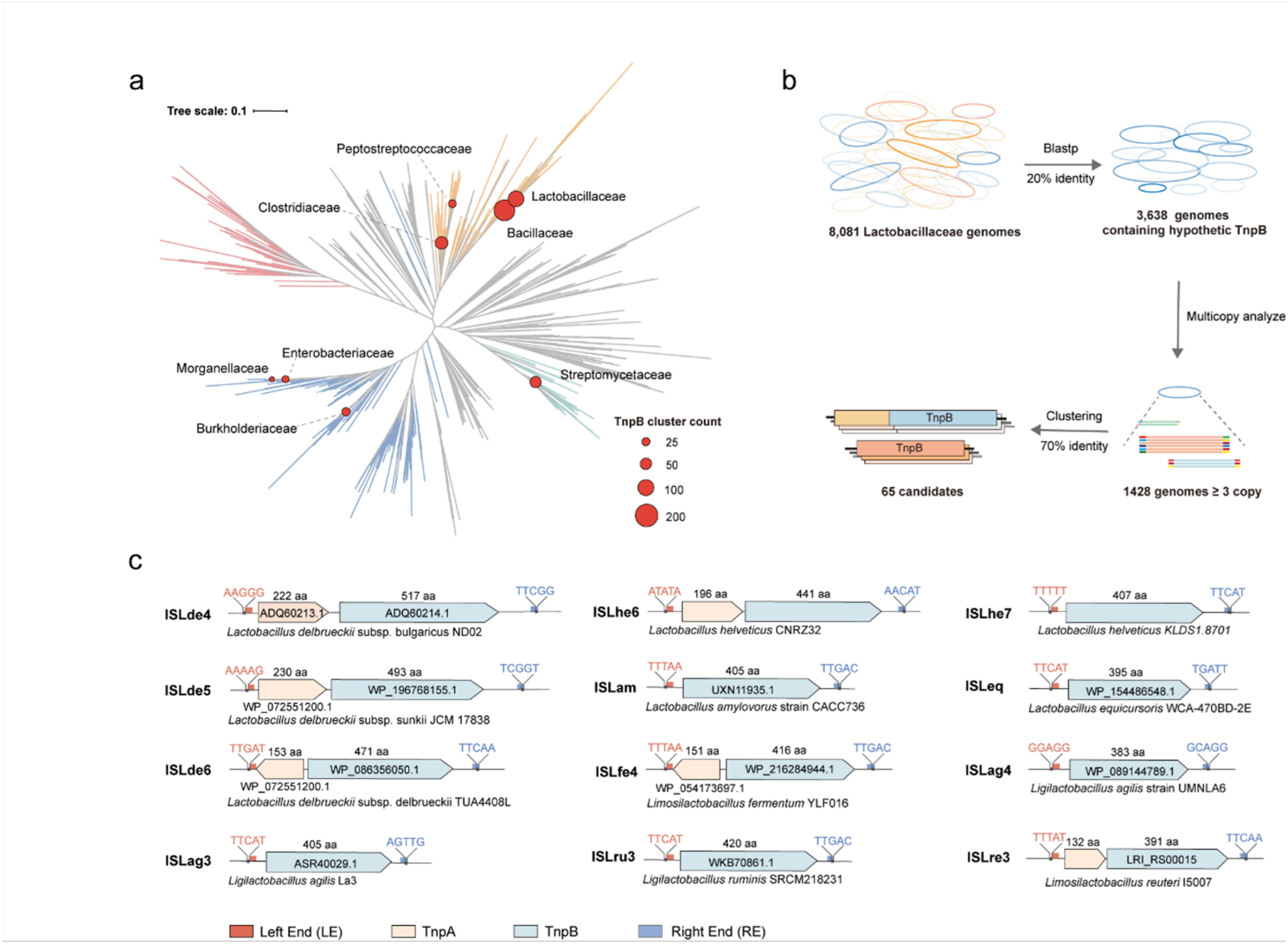


Figure 1

Phylogeny-guided identification of Lactobacillaceae TnpB scaffolds.

a, Phylogenetic tree constructed based on 16S rRNA gene sequences from representative bacterial genomes. Red dot size indicates the number of TnpB clusters identified from each family. Additional details are provided in Extended Data Fig. 1a. **b**, Workflow for identifying multicopy TnpB systems from Lactobacillaceae. A total of 3,638 genomes encoding TnpB were screened by aligning 8,081 complete genomes from the NCBI database against the ISfinder TnpB database using BLASTp (≥20% sequence identity). These TnpB-encoding sequences were further clustered at 70% identity, generating 65 representative TnpB systems. **c**, Structural compositon of twelve selected representative TnpB systems from Lactobacillaceae in their source genomes. These 12 TnpB systems (selected from 65 candidates based on key residue variations or phylogenetic distinctness) exhibit structural diversity. LE (left end) and RE (right end) denote palindromic terminal sequences flanking the TnpBsystem, and LE and RE are essential for transposition. The five nucleotides upstream LE (red) were used to infer the conserved target-associated motif (TAM), and the five nucleotides at the 3' end of RE (blue) correspond to the 3' terminus of the reRNA of TnpB.

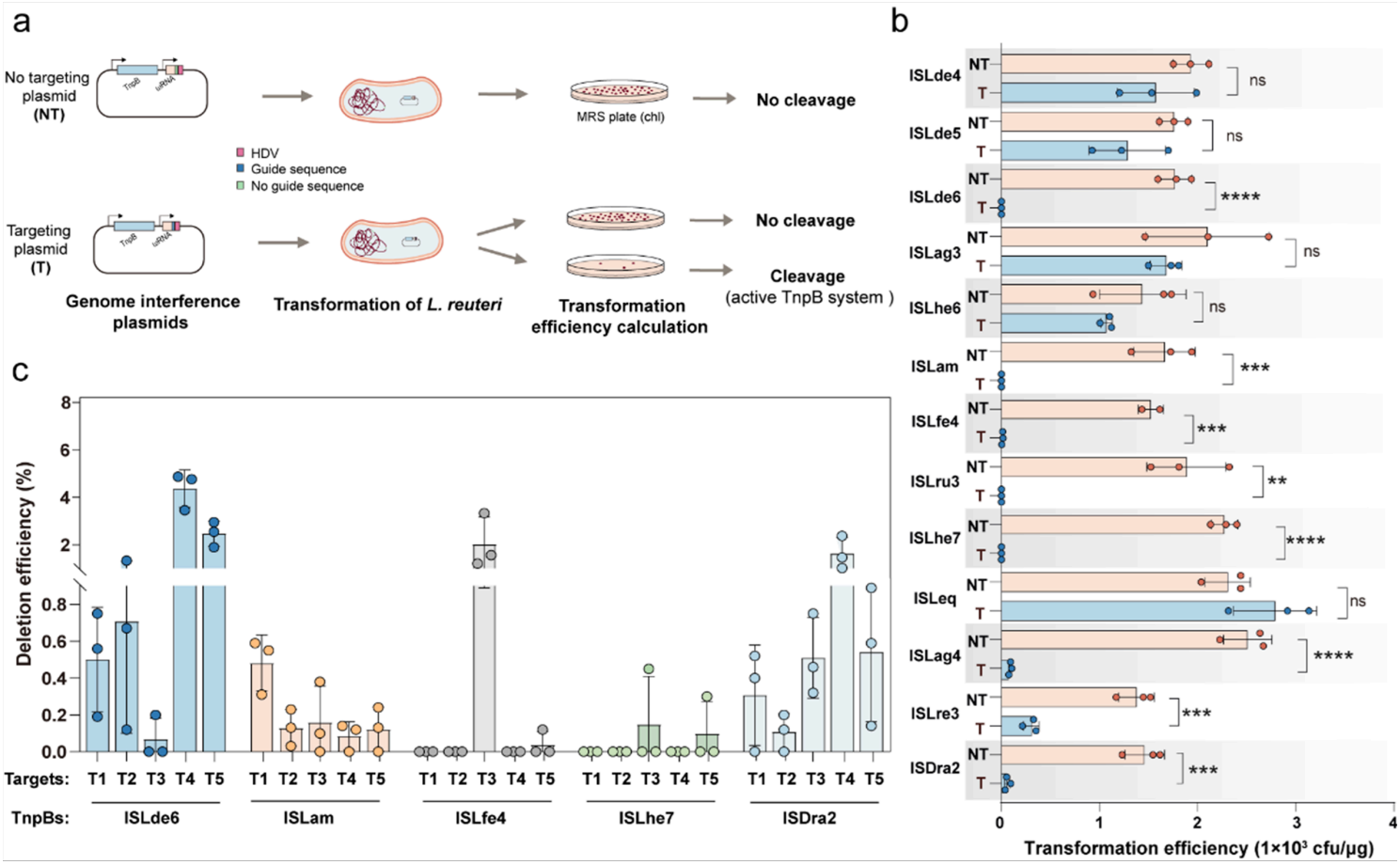


Figure 2

Identification of active TnpB candidates by cross-host screening. a, Schematic of active TnpB system screening assays. Non-targeting (NT) plasmids expressed the TnpB/ωRNA complex without guide sequences, while targeting (T) plasmids expressed this complex with guide sequences. HDV indicates hepatitis delta virus ribozyme. Plasmids were transformed into *Limosilactobacillus reuteri,* and active TnpB systems cleaved the genomic DNA, causing cell death and reducing transformation efficiency. **b,** Screening assay of 12 TnpB candidates in *L. reuteri.*RNA-guided nuclease activity of 12 TnpB systems was evaluated, organge and blue bars respectively denote the transformation efficiency of NT and T groups for each of 12 TnpB systems. **c,** Editing efficiencies of 5 TnpB systems in haploid *S. cerevisiae* S288C. Only TnpBs with detected activity are shown. Data are presented as mean ± SD (n ≥ 3 biological replicates).

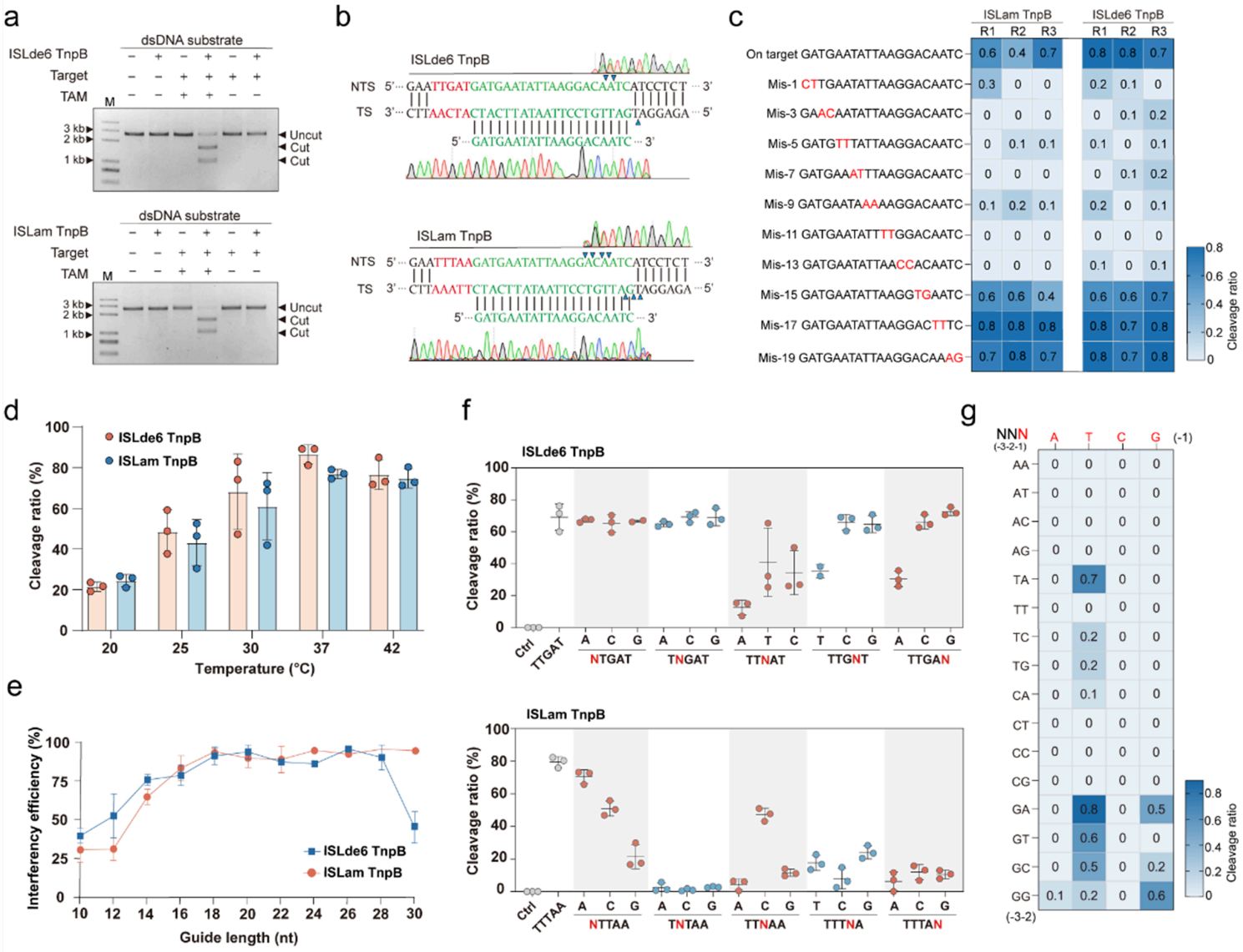


Figure 3

Functional characterization of ISLde6 TnpB and ISLam TnpBs. a *In vitro* cleavage of dsDNA substrates (TAM+/Target+, TAM−/Target+, and TAM+/Target−) by the ISLde6 and ISLam TnpBs at 37°C. Both the TAM sequence and the target sequence complementary to guide were required for cleavage. **b,** Identification of cleavage sites of ISLde6 and ISLam TnpBs. Blue triangles indicate cleavage sites (detected by Sanger sequencing) on the non-target strand (NTS) and target strand (TS) of cleaved plasmid products. C, Cleavage efficiencies (at 37°C) on dsDNA substrates containing sequential 2-nt mismatches within the target region. **d,** Cleavage efficiencies of ISLde6 and ISLam TnpB endonucleases across different temperatures (20, 25, 30, 37, and 42°C). **e,** Guide length dependence of ISLde6 and ISLam TnpB endonucleases. **f,** TAM sequence preference of ISLde6 and ISLam TnpB endonucleases. We measured cleavage efficiencies (at 37°C) on dsDNA substrates containing 5′-TTGAT or 5′-TTTAA TAMs with sequential single-nucleotide substitutions at each position. **g,** TAM recognition profile of ISLde6 TnpB. We measured cleavage efficiencies on substrates containing each of all 64 5′-NNN TAMs at 37°C. Data are presented as mean ± SD (n ≥ 3 biological replicates).

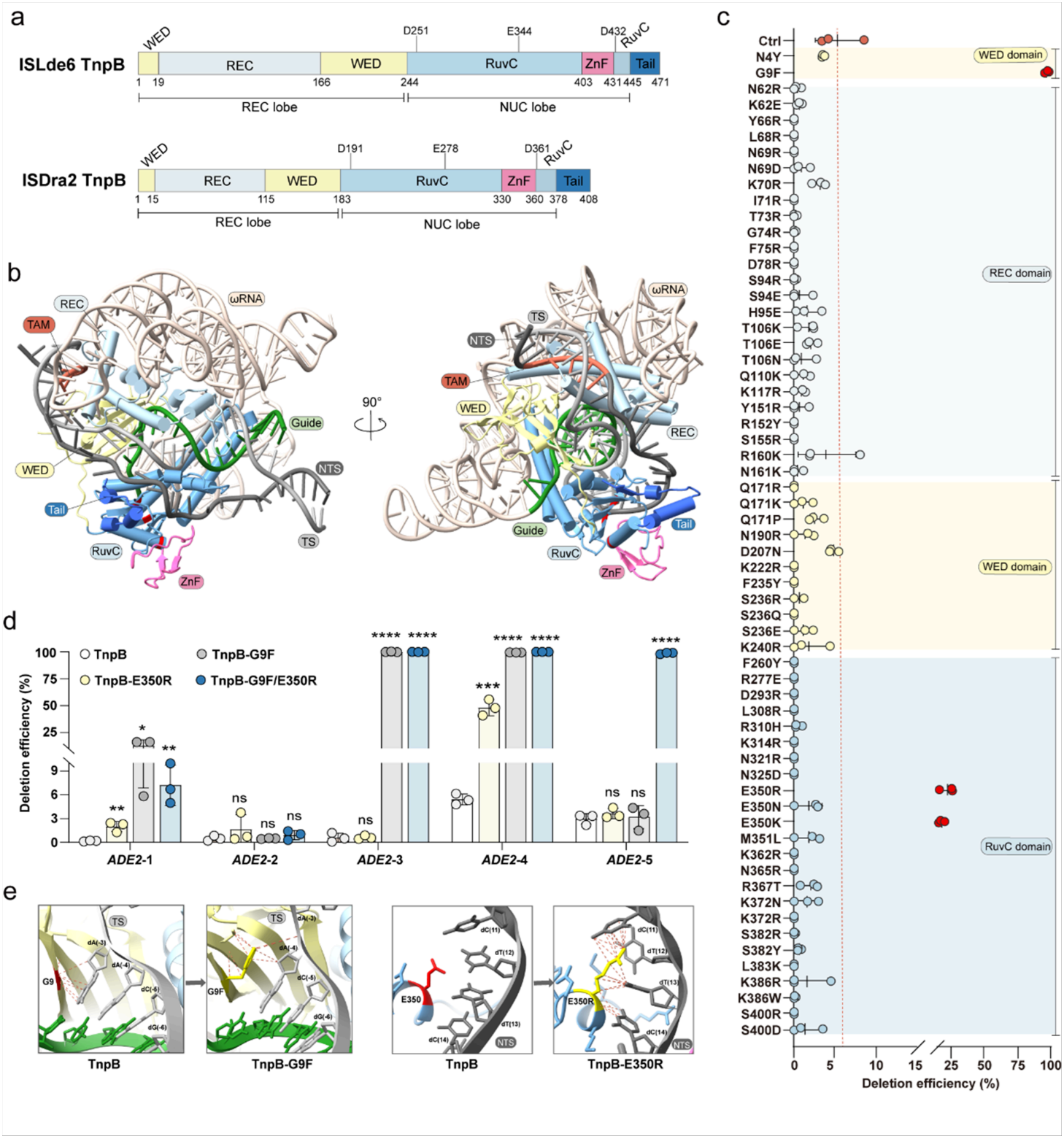


Figure 4

Structure-guided enhancement of ISLde6 editing activity. a, Domain comparison of ISLde6 and ISDra2 TnpBs. **b,** AlphaFold3 predicted ternary complex structure of ISLde6 - ωRNA - target dsDNA. **c,** Editing efficiencies of 62 ISLde6 TnpB mutants at the *ADE2*-4 site in *S. cerevisiae* S288C. **d,** Editing efficiencies of TnpB wild-type, TnpB-G9F, TnpB-E350R single mutants and TnpB-G9F/E350R double mutant at different sites in ADE2. **e,** Structural comparison of TnpB and its G9F and E350R mutants. Structures

were predicted by AlphaFold3 and analyzed by ChimeraX. The G9F mutation in REC domian created additional contacts with the target-strand phosphate at position dA (-4). The E350R mutation in RuvC domain established new contacts with non-target-strand phosphates at positions dC (11)–dC (14). New contacts are shown as dashed lines. Nucleotide positions of NTS are numbered 1 to 20 (5' to 3' direction), and nucleotide positions of TS are numbered -1 to -20 (3' to 5' direction). TAM, transposon-associated motif; NTS, non-target strand; and TS, target strand. Data are presented as mean ± SD (n ≥ 3 biological replicates).

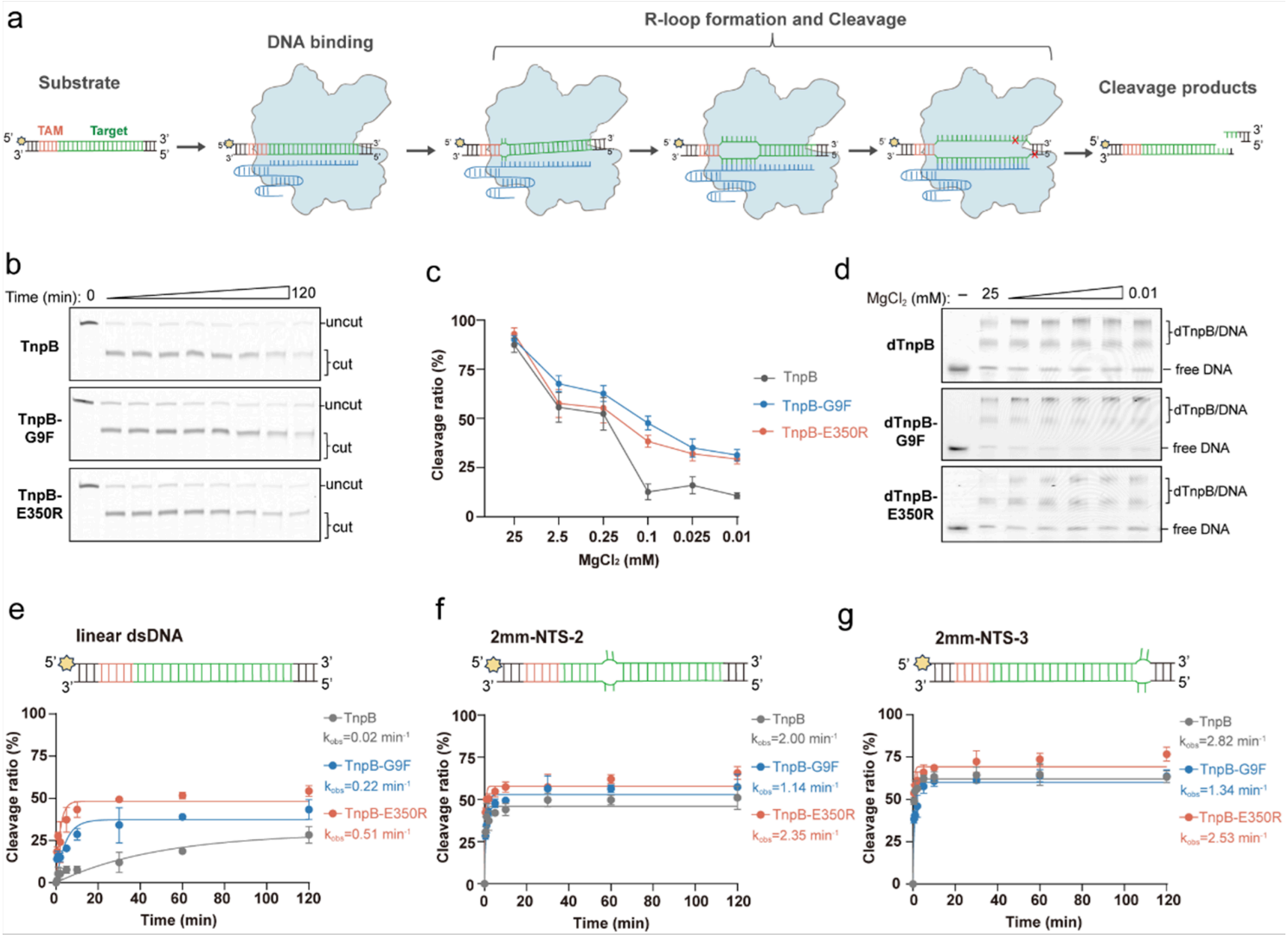


Figure 5

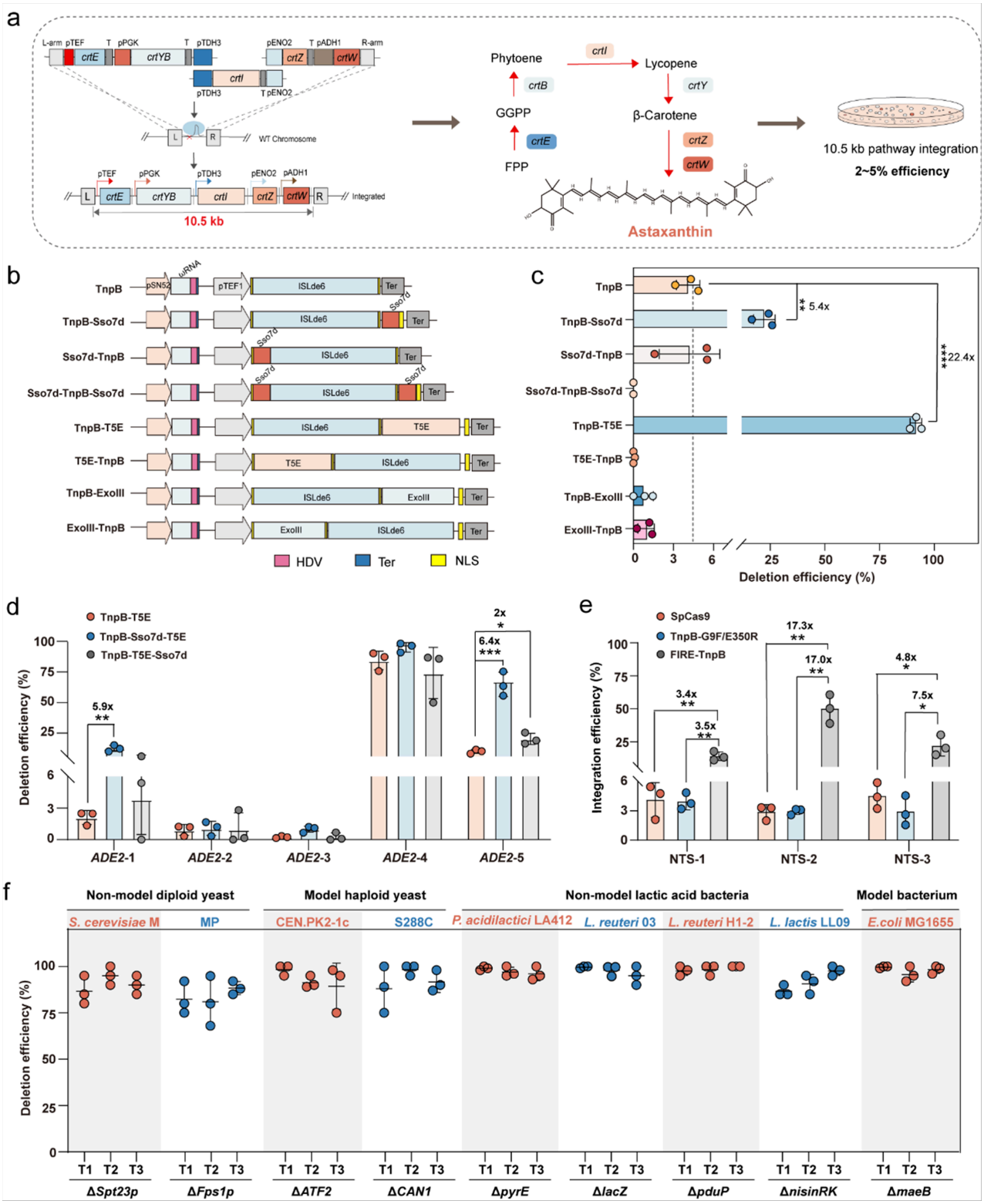


Figure 6

FIRE-TnpB-mediated large-fragment integration and broad editing across non-model industrial microorganisms. a, Strategy for one-step integration of the 10.5-kb astaxanthin biosynthetic pathway by TnpB-G9F/E350R. The pathway was assembled from three co-transformed DNA fragments (L arm + *crtE* + *crtYB*; *crtI*; and *crtZ* + *crtW*+ R arm) and integrated into the multicopy rDNA non-transcribed spacer (NTS) locus of S. cerevisiae strain M, yielding visibly pigmented colonies with 2 - 5% integration

efficiency. **b,** Schematic of TnpB and its functional domain fusion derivatives. **c**, Editing efficiencies of the TnpB-derived fusion editors at the *ADE2*-4 site. **d**, Editing efficiencies of the TnpB-T5E, TnpB-Sso7d-T5E, and TnpB-T5-Sso7d at various target sites in the *ADE2*. **e**, Integration efficiencies of astaxanthin biosynthetic pathway by SpCas9, TnpB-G9F/E350R, and the FIRE-TnpB (ISLde6-G9F/E350R-Sso7d-T5E) across three sites within multicopy NTS. **f,**FIRE-TnpB-mediated high-efficiency gene deletion across diverse non-model industrial and model microorganisms. Data were presented as mean ± SD. (n ≥ 3 biological replicates).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaltable1RawData.xlsx
- supplementaltable3primer.xlsx
- supplementaltable2StrainplasmidDNAsubstrateanddonorsequneces.xlsx
- SupplementaryFigureNature20260421.pdf