RESEARCH ARTICLE SUMMARY

MOLECULAR BIOLOGY

Hydrolytic endonucleolytic ribozyme (HYER) is programmable for sequence-specific DNA cleavage

Zi-Xian Liu†, Shouyue Zhang†, Han-Zhou Zhu†, Zhi-Hang Chen†, Yun Yang†, Long-Qi Li†, Yuan Lei, Yun Liu, Dan-Yuan Li, Ao Sun, Cheng-Ping Li, Shun-Qing Tan, Gao-Li Wang, Jie-Yi Shen, Shuai Jin, Caixia Gao, Jun-Jie Gogo Liu*

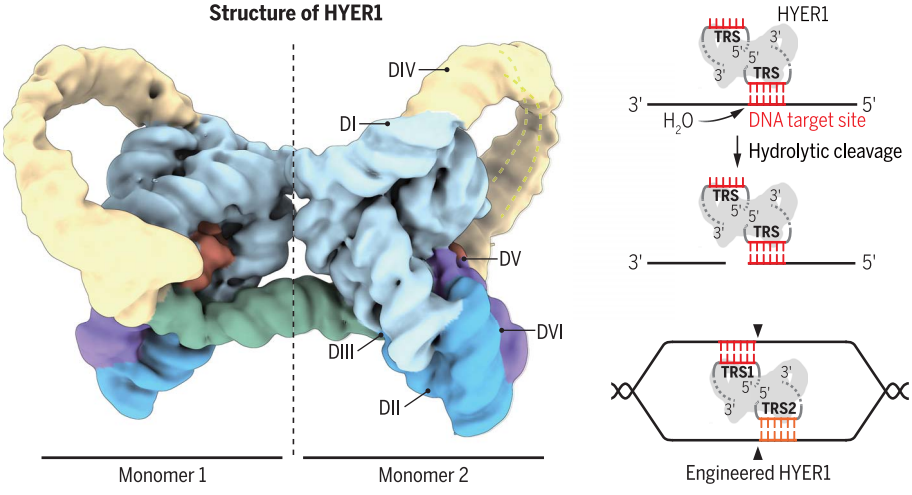
INTRODUCTION: Ribozymes play catalytic roles in diverse biochemical processes. Group II (GII) intron ribozymes are well-studied catalysts for splicing and retrotransposition with the assistance of intron-encoded proteins (IEPs) through transesterification. When IEPs are not provided, GII introns can cleave their precursor RNA targets through hydrolysis, forming mature and linearized ribozymes. These ribozymes also exhibit hydrolytic endonuclease activity on RNA or DNA-modified RNA substrates. However, substantive hydrolytic cleavage of pure DNA is rarely observed within reported intron ribozymes. These clues suggest that uncharacterized intron ribozymes encoded by "propagating" retrotransposons, especially those that naturally lack the IEP components [open reading frame (ORF)–less introns], might hold the potential to cleave pure DNA targets through hydrolysis. These hypothetical ribozymes are called HYdrolytic Endonucleolytic Ribozymes (HYERs). If available, RNA-based HYERs could be employed for gene editing, expanding the current toolbox established by protein enzymes such as TALEN and CRISPR-Cas.

RATIONALE: In our preliminary observation, several compact ORF-less GII-C introns show multiple copies within their resident genomes, suggesting that these introns might be active for genomic propagation including target DNA recognition and cleavage reactions. Bioinformatic mining identified many "propagating" ORF-less GII-C introns as HYER candidates. Biochemical assays confirmed HYER's hydrolytic cleavage activity on RNA and DNA substrates. HYER was active for DNA cleavage in both bacterial and mammalian cells. Cryo–EM analysis revealed the high-resolution structure of HYER, elucidating its catalytic mechanism and providing structural insights for engineering.

RESULTS: Using the primary sequences and secondary structural features of core domains I to III and V to VI in reported GII-C introns, we built the covariance models for systematic mining and identified 93 HYER candidates from the public bacterial genomes. Nine candidates (HYERs 1 to 9) were rationally selected for experimental characterization. HYERs 1 to 6 exhibited sequence-specific endonuclease activity in vitro. Biochemical screening sugge… that the highly active HYER1 and HYER2… both Mg^{2+}-dependent hydrolytic ribozymes, which cleaved the DNA targets in a broad range of biochemical conditions including various temperatures (25° to 50°C) and potassium concentrations (10 to 500 mM). Of note, HYER1 and HYER2 were active for plasmid interference in *E. coli*. Within the reporter cell line containing the out-of-frame *puromycin* gene, HYER1 could generate genomic dsDNA breaks and yield DNA insertions and deletions to restore puromycin resistance. Single particle cryo-EM analysis illustrated that HYER1 forms a C2-symmetric homodimer mediated by palindromes and A-stacking interactions between the two monomers. Each HYER1 monomer contains a 6-nucleotide (nt) target recognition site (TRS) and a hydrolytic pocket for capturing and cleaving DNA substrates. HYER2, which does not have the palindrome, exhibits monomeric architecture and contains a catalytic pocket highly identical to the monomer unit of HYER1. Full complementarity between HYER TRSs and DNA targets is essential for efficient cleavage. Revised TRSs could redirect HYERs to cleave different DNA targets.

Three rational designs, which aimed to improve the DNA-targeting performance by HYER1, were explored. Extending the TRS improved HYER1's DNA recognition specificity but decreased the cleavage activity. The insertion of a 14-nt recruiting sequence (RS), which is sequentially distant but spatially proximal to TRS, increased both the DNA recognition length and the cleavage activity by HYER1. Mutating the palindromes allowed HYER1 to form the heterodimer, which can carry two distinct TRSs and generate customizable dsDNA breaks with cleavage products bearing 5' overhangs, blunt ends, or 3' overhangs.

CONCLUSION: We discovered HYERs as sequence-specific and hydrolytic endonucleases. HYER1 showed plasmid interference activity in bacterial cells and genome-editing capability in mammalian cells. Cryo-EM analysis revealed the homodimer structure of HYER1, which captures the complementary DNA target via TRS and cleaves the target using the hydrolytic mechanism. TRS extension, RS insertion, and heterodimerization improved the specificity and flexibility of DNA targeting by HYER1. As an RNA-based DNA manipulation platform, HYER is highly programmable and holds the potential to be further developed and utilized in multiple genetic scenarios. ■



Hydrolytic DNA cleavage by HYER1. During the DNA cleavage reaction, the DNA is captured by the TRS region of HYER1 via complementary base pairing. Then, the nucleophilic attack by the water molecule within the catalytic pocket leads to the cleavage of DNA substrate. The engineered HYER1 heterodimer, which contains two distinct TRSs, can generate various forms of dsDNA breaks.

The list of author affiliations is available in the full article.
*Corresponding author. Email: junjiegogoliu@tsinghua.edu.cn
†These authors made equal contributions to this work.
Cite this article as Z.-X. Liu *et al.*, *Science* **383**, eadh4859 (2024). DOI: 10.1126/science.adh4859

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adh4859

Downloaded from https://www.science.org at Donghua University on February 01, 2024

RESEARCH ARTICLE

MOLECULAR BIOLOGY

Hydrolytic endonucleolytic ribozyme (HYER) is programmable for sequence-specific DNA cleavage

Zi-Xian Liu[1]†, Shouyue Zhang[1]†, Han-Zhou Zhu[1]†, Zhi-Hang Chen[1]†, Yun Yang[1]†, Long-Qi Li[1]†, Yuan Lei[2,3], Yun Liu[1], Dan-Yuan Li[1], Ao Sun[1], Cheng-Ping Li[1], Shun-Qing Tan[1], Gao-Li Wang[1], Jie-Yi Shen[1], Shuai Jin[2], Caixia Gao[2,3], Jun-Jie Gogo Liu[1]*

Ribozymes are catalytic RNAs with diverse functions including self-splicing and polymerization. This work aims to discover natural ribozymes that behave as hydrolytic and sequence-specific DNA endonucleases, which could be repurposed as DNA manipulation tools. Focused on bacterial group II-C introns, we found that many systems without intron-encoded protein propagate multiple copies in their resident genomes. These introns, named HYdrolytic Endonucleolytic Ribozymes (HYERs), cleaved RNA, single-stranded DNA, bubbled double-stranded DNA (dsDNA), and plasmids in vitro. HYER1 generated dsDNA breaks in the mammalian genome. Cryo–electron microscopy analysis revealed a homodimer structure for HYER1, where each monomer contains a Mg^{2+}-dependent hydrolysis pocket and captures DNA complementary to the target recognition site (TRS). Rational designs including TRS extension, recruiting sequence insertion, and heterodimerization yielded engineered HYERs showing improved specificity and flexibility for DNA manipulation.

Many natural catalytic RNAs, known as ribozymes, catalyze diverse biochemical reactions, such as RNA splicing and polymerization (*1*–*6*). Group II (GII) retrotransposons—specifically the intron RNAs they encode—are well-documented ribozymes capable of catalyzing the transesterification reactions during "copy and paste" retrotransposition in the genome (*7*, *8*). GII retrotransposons are categorized into three classes (A, B, and C) based on the sequence and structural features of the intron RNAs. GII-C introns, being only half the size of GII-A and GII-B introns, are considered more ancestral (*9*–*11*). GII introns typically consist of six domains (fig. S1A) (*12*). Domain I (DI), the largest, facilitates RNA exon recognition through base pairing with the exon binding site (EBS). Domain II and III (DII and DIII) are structural motifs, maintaining the stable ribozyme architecture. Domain IV (DIV) often contains an open reading frame (ORF) encoding an intron-encoded protein (IEP), which houses DNA endonuclease (EN) and reverse transcriptase (RT) domains (*7*, *8*). The crucial nucleotides forming the RNA-splicing core are located in domain V (DV), and the bulged adenosine (A) responsible for the transesterification reaction in splicing resides in domain VI (DVI) (*7*, *12*, *13*).

During retrotransposition, the intron RNA is initially transcribed with the exons in a precursor RNA form (pre-RNA). Assisted by the IEP, intron RNA undergoes self-splicing through two sequential transesterification reactions, resulting in ligated exons and lariat intron RNA (*7*). The lariat intron is then inserted into the DNA sense strand within the propagation target site through two-step reverse splicing. Subsequently, in the target-primed reverse transcription (TPRT) process, the DNA antisense strand is cleaved by the EN domain to prime the intron cDNA synthesis by the RT domain of IEP. After TPRT, the intron RNA template is removed and its DNA copy is pasted into the propagation site by the host's DNA repair system (fig. S1B) (*7*, *8*). Because of the sequence-specific transposition activity, GII retrotransposon consisting of IEP and intron RNA has been repurposed as a bacterial gene manipulation tool, Targetron, which can integrate the intron sequence into a targeted site to disrupt gene functions (*14*–*17*).

Natural GII retrotransposons lacking an ORF have been discovered (*9*, *18*–*20*). These ORF-less GII introns have often been regarded as "degenerate and nonfunctional descendants" of GII retrotransposons and remain poorly characterized. To date, widely used GII intron databases list only 42 ORF-less GII-A and GII-B systems identified by bioinformatic analyses (*9*, *18*–*20*). To our knowledge, ORF-less GII-C introns have not been reported. Our preliminary screening of publicly available bacterial genome sequences led us to discover many GII-C retrotransposons lacking an ORF for IEP encoding. Several of these introns exist in multiple copies in the genome, suggesting that IEP-free retrotransposons might be active and might have propagated through a "copy and paste" mechanism (fig. S1C).

Previous in vitro experiments showed that, when IEP is not provided, the ORF-containing GII introns can recognize and cleave the exons in the pre-RNA through hydrolysis instead of transesterification, producing mature and linearized intron ribozymes (*12*, *13*, *21*, *22*). The linearized intron ribozymes can use water molecules to attack RNA-RNA or DNA-RNA linkages, exhibiting hydrolytic endonuclease activity on the EBS-complementary RNA or DNA-modified RNA substrates (*13*, *23*, *24*). However, to our knowledge, substantive hydrolytic cleavage activity on pure DNA is rarely observed for the reported intron ribozymes.

These clues imply that uncharacterized "propagating" ORF-less GII-C introns, which naturally do not have the IEP, presumably possess the capability to cleave the DNA in the propagation sites by means of hydrolysis, in addition to their RNA cleavage capability for maturation (fig. S1, C and D). We term these hypothesized intron ribozymes as HYdrolytic Endonucleolytic Ribozymes (HYERs), in contrast to reported intron ribozymes showing endonucleolytic activity on DNA through transesterification (fig. S1E). We set out to test our hypothesis and discover HYERs.

Results

Large-scale screening revealed a population of multicopy ORF-less GII-C introns

We focused on the compact GII-C introns for the systematic discovery of HYERs in the public bacterial genome database (*25*, *26*). We used 92 reported GII-C introns and excluded the ORF-residing DIV to construct two sequence and structure-based covariance models, one for DI to DIII and one for DV to DVI, for mining intron candidates (Fig. 1A and data S1 and S2) (*9*, *19*, *20*, *27*). These candidates were required to align with both covariance models with high confidence (E-value $<10^{-5}$) and be shorter than 4000 nucleotides (nt), as even reported introns with ORF are typically less than 4000 nt. In total, we identified 5635 loci in the genomes from the Earth's Microbiomes (*26*).

Based on the assumption that an active retrotransposon was likely to propagate multiple copies in its resident genome (*28*), we used BLASTN to cluster highly similar copies (coverage >90% and percent identity >90%) in the same species-level operational taxonomic units (OTUs) and reasoned that each cluster represented a single intron candidate (Fig. 1A). Among the 5635 loci, 469 GII-C intron candidates with at least two copies in the genomes were identified from 400 OTUs (Fig. 1A and data S3). Similar to reported GII-C introns, newly identified candidates often reside downstream of transcription terminators with conserved hairpin features (fig. S2A) (*8*, *22*, *29*). Furthermore, we constructed a hidden Markov model (*30*) of IEPs to determine whether the intron

[1]Beijing Advanced Innovation Center for Structural Biology, State Key Laboratory of Membrane Biology, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]New Cornerstone Science Laboratory, Center for Genome Editing, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [3]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China.
*Corresponding author. Email: junjiegogoliu@tsinghua.edu.cn
†These authors made equal contributions to this work.

Downloaded from https://www.science.org at Donghua University on February 01, 2024

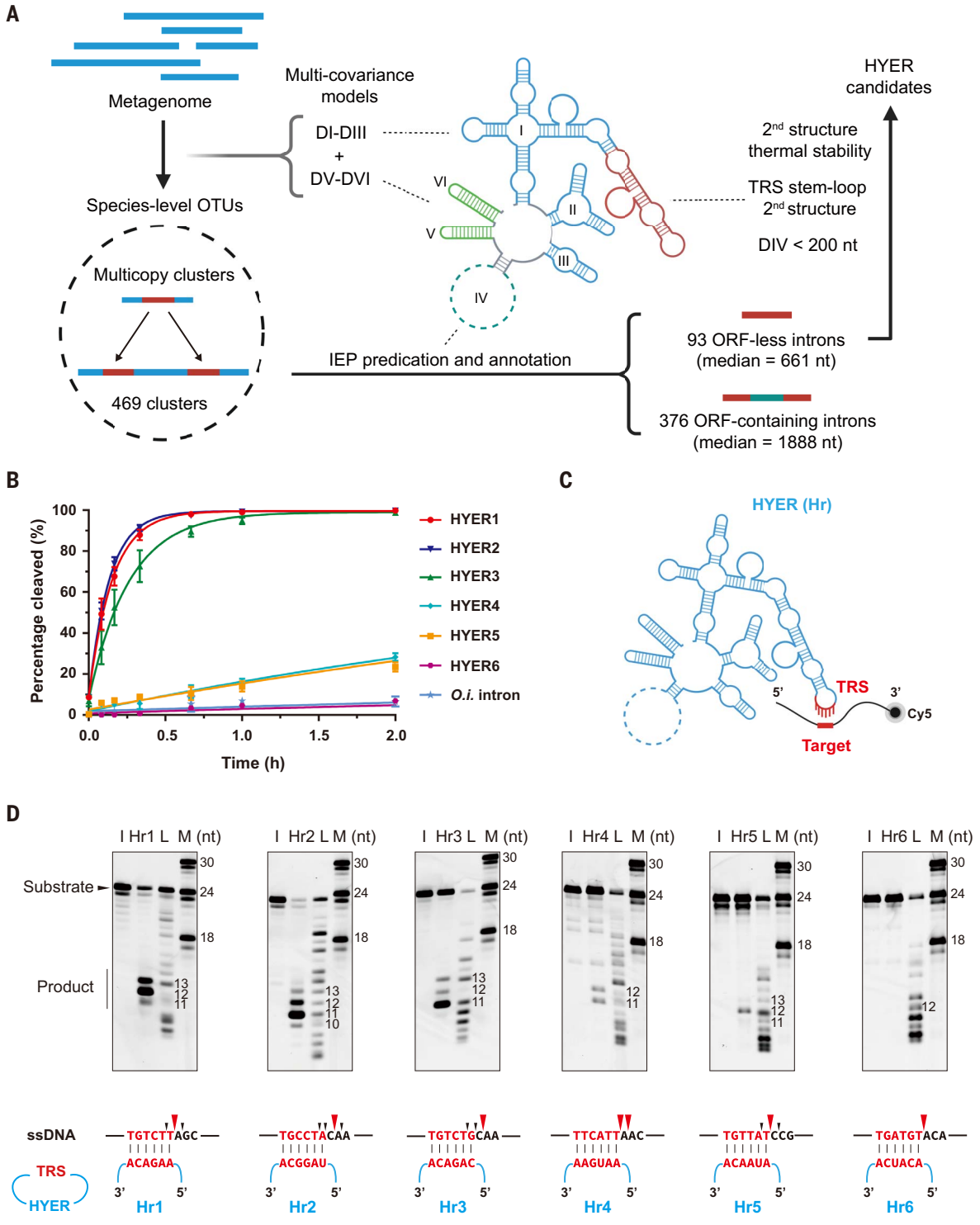


Fig. 1. Identification of ORF-less GII-C introns with DNA cleavage capability. (**A**) Pipeline for the identification of GII-C introns and HYER candidates. HYER is often abbreviated as Hr in following figures and legends for concise presentation. (**B**) The DNA cleavage plot of HYERs on the ssDNA with five target sites (n = 3 each; means ± SD); aliquots were collected at the following time points: 0 min, five min, 10 min, 20 min, 40 min, 1 hour, and 2 hours. The single-turnover rate constants of HYERs 1 to 6 and *O.i.* intron are 6.443 hour^{-1}, 7.466 hour^{-1}, 3.776 hour^{-1}, 0.2151 hour^{-1}, 0.1455 hour^{-1}, 0.1460 hour^{-1}, and 0.02437 hour^{-1}. (**C**) Diagram of in vitro DNA cleavage assay of HYER on substrate with one target. The ssDNA was labeled with Cy5 at the 3′ end. (**D**) Top, cleavage site identification of HYERs 1 to 6; lane I, input of the ssDNA; lane Hr1 to Hr6, cleavage of ssDNA by HYERs 1 to 6; lane L, DNA ladder generated through DNase I digestion; lane M, Cy5-labeled ssDNA marker. Bottom, graphical representation of the cleavage sites of HYERs 1 to 6 on ssDNA; the ssDNA, HYER molecules and TRS are labeled according to (C); red letters show the base pairs between TRS and the target on ssDNA; red and black triangles indicate the major and minor cleavage sites, respectively.

Downloaded from https://www.science.org at Donghua University on February 01, 2024

candidate contains an IEP-coding sequence (ORF) in DIV, which enabled us to distinguish 93 ORF-less introns from 376 ORF-containing introns (Fig. 1A). The length of ORF-less GII-C introns (median 661 nt) was significantly shorter than the ORF-containing GII-C introns (median 1888 nt) ($P < 0.0001$) (fig. S2B). Phylogenetic tree analysis on these newly identified and previously reported multicopy GII-C introns showed that ORF-containing and ORF-less introns could be clustered separately, albeit with some exceptions (fig. S2C). Uniform manifold approximation and projection (UMAP) (*31*) on the multicopy members also revealed the ORF-containing and ORF-less introns to be two distinct groups (fig. S2D). These analyses presumably indicate that ORF-containing and ORF-less introns are two different evolutionary populations. However, UMAP analysis on these introns—excluding the ORF sequences in DIV—showed that both populations are closely related, which suggests that the ORF region is responsible for the major differences between the ORF-containing and ORF-less introns (fig. S2D). To select HYER candidates for biochemical characterization, we identified compact ORF-less GII-C introns that may fold as a stable and active ribozyme, specifically focusing on the top nine intron RNAs with a DIV less than 200 nt, a recognizable EBS in DI, and high thermal stability based on secondary-structure prediction (Fig. 1A, fig. S2C, and data S4).

Identification of active HYERs

Similar to the linearized ORF-containing *Oceanobacillus iheyensis* (*O.i.*) intron (*12*, *13*) without IEP in the reaction, in vitro transcribed HYER candidates generated cleavage products shorter than their RNA targets (fig. S3, A and B), indicative of hydrolytic cleavage, as opposed to transesterification cleavage which produces longer, intron-ligated RNA products (*32*). Six out of nine candidates showed endonuclease activity on the single-stranded DNA (ssDNA) containing five copies of a sequence complementary to the introns' EBS (fig. S3, C and D), which we refer to as target recognition sites (TRSs) in the rest of the study (fig. S3E). Whereas HYERs 4 to 6 exhibited modest cleavage activity, HYERs 1 to 3 were highly active and cleaved DNA within one hour (Fig. 1B and fig. S3F). The *O.i.* intron (without IEP provided) also showed weak DNA cleavage activity comparable to the least active HYER6 (Fig. 1B and fig. S3, C, D, and F). Other newly identified ORF-containing introns, FJAT-27916 sp. intron and *Bacillus_W* sp. (*B.*sp.) intron, which are phylogenetically close to HYER1, were inactive for DNA cleavage (fig. S3F). HYERs 1 to 6 only cleaved DNA targets containing TRS-complementary sequences, suggesting that HYERs are TRS-guided and sequence-specific DNA endonucleases (fig. S3C). To further determine the cleavage sites, we redesigned the ssDNA to contain a single TRS-complementary sequence (Fig. 1C). Sequencing gel analysis of the cleaved products showed that major cleavage sites were located at 0 or 1 nt downstream of the TRS-DNA complementary region (Fig. 1D).

Biochemical characterization of HYERs

Next, we characterized the biochemical properties of the highly active HYER1 and HYER2. Similar to most RNA ribozymes requiring Mg^{2+} to maintain catalytic cores and three-dimensional (3D) structures (*33*–*37*), increased Mg^{2+} concentration promoted the DNA cleavage activity of HYER1 and HYER2 (fig. S4A). Reported ORF-containing GII-C intron ribozymes often prefer high concentrations of monovalent ions such as K^+—which help maintain the stable structures for catalysis (*13*, *38*)—whereas HYER1 and HYER2 displayed enhanced DNA cleavage activity in the buffer with low K^+ concentration (fig. S4B). High Na^+ concentration seemed toxic as it inhibited DNA cleavage by HYER1 and HYER2 (fig. S4C). HYER1 and HYER2 effectively cleaved DNA from 25 to 50°C (fig. S5, A and B). In addition, ssDNA cleavage assay with excess substrate showed that HYER1 and HYER2 were multiple-turnover ribozymes (fig. S5, C and D). Further, we compared the activity of HYER1 and HYER2 with well-known RNA-guided endonucleases (protein-based), SpyCas9 and Cas12s. Each cleavage reaction was performed under the optimal working conditions reported for each enzyme (*39*–*41*). The HYER (1.5 μM) or Cas effector (1.5 μM) ratio to DNA (100 nM) was 15:1, under the single-turnover condition. The SpyCas9 effector [1368-amino acid (aa) protein with 118-nt guide RNA] behaved most efficiently for DNA cleavage. HYER1 (633 nt) and HYER2 (621 nt) showed comparable activity to PlmCas12e effector (978-aa protein with 121-nt guide RNA) and AbCas12l1 effector (867-aa protein with 173-nt guide RNA) (fig. S6, A and B).

We next examined whether HYER1 and HYER2 could cut double-stranded DNA (dsDNA). Cleavage assays showed that they only cleaved the bubbled dsDNA but not the fully annealed dsDNA, implying that the native version of HYER1 and HYER2 could not unwind or invade dsDNA (fig. S6C). We also tested HYERs 3 to 6 and found they were inactive for cleaving fully annealed and bubbled dsDNA (fig. S6D). Because plasmids usually contain denatured bubbles as a result of supercoiled tension (*42*), we tested and found that HYER1 and HYER2 were able to nick a plasmid that contained a TRS-complementary sequence (Fig. 2A). Although inactive for bubbled dsDNA cleavage, HYERs 3 to 6 all exhibited such plasmid nicking capabilities (fig. S6E).

HYER cleaves plasmid and genomic DNA in vivo

To test whether HYERs were active in cells, we adopted the *ccdB* reporter system for activity validation. ccdB toxin interferes with DNA replication in *E. coli* and leads to bacterial death (*43*, *44*). *E. coli* could survive, however, if the *ccdB* plasmid was targeted and depleted instead (Fig. 2B). Nicking and retention of RNA-guided nuclease in the target plasmid has been reported to inhibit plasmid replication, resulting in depletion of the plasmid in cultured bacteria (*45*, *46*). We tested HYERs 1 to 6 and SpyCas9 controls designed to target the *ccdB* plasmid. Similar to wild-type SpyCas9 (wtCas9) and nickase SpyCas9 (nCas9), HYER1 and HYER2 yielded survival clones, but the blank control and HYERs without TRS did not (Fig. 2C). This correlates with reduced ccdB expression due to plasmid depletion by HYER1 and HYER2 (Fig. 2D and fig. S7A). Furthermore, the depletion of the *ccdB* plasmid, which also carries the ampicillin resistance gene (*AmpR*), led to bacterial death in the presence of ampicillin (Amp^+) without ccdB induction (fig. S7B). Growth curve analysis demonstrated that HYER1 exhibited minimal toxicity to bacteria, confirming that cell death under Amp^+ selection was caused by plasmid depletion by HYER1 (fig. S7C). HYERs 3 to 6 were inactive in our assays (fig. S7, D to G).

Although HYERs cannot unwind dsDNA in vitro, active cellular events such as transcription may generate a bubbled window at the target site, allowing DNA access and cleavage by HYERs. We evaluated the ability of HYERs to cleave genomic DNA in a reporter cell line harboring an out-of-frame puromycin resistance gene (*puro**) in its genome. Upstream of the *puro** gene, we positioned a target cassette with three interspaced and identical target sites, all bearing the same TRS-complementary sequence on both top and bottom DNA strands (Fig. 2E). HYER-induced dsDNA breaks are expected to be repaired through endogenous pathways, potentially leading to target-site insertions and deletions (INDELs) that realign the *puro** into a functional ORF (*puro*) which then confers puromycin resistance to the cells (Fig. 2F). Though HYER2 was inactive in mammalian cells, HYER1 yielded many survival clones after puromycin selection. The nontargeting variants of HYER1 and HYER2 (NT-HYER), each harboring a mutated TRS noncomplementary to the target site, failed to produce any survival clones (fig. S8A). In agreement, T7E1 assay (*47*) on PCR products of HYER1-edited genome fragments covering the entire target cassette yielded cleaved DNA bands with expected sizes (fig. S8B). Deep sequencing focusing on each site revealed 9.18% INDEL efficacy for target 1, 7.35% for target 2, and 0.01% (<1%, not appreciable) for target 3 by HYER1 (Fig. 2G). Within the target sites, insertions of 1 to 2 nt and deletions of 1 to 25 nt were ob served (Fig. 2G). Sequencing analysis covering the entire target cassette identified the largest deletion, spanning 85 nt from target 1 to target 2 (fig. S8C). Notably, no INDELs were detected in regions flanking the target cassette, indicating that HYER1 editing is TRS-guided

Downloaded from https://www.science.org at Donghua University on February 01, 2024

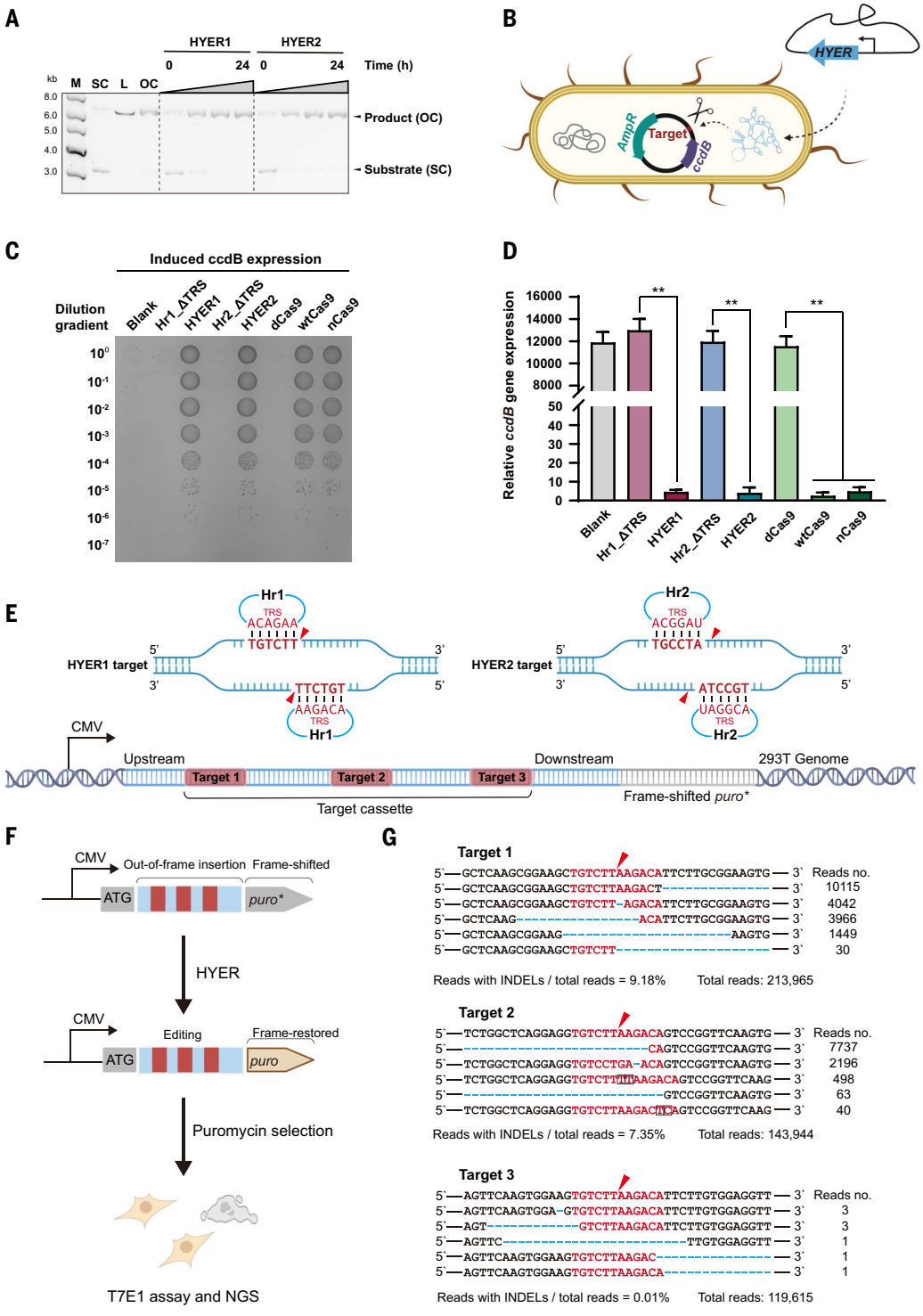


Fig. 2. HYERs cleave DNA in multiple scenarios. (**A**) In vitro cleavage of plasmid. SC, supercoiled plasmid; L, linearized plasmid; OC, open-circular plasmid. Aliquots were collected at the following time points: 0, 3, 8, and 24 hours. (**B**) Schematics of the plasmid interference assay. The HYER plasmid, target plasmid, Ampicillin-resistance gene (*AmpR*) and *ccdB* gene are labeled. (**C**) Survival of bacteria on the culture plate containing 50 μg/mL streptomycin and 10 mM arabinose. The plasmid used for transformation harbored a streptomycin resistance gene and arabinose was used to induce ccdB expression. HYER1 recognizes the TGTCTT site within the *ori* region of the *ccdB* plasmid. HYER2 recognizes the TGCCTA site within the *araC* gene region of the *ccdB* plasmid. The blank column indicates the empty vector; Hr1_ΔTRS, HYER1 without TRS; Hr2_ΔTRS, HYER2 without TRS; dCas9 (dead SpyCas9, D10A, H840A), dCas9 with nontargeting guide RNA; nCas9 (SpyCas9 nickase, H840A), nCas9 with *ccdB* plasmid-targeting guide RNA; wtCas9, wtCas9 with *ccdB* plasmid-targeting guide RNA. Dilution gradient is shown on the left. (**D**) Relative expression of the *ccdB* gene. Labels are described in (C). The significance of relative gene expression analysis was determined through one-way analysis of variance (ANOVA, $^{**}P < 0.01$). (**E**) Schematics of HYER-targeted "out-of-frame" *puro** sequence. The recognition details in each target sites for HYER1 and HYER2 are shown in the top panel. Cleavage of both strands would produce blunt-ended dsDNA breaks. The cleavage sites are indicated by red triangles. A target cassette consisting of three repeated target sites (target 1, target 2, and target 3) is drafted in the bottom panel, followed by a 3 *n*+1 frame-shifted *puro**. The whole sequence is driven by the CMV promoter and stably integrated into the HEK293T genome. (**F**) Workflow of HEK293T-*puro** cell-editing assay by HYER. The frame-restored puromycin resistance gene is indicated as *puro*. (**G**) Alignment of top 5 reads with INDELs generated by HYER1 at three target sites ordered by abundance, with number of reads (reads no.) shown on the right. Red letters show the HYER1 target sites and red triangles mark the cleavage sites. Dashes and boxed letters indicate deleted and inserted bases, respectively.

and confined to specific target sites (fig. S8D). To further test the unenriched editing efficacy, we collected genomic DNA from HYER1-transfected cells before puromycin selection (fig. S9A). Edited bands were not detected via T7E1 assay (fig. S9B). The INDEL rates revealed by deep sequencing were minimal, ranging from 0.09 to 0.2% for the three target sites, suggesting a limited genome editing capability of native HYER1(fig. S9, C and D). Collectively, our results indicated that TRS-guided HYER was capable of targeting and cleaving DNA in cells. However, the genome-editing efficacy in mammalian cells of the native version is weaker than that of protein-based platforms such as CRISPR-Cas.

The overall architecture of HYER

To elucidate the DNA cleavage mechanism of HYER and provide information for potential engineering, we analyzed the structure of HYER1

Downloaded from https://www.science.org at Donghua University on February 01, 2024

using cryogenic electron microscopy (cryo-EM) (fig. S10). Unlike the reported intron ribozymes that often have monomeric status, HYER1 forms a homodimer (Fig. 3A, fig. S11, and movie S1). 3D classification revealed two structural populations of HYER1 dimers: one in an open conformation and one in a closed conformation (fig. S10C). HYER1 in the closed conformation exhibits a C2 symmetry. A refined EM map in the closed conformation at 3.0 Å resolution was obtained for ab initio atomic model building and structural elucidation (figs. S10C, S11, and S12, A and B).

Based on the map, DI, DII, DIII, DV, and DVI form a stable core with clearly visible and high-resolution density (Fig. 3A and fig. S10, C and D). DIV forms a semi-ring structure and anchors around the stable core, leading to a flexible local density with limited resolution in the EM map (Fig. 3A and fig. S10D). We therefore modeled the entire stable core and the confident regions within DIV (fig. S12A).

Canonical structural elements of the GII-C intron that maintain the overall architecture of each monomer, such as the α-α', γ-γ', ϵ-ϵ', λ-λ', θ-θ', κ-κ', and ζ-ζ' interactions (*12*), are observed in HYER1 (fig. S11). The bulged A627 forms a stack interaction with G1 in the linearized HYER1 instead of covalently ligating to G1 as seen in lariat introns (fig. S12B) (*7*, *8*). HYER1 contains several other distinct structural features (fig. S12). In the recently reported ORF-containing *Eubacterium rectale* (*E.r.*) intron, the interaction between DIV and DI is mediated by IEP (fig. S12C) (*8*). However, each monomer of HYER1's DIV directly interacts with DI, displaying an anchor-back architecture (Fig. 3, B and C, and fig. S12, B and C). This anchor-back region is stabilized by multiple hydrogen bonds such as the G·A-pairing by G206 (DI) and A471 (DIV), the triplex-pairing by U473 (DIV), A207 (DI), and U469 (DIV), as well as stacking interactions (Fig. 3C). Moreover, within DIII of HYER1, the $U_{361}CUAGA_{366}$ palindrome provides six continuous base pairs between the two monomers and mediates dimerization (Fig. 3, B and D, fig. S10, and fig. S12B).

Primary sequence analysis revealed seven distinct palindromes within DIII of 45 newly identified ORF-containing and ORF-less introns, and the sequence UCUAGA is shared by 17 introns including HYER1 and HYER5 (fig. S13A). This suggests that the palindrome might be a common feature for a particular population of GII-C introns. Among the six active HYERs, HYER1 and

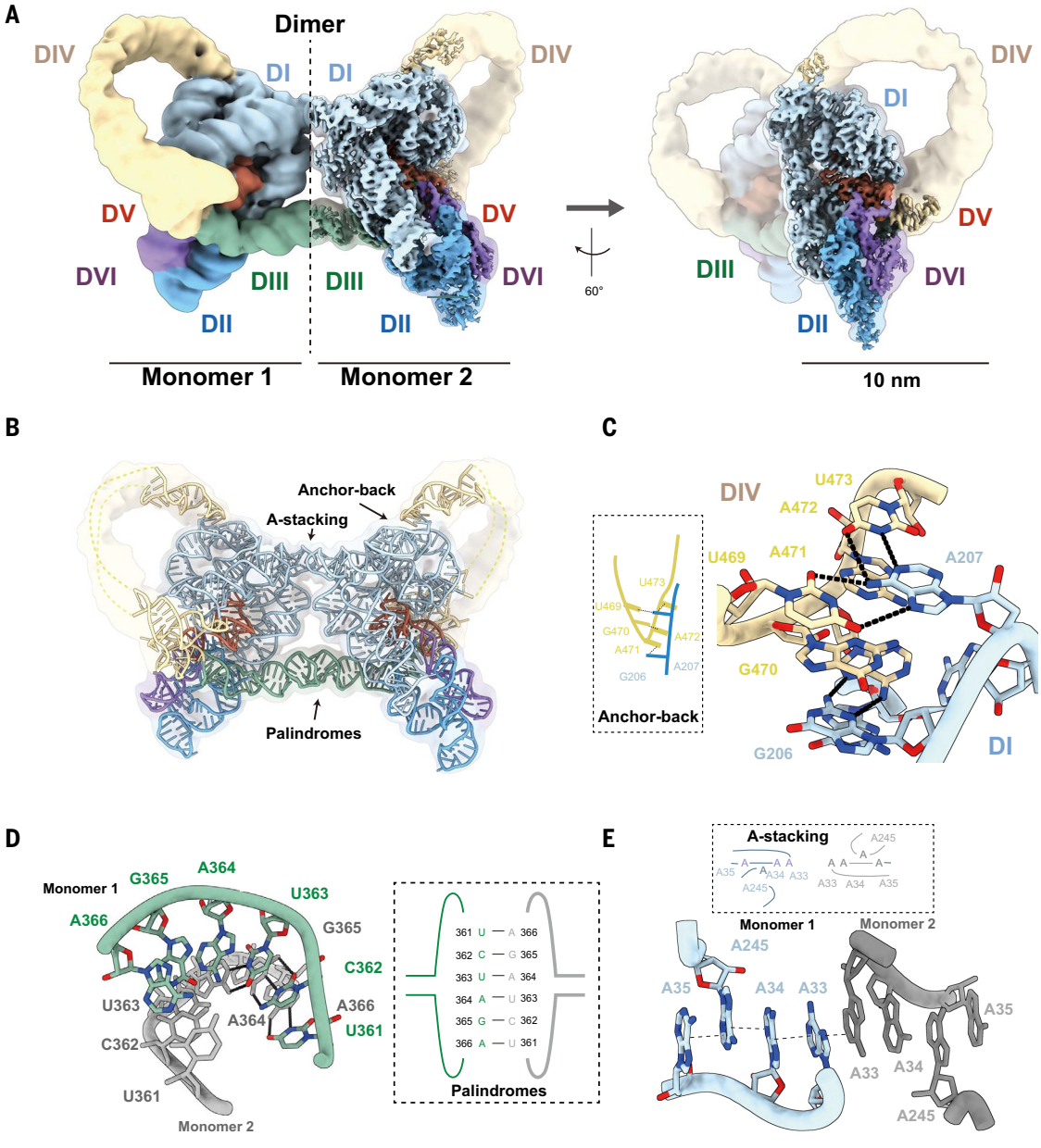


Fig. 3. The overall structure of HYER1. (**A**) Cryo-EM maps of HYER1 in closed conformation. Both the front and side views are shown. DI is light blue, DII is navy blue, DIII is light green, DIV is yellow, DV is red, and DVI is purple. For clear presentation of both the high-resolution details and the overall architecture, the 3.0-Å map is fitted into the 6-Å low-pass-filtered map. Monomer 2 in the low-pass-filtered map is semi-transparent. (**B**) Atomic model of HYER1. The model is fitted into a 6-Å low-pass-filtered transparent map. (**C**) Molecular details and illustrated model of the anchor-back region. The nucleotides involved in the anchor-back interaction are labeled. (**D**) The molecular details and diagram of the $U_{361}CUAGA_{366}$ palindrome. The palindrome in each monomer is colored in green or gray, respectively. (**E**) The molecular details and illustrated model of the A-stacking region. The nucleotides from each monomer are colored in light blue or gray, respectively.

Downloaded from https://www.science.org at Donghua University on February 01, 2024

HYER5—which contain identical palindromes—both displayed dimerized architectures based on size exclusion chromatography (SEC) and cryo-EM analysis (fig. S13). HYERs 2, 3, 4, and 6, without palindromes in DIII, all exhibited monomeric architectures (fig. S13, B and C). 3D refinement and modeling of the monomeric HYER2 revealed a similar overall structure to a HYER1 monomer but with some differences including additional interactions formed by DIII (A351 and U353), DI (A144), and DV (C562-G563 and G583-U585), which may further stabilize the catalytic core in HYER2 (fig. S14). Although mutating the palindromes disrupted the dimerization of HYER1 monomers, it had a minimal effect on the DNA cleavage activity both in vitro and in bacteria (fig. S15). Whether this palindrome-mediated dimer architecture has further biological significance in HYERs' native hosts remains to be studied.

Beyond the palindromes, an 8-nt stacking region was formed by A35-A245-A34-A33 (DI of monomer 1) and A33-A34-A245-A35 (DI of monomer 2), which stabilizes the HYER1 dimer in a closed conformation with a C2 symmetry (Fig. 3, B and E, and figs. S11A and S12B). However, the stacking interaction seems to be lost in the unmodeled open conformation map, allowing the detachment of DIs within the two monomers and resulting in a flexible and asymmetric dimer (fig. S10C).

Molecular details of DNA recognition and cleavage by HYER

Within HYER1, we observed a catalytic pocket built by DI (G1-G5 and A106-G107), DII (J2/3 linker A337-C339), and DV (C565, catalytic triad C566-C568, G582-U583, 2-nt bulge A584-C585), similar to ORF-containing introns (Fig. 4, A and B, and fig. S12A) (*8*, *12*, *25*, *48*–*50*). To visualize how DNA is captured and cleaved, excess oligos that mimic the pre-cleaved DNA (5′-T_1GTCTT$_6$T$_7$CC-3′) and postcleaved DNA (5′-T_1GTCTT$_6$-3′) were incubated with HYER1 at 8°C for 10 min for cryo-EM analysis (figs. S16 and S17). The refined maps suggest that the pre-cleaved and postcleaved DNA oligos stably reside in the catalytic pocket (Fig. 4, A and B, and figs. S16 and S17). Within this pocket, the DNA is captured by TRS (A_{181}AGACA$_{186}$) in DI through 6-bp continuous complementarity, and the DNA-DNA linkage between dT6 and dT7 is further presented to the reaction center formed by the catalytic triad (C566-C568) and 2-nt bulge (A584 and C585) in DV (Fig. 4C and movie S2). The single mismatch within the 6-bp pairing region decreased or eliminated the DNA cleavage activity, indicating precise DNA recognition by HYER1 TRS (Fig. 4D and fig. S18).

Two extra densities in the catalytic center can be observed which are presumed to be the catalytic Mg^{2+} (M-1 and M-2) (*12*, *50*). This indicates that HYER1 might hydrolyze DNA using the canonical two-metal mechanism, in which the two Mg^{2+} can recruit a water molecule as a nucleophile to attack the scissile phosphodiester bond in the DNA backbone (Fig. 4, B and C) (*12*, *50*). In addition to the often-observed Mg^{2+} coordination interactions in ORF-containing introns, U633 in HYER1 DVI also interacts with M-2, which may help to further stabilize this catalytic Mg^{2+} in the reaction center for DNA hydrolysis (Fig. 4, B and C) (*7*, *8*). Structural comparison focusing on the monomer cores shows very similar overall architectures and catalytic pockets [root mean square deviation (RMSD) <1.2 Å] among the apo- and cleavage states, indicating the stable structural folding of the HYER1 monomer, which might also be related to its robust DNA cleavage activity (fig. S19 and movie S2). For apo-HYER1, 61% of the refined particles are in open conformation, and this percentage of the particles in the open conformation increases to 63 and 73% for pre- and postcleavage states, respectively (figs. S10, S16, and S17).

HYER is programmable for sequence-specific DNA cleavage

Our biochemical and structural results confirmed the essential role of DNA recognition by TRS, which is structurally exposed and stays in a single-stranded state in either dimeric or monomeric HYERs (Fig. 5A and fig. S14, A and B). This suggested that the target specificity of HYERs was programmable. To test this, we mutated the TRS sequences of HYERs 1 to 6 and redirected them to cleave new DNA targets (fig. S20A). For HYER1 specifically, an extensive activity comparison was conducted, demonstrating that all four TRS-reprogrammed constructs cleaved their complementary DNA, with two showing more than twofold increased activity compared with native HYER1 (Fig. 5B and fig. S20B). To explore whether the length of TRS could be manipulated, we designed HYER1 without TRS or with 2, 4, 8, 10, 12, and 14-nt TRS. Although the activity was reduced, constructs with 8 to 12-nt TRS still cleaved the DNA targets in vitro (Fig. 5C and fig. S20, C and D). Similar to native HYER1, constructs with 4, 8, and 10-nt TRS showed plasmid depletion activity in bacteria (fig. S20, E to H). Structural analysis indicated that HYER1 with 10-nt TRS had the same dimer architecture and catalytic pocket as native HYER1 but contained a more flexible TRS due to the length extension (fig. S21). Moreover, every single mismatch in the 10-nt TRS pairing region eliminated the plasmid interference activity in bacteria, suggesting a precise DNA recognition contributed by all nucleotides in the extended TRS of HYER1 (fig. S22).

To expand the DNA-recognition ability beyond the TRS, we substituted A223 with a 14-nt recruiting sequence (RS), positioned next to TRS (Fig. 5D and fig. S23). RS-inserted HYER1 showed increased DNA cleavage activity compared with native HYER1 (Fig. 5, E and F). Of note, cleavage was abolished when the DNA matched only the TRS but not the RS (Fig. 5, E and F). We hypothesized that RS might reshape the DNA-binding pocket to influence the role of TRS, becoming essential for DNA recruitment and cleavage.

HYER1 naturally forms a homodimer mediated by the palindromes and contains two identical catalytic pockets and TRS sequences (Fig. 5A). We mutated the palindrome to UCUCCU in monomer 1 (M1*) and AGGAGA in monomer 2 (M2*) (Fig. 5G). Additionally, we redesigned the TRS sequences to create two distinct TRSs (TRS1 in M1* and TRS2 in M2*) (Fig. 5G). Although M1* and M2* could not form dimers individually, combining them in an equimolar ratio yielded a stable M1*-M2* heterodimer (Fig. 5, H and I). By designing the sequences of bubbled dsDNA, M1*_TRS1 and M2*_TRS2 could selectively target the top and bottom DNA strands (Fig. 5J). Through controlling the length of the spacer sequence between the two cleavage sites, the M1*-M2* heterodimer generated cleavage products with 5′ overhangs, blunt ends, or 3′ overhangs (Fig. 5J and K). Our results suggest that the HYER1 heterodimer provides improved DNA manipulation capability.

Discussion

In this study, we obtained a population of multicopy intron RNAs, called HYERs, capable of DNA cleavage through hydrolysis instead of transesterification. However, whether HYERs naturally cleave DNA in their hosts remains unclear, and whether such HYER-encoding retrotransposons propagate through retrotransposition or other mechanisms such as gene conversion has yet to be explored (*51*–*53*). If HYERs behave similarly to canonical retrotransposons, understanding what components in the native hosts may catalyze the reverse transcription of HYER RNAs requires additional investigation.

Unlike many reported GII ribozymes, which prefer the high cation concentration and extreme temperature for catalysis (*10*, *12*, *21*, *22*), HYER seems tolerant to a wide range of cation concentrations and temperatures and cleaves both the RNA and DNA. The highly active HYER1 cleaved plasmid in the buffer with 10 mM Mg^{2+} and 150 mM K^+, showed plasmid-interference activity in bacteria and generated dsDNA breaks in the mammalian genome. Moreover, by combining biochemical and structural analysis, we elucidated the working mechanism by HYER, which is comparable to the well-known RNA-guided CRISPR-Cas nucleases that use a reprogrammable single-stranded RNA for target recognition and cleave the target through hydrolysis (*39*–*41*). These results suggest the potential of HYER as DNA and RNA manipulation tools without ligating itself into the target site. HYERs (~0.6 kb) are much more compact than widely used CRISPR nucleases

Downloaded from https://www.science.org at Donghua University on February 01, 2024

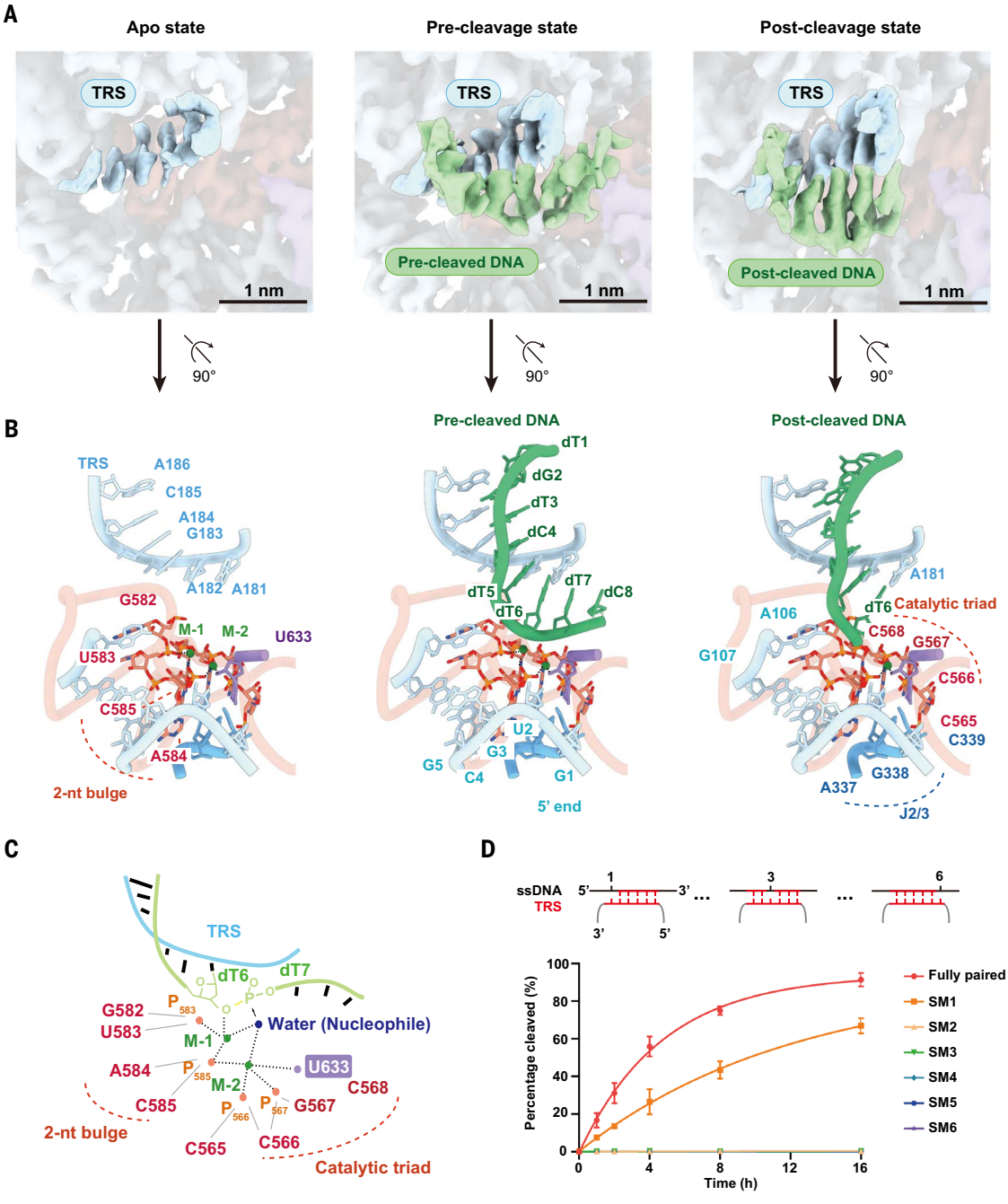


Fig. 4. The structural mechanism of DNA cleavage by HYER1. (**A**) Cryo-EM maps of HYER1 in the apo state, precleavage state, and postcleavage state. DNA density is colored in green and TRS density is colored in light blue. Other domains are indicated in semi-transparent colors according to Fig. 3A. (**B**) The atomic models of HYER1 catalytic core in the apo state, precleavage state and postcleavage state. The different domains are colored according to Fig. 3A. Additionally, the illustrated model of DV is rendered transparent and the Mg^{2+} (M-1 and M-2) in forest green. (**C**) Model of DNA hydrolytic cleavage. The Mg^{2+} ions are displayed as green dots. The phosphate groups coordinating the Mg^{2+} are displayed as orange dots and linked by the gray lines with their associated RNA nucleotides. The nucleophilic water is displayed as a navy blue dot. The TRS (light blue) and DNA (green) are paired for 6 bp. The scissile phosphodiester bond between dT6 and dT7 is specifically shown in detail. (**D**) The cleavage plot of HYER1 with 6-nt TRS on DNA with single mismatches (n = 3 each; means ± SD). Aliquots were collected at 0, 1, 2, 4, 8, and 16 hours. The TRS and paired DNA are labeled in red and unpaired bases in DNA are labeled in black. Single mismatches at 6 positions are indicated as SM1-6 and labeled on the DNA strands. The single-turnover rate constants for the cleavage of fully paired and SM1 substrates are 0.2112 $hour^{-1}$ and 0.08238 $hour^{-1}$ (2.6-fold decrease for SM1 substrates compared with the fully paired substrates), whereas other substrates are not cleaved.

(3 to 4 kb) (*39*–*41*, *54*–*57*). The HYER molecules are also considerably easier to obtain by in vitro transcription than protein nucleases that are often expressed and purified from cells with complex protocols and considerable effort. Whereas CRISPR-Cas typically recognizes a 20-nt DNA region using a guide RNA spacer, native HYER only recognizes a 6-nt DNA region using a TRS element, limiting HYER's performance for precise genome editing. Our structure-based engineering extended the DNA recognition length up to 20 nt by TRS extension and RS insertion. The heterodimer design with two distinct TRSs also doubled the length of DNA recognition and yielded engineerable cleavage patterns on bubbled dsDNA, suggesting HYER as a highly engineerable platform with considerable room for improvement. Currently, the in vivo editing activity of native HYERs is minimal, most likely a result of its limited capacity of dsDNA unwinding. However, as seen for many protein nucleases, such as CRISPR-Cas9 and CRISPR-Cas12f1, whose genome editing efficacies have been vastly enhanced after intensive engineering (*58*, *59*), advanced structural engineering, integration of functional RNA elements for recruiting endogenous helicases, directed evolution, and screening of new candidates or RNA modules capable of dsDNA unwinding can be applied to improve HYER's performance in various applications.

The DNA cleavage activity of the ORF-less introns identified in this study is normally

Downloaded from https://www.science.org at Donghua University on February 01, 2024

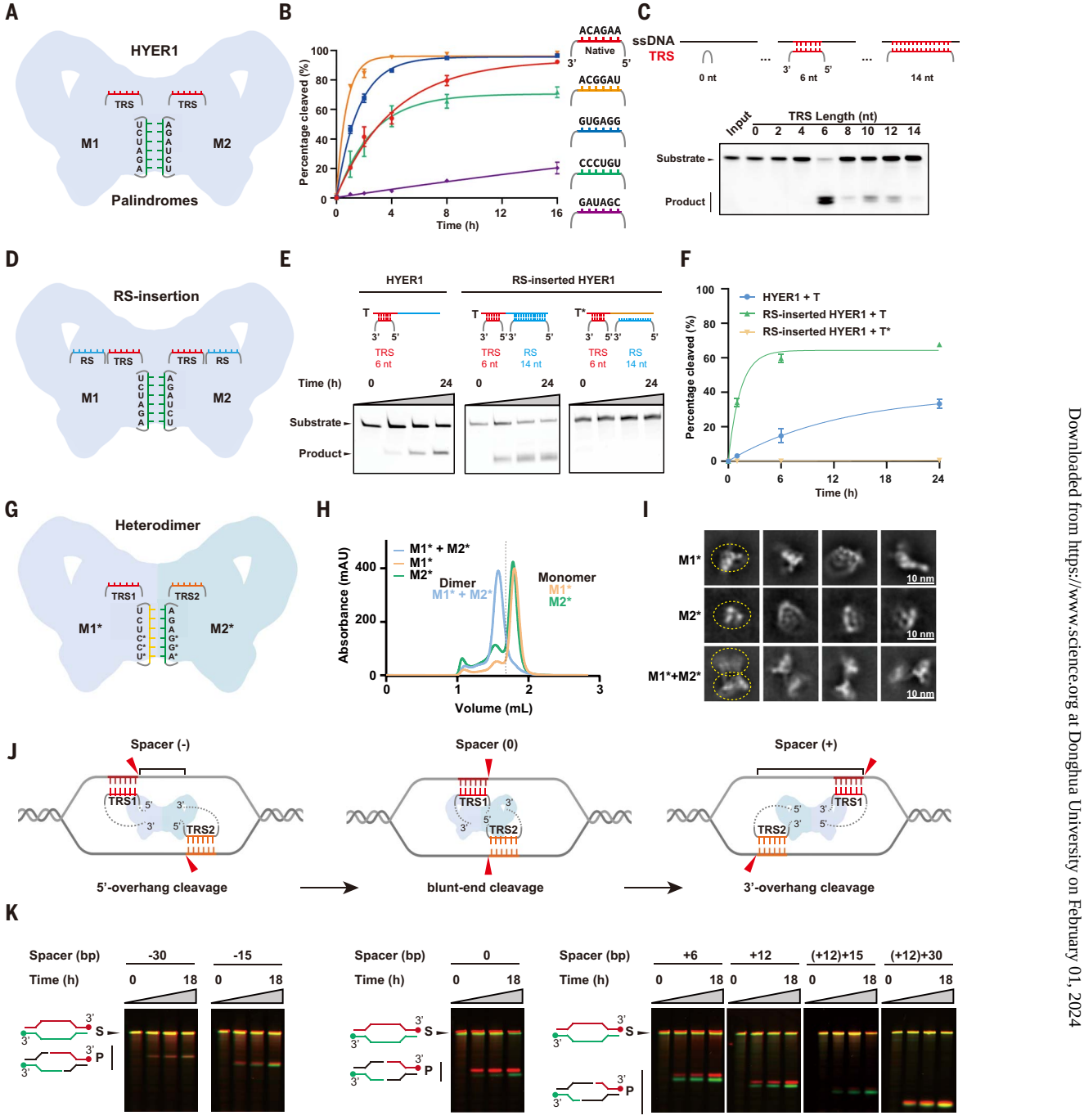


Fig. 5. Engineering of HYER1. (**A**) Illustrated model of HYER1 homodimer. The TRS and palindromes are outlined and labeled. (**B**) The DNA cleavage plot of HYER1 with reprogrammed TRS (n = 3 each; means ± SD). Aliquots were collected at 0, 1, 2, 4, 8, and 16 hours. The sequences of tested TRSs are shown in the right panel. The single-turnover rate constants of HYER1 with UAGGCA, GGAGUG, UGUCCC, and CGAUAG TRS are 1.393 hour^{-1}, 0.6206 hour^{-1}, 0.3815 hour^{-1} and 0.01634 hour^{-1} [5.9-fold increase, 2.6-fold increase, 1.6-fold increase and 14.6-fold decrease compared with the native TRS (0.2380 hour^{-1})]. (**C**) The cleavage assay of HYER1 with different lengths of TRS. HYER1 and ssDNA were incubated at 37°C for 16 hours. The complementary nucleotides within TRS and DNA are labeled in red. (**D**) Illustrated model of RS-inserted HYER1. (**E**) Target and nontarget DNA cleavage assay of HYER1 and RS-inserted HYER1, aliquots were collected at 0, 1, 6, and 24 hours. T, the ssDNA with TRS and RS-complementary sequences. T*, the ssDNA with TRS-complementary sequence and RS-noncomplementary sequence. (**F**) The cleavage plot of HYER1 and RS-inserted HYER1 (n = 3 each; means ± SD). The single-turnover rate constants of HYER1 on T, RS-inserted HYER1 on T, and RS-inserted HYER1 on T* are 0.03968 hour^{-1}, 0.7296 hour^{-1} and 6.17 × 10^{-5} hour^{-1}. (**G**) Illustrated model of the HYER1 heterodimer. (**H**) The size exclusion chromatogram of HYER1 monomers and heterodimer. The UV254 absorbance curve is respectively colored for each sample. (**I**) The 2D class averages of monomers and heterodimer. The monomer units are highlighted with dashed circles. (**J**) Diagram of bubbled dsDNA cleavage by HYER1 heterodimer. The cleavage sites are indicated by red triangles. (**K**) Cleavage of bubbled dsDNA with various spacer lengths between the two cleavage sites. The lengths of spacers are indicated on the top. The top and bottom strands are labeled with Cy5 (red dots) or FAM (green dots) on the 3' end, respectively. The DNA strands with fluorophores are colored accordingly. Fluorescence channels (Cy5, red; FAM, green; merge, yellow) are merged through ImageJ. S, substrate. P, product. Aliquots were at 0, 3, 6, and 18 hours.

Downloaded from https://www.science.org at Donghua University on February 01, 2024

performed by the IEP encoded in DIV of the ORF-containing intron. In the ORF-containing intron complex, IEP connects DIV and DI to maintain the stable architecture of the retrotransposon complex whereas in the ORF-less intron HYER1, DIV itself can interact with DI, achieving an overall architecture similar to the IEP-bound introns (*7*, *8*, *60*). Inspired by the RNA world hypothesis that envisions RNA as the original form of life, we speculate that during evolution, structural and catalytic functions of RNAs were gradually taken over by proteins (*61*–*64*). In the case of the GII-C intron system, we envision a multi-step molecular evolution (fig. S24). (i) In the primitive RNA world, intron RNA played roles such as maturation, new site recognition, and insertion in the RNA genome. (ii) When DNA began to emerge, genetic information storage in the form of DNA became more stable and intron RNA might have adapted to propagate in DNA-RNA chimeric or pure DNA genome, although not as efficiently. (iii) Continuing evolution, especially the extension of DIV, gradually generated ORF-containing DIV, which were able to code for short peptides or proteins. These peptides or proteins then acted as *cis*-elements to interact with DIV and enhance intron RNA's structural stability and catalytic activity. (iv) As evolution went further, the ORF of DIV became longer and more mature and the protein it encoded not only played the function of stabilizing the structure, but also acquired DNA-cleavage and reverse transcription activities to functionally replace the RNA ribozymes, as seen in the current ORF-containing GII-C systems that actively propagate in bacterial genome through retrotransposition (fig. S24) (*7*, *25*, *65*, *66*).

Materials and Methods

Group II-C intron identification and phylogenic analysis

Multi-covariance models were built based on 92 reported GII-C systems (*9*, *19*, *20*). The GII-C sequences were first divided into two parts, DI to DIII and DV to DVI, based on their secondary structures. After removing duplicated sequences, we performed sequence alignments using Clustal Omega (1.2.4) (*67*). Subsequently, covariance models for DI to DIII and DV to DVI were independently constructed based on the sequence alignment results using INFERNAL (1.1.4) with default parameters (*27*). The covariance models are listed in data S1 and S2.

Candidates that matched both models with E-value $<10^{-5}$ and were shorter than 4000 nt were selected from the genome of 45,599 species-level operational taxonomic units (OTUs) in the Earth's Microbiomes catalog (*26*). The hidden Markov model of ORF-encoded IEPs was built using HMMER (3.3.2) based on identified IEP sequences from reported GII-C systems (*30*). Proteins encoded in the metagenome were predicted by Prodigal (2.6.3) (*68*).

To determine the multicopy introns, we employed BLASTN (2.12.0+) (*69*) to cluster and find highly identical copies (both query coverage per subject and percentage of identical matches higher than 90%) in each OTU. The significance of GII-C intron size analysis was determined through Mann-Whitney U test. The sequences of newly identified multicopy GII-C introns are listed in data S3.

To identify more suitable candidates for experimental evaluation, the subsequent filtering criteria were implemented: (i) The DIV was restricted to be less than 200 nt in order to obtain more compact introns, 64 introns left. (ii) Secondary structures predicted by RNAfold (2.4.17) were manually screened to find candidates with clearly recognized TRS loop, 25 introns left (*70*). (iii) Thermal stabilities were ranked among these 25 introns based on secondary structure prediction (ensemble diversity normalized by length using RNAfold) to identify introns exhibiting high stability. Following the application of these criteria, the top nine introns, which possess compact sizes, exhibit clear TRS structures and demonstrate high thermal stability, were selected as HYER candidates for the wet-lab characterization (data S4).

The phylogenetic tree reconstruction was performed with the full sequences of 469 multicopy introns identified in this study and 54 reported multicopy introns using IQ-TREE2 (*71*) and iTOL (*72*), based on the alignments by Clustal Omega (1.2.4) (*67*).

To identify the 469 multicopy intron RNAs with or without ORFs, a 5-mer frequency matrix was constructed by Jellyfish (*73*). The log normalization and dimension reduction were conducted by Seurat (4.3.0) (*74*). The top 10 principal components (PCs) were selected for dimension reduction by UMAP (*31*).

Identification of the conserved 5′-exon hairpin

The secondary structures of the upstream 50-nt sequences of the multicopy intron RNAs were predicted by RNAfold. We then analyzed the occurrence frequency of the 5′ paired, 3′ paired, and nonpaired bases at each position to indicate the secondary structure conservation. The graphical representation of sequence conservation was generated by WebLogo (2.8.2) (*75*, *76*).

Plasmid construction

The RNA sequences of HYERs 1-9, *O.i.* intron, FJAT-27916 sp. intron and *B.*sp. intron are listed in table S1, and the genes were ordered from Sangon Biotech or Tsingke Biotechnology. The mutations in the palindrome or the TRS of HYERs were introduced through site-directed mutagenesis by PCR. For bacterial plasmid interference, the *ccdB* toxic plasmid (p11-LacY-wtx1, Addgene no. 65777) was kindly provided by W. Li in the Institute of Zoology, Chinese Academy of Sciences. The trc promoter activated Cas9, nCas9 (H840A), and dCas9 (D10A and H840A) gene cassettes and the J23119 promoter activated guide RNA (spacer: GCGAUAAGUCGUGUCUUACC) cassette were inserted into the pCDFDuet-1 vector by Gibson assembly, and the original T7 promoters in pCDFDuet-1 were removed. The J23119 promoter activated ribozyme cassettes were inserted into the pCDFDuet-1vector by Gibson assembly, and the original T7 promoters in pCDFDuet-1 were removed. The PiggyBac integration plasmid and transposase plasmid were kindly provided by N. Liu in the School of Life Sciences, Tsinghua University. Plasmids and related details are listed in table S2.

In vitro transcription

T7 RNA polymerase promoter upstream of the DNA that encodes RNA was added by PCR. The PCR-amplified products were purified using VAHTS DNA Clean Beads (Vazyme) and then used as the templates for in vitro transcription (IVT). All reactions were performed in IVT buffer (30 mM Tris 8.1, 25 mM $MgCl_2$, 0.01% Triton X-100, 2 mM spermidine, 5 mM DTT) with 5 mM of each NTP, 0.01 mg/ml RNase inhibitor and 0.02 mg/ml T7 RNA polymerase. After DNase I digestion followed by proteinase K treatment to remove DNA templates and proteins, the transcribed products were washed and concentrated using 100 kDa molecular weight cut-off centrifugal filters (Merck Millipore) in a buffer containing 50 mM MOPS (pH 7.5), 10 mM $MgCl_2$ and 15% glycerol. The HYERs were then flash-frozen in liquid nitrogen and stocked in −80°C. The RNA samples were usually stocked with a concentration of 20 μM. The RNA quality was examined by electrophoresis using a denaturing 8% Urea-PAGE gel. The RNA sequences and related descriptions are listed in table S1.

In vitro cleavage assays for screening HYER candidates

The 3′-NH_2 labeled ssRNA was ordered from RuiBiotech and were further labeled with Cy5 through the reaction between -NH_2 and Cy5-NHS. 3′-NH_2 labeled ssDNA was ordered from Sangon Biotech and labeled with Cy5 following the same protocol. For the cleavage assay of ssRNA and ssDNA with 5 copies of target site, 0.1 μM substrates were incubated with 1.5 μM HYERs at 50°C in a buffer containing 125 mM $MgCl_2$, 500 mM NH_4Cl, 40 mM MOPS (pH 7.5) and 100 ng/μL BSA. The reaction was terminated by adding EDTA, which was 1.25 times as much as the amount of Mg^{2+}. For the ssRNA cleavage assay, formamide was directly added to a final concentration of 50% before incubation at 95°C for 1 min and cooled down on the ice. For ssDNA cleavage assay, RNase digestion and proteinase K treatment were

Downloaded from https://www.science.org at Donghua University on February 01, 2024

performed before the addition of formamide. The reaction samples were analyzed in 20% Urea-PAGE for ssRNA cleavage and 15% Urea-PAGE for ssDNA cleavage. The gels were visualized using Amersham Typhoon 5 (GE Healthcare). The percentage cleaved was quantified using ImageJ and the curves were plotted using the One-Phase-Association model $P = P_0 + (Plateau - P_0) \times (1 - e^{-k \times t})$ in GraphPad Prism 7 [P indicates the percentage cleaved (%), P_0 indicates the percentage cleaved (%) at t = 0 hours, Plateau indicates the maximum percentage cleaved (%), k indicates the rate constant ($hour^{-1}$), t indicates the time of the cleavage reaction (hour)]. All experiments were performed with at least three replicates. Oligonucleotides used in this study and related descriptions are listed in table S3.

Cleavage site determination

The 3′-NH_2 labeled ssDNA marker and ssDNA with one target site was ordered from Sangon Biotech and labeled with Cy5 as described above. The ssDNA random hydrolysis ladder was prepared by DNase I treatment of Cy5-labeled ssDNA. The HYER and ssDNA with one target site were incubated at 50°C for 24 hours in the condition described above and analyzed by electrophoresis using 20% Urea-PAGE. The gels were visualized using Amersham Typhoon 5 (GE Healthcare). Oligonucleotides used in this study and related descriptions are listed in table S3.

In vitro cleavage assays for biochemical characterization of HYERs

To investigate the effect of Mg^{2+} concentration on HYERs catalytic capability, 0.1 μM ssDNA with one target site was incubated with 1.5 μM HYERs at 37°C in a buffer containing 10 to 125 mM $MgCl_2$, 150 mM KCl, 40 mM MOPS (pH 7.5) and 100 ng/μL BSA. To investigate the effect of monovalent ion concentration on HYERs' catalytic capability, 0.1 μM ssDNA was incubated with 1.5 μM HYERs at 37°C in a buffer containing 50 mM $MgCl_2$, 40 mM MOPS (pH 7.5), 100 ng/μL BSA and 10 to 500 mM KCl or NaCl. For cleavage assay at different temperatures, 0.1 μM ssDNA was incubated with 1.5 μM HYERs at 25°C, 37°C, 42°C, 50°C or 60°C in a buffer containing 50 mM $MgCl_2$, 40 mM MOPS (pH 7.5), 100 ng/μL BSA and 10 mM KCl. The reactions were terminated as described above and the reaction samples were analyzed using 20% Urea-PAGE. The gels were visualized and analyzed as described above. All experiments were performed with at least three replicates. Oligonucleotides used in this study and related descriptions are listed in table S3.

DNA cleavage activity comparison among HYERs and CRISPR-Cas nucleases

For DNA cleavage by HYERs, 0.1 μM ssDNA was incubated with 1.5 μM HYERs at 37°C in a buffer containing 50 mM $MgCl_2$, 10 mM KCl, 40 mM MOPS (pH 7.5) and 100 ng/μL BSA. The reaction samples were analyzed in 20% Urea-PAGE. For DNA cleavage by Cas nucleases, dsDNA, proteins, and guide RNAs were prepared as described before (*39*–*41*). Cleavage reactions were initiated by mixing 1.5 μM Cas effector (SpyCas9 or AbCas12l1 effector) with 0.1 μM dsDNA at 37°C in a buffer containing 150 mM NaCl, 10 mM HEPES (pH 7.5), 1 mM TCEP and 10 mM $MgCl_2$. For the PlmCas12e effector, the reaction buffer contains 400 mM KCl, 20 mM Tris 7.8, 1 mM DTT and 5 mM $MgCl_2$. The reaction samples were analyzed in 15% Urea-PAGE. The gels were visualized and analyzed as described above. All experiments were performed with at least three replicates. Oligonucleotides used in this study and related descriptions are listed in table S3.

In vitro plasmid and dsDNA cleavage by HYERs

For the dsDNA cleavage assay, the nonlabeled complementary ssDNA was ordered from Tsingke Biotechnology and annealed with Cy5-labeled ssDNA containing one target site to form bubbled or fully annealed dsDNA. 0.1 μM dsDNA with one target site were incubated with 1.5 μM HYERs at 37°C in a buffer containing 50 mM $MgCl_2$, 10 mM KCl, 40 mM MOPS (pH 7.5) and 100 ng/μL BSA. After the addition of EDTA, RNase digestion and proteinase K treatment described above, the reactions were analyzed using 15% Urea-PAGE. Oligonucleotides used in this study and related descriptions are listed in table S3.

For plasmid cleavage assay, 0.03 μM *ccdB* plasmid carrying the targets was incubated with 1.5 μM HYERs at 37°C in a buffer containing 10 mM $MgCl_2$, 150 mM KCl, 40 mM MOPS (pH 7.5), and 100 ng/μL BSA. The reactions were terminated by the addition of EDTA, RNase digestion and proteinase K treatment described above. Glycerol was added to a final concentration of 15% before performing electrophoresis in 0.9% agarose gels. Open-circular and linearized plasmid were prepared using Nt.BspQI and EcoRI (NEB) following the manufacturer's protocols.

Plasmid interference in bacteria

E. coli BW25141 cells were requested from Prof. G. Shang in College of Life Sciences, Nanjing Normal University. The *ccdB* plasmid-containing BW25141 competent cells were prepared following the protocol previously described (*44*). For each group, 200 ng *HYER* plasmid was electroporated into 50 μL *ccdB* plasmid-containing competent cells with 0.2 cm cuvette (BIO-RAD) under 2.0 kV using Eppendorf eporator. After 1.5 hours of recovering in 5 mL SOC medium (Sangon Biotech) at 37°C, bacteria enriched by centrifugation were resuspended with 5 mL liquid LB medium plus 50 μg/mL streptomycin and 0.2% glucose (w/v). After every 12 hours of culturing, 50 μL culture was further seeded into 5 mL fresh liquid medium for the next 12 hours of culturing. After continuous culturing of 72 hours in total, 5 μL of bacterial culture in exponential phase (OD_{600} = 0.5-0.6) was spotted onto LB agar plates with gradient dilutions from 10^0 to 10^{-7} and incubated overnight at 37°C. 1 μL edited bacterial culture was used for PCR validation of the plasmid interference with Phanta Max Super-Fidelity DNA polymerase Mastermix (Vazyme).

Meanwhile, 1 mL culture was used for bacterial RNA extraction (*77*), which was further reverse transcribed into cDNA using TIANScript II RT Kit (Tiangen Biotech). The cDNA served as templates for real-time fluorescent quantitative PCR (RT-qPCR) to detect the mRNA level of *ccdB* gene relative to the 16S rRNA. RT-qPCR reactions using the Talent qPCR PreMix (SYBR Green) reagent (Tiangen Biotech) on a CFX96 real-time system (BIO-RAD) were performed with the following settings: preheating (95°C for 3 min); 40 cycles of alternating within denaturation (95°C for 5 s), annealing (55°C for 10 s) and extension (72°C for 15 s). The relative mRNA level of *ccdB* gene was calculated by the comparative threshold cycle (CT) method normalized to the 16S rRNA in the same sample. The relative gene expression was defined as $2^{-\Delta\Delta CT}$ (*78*). The significance of relative gene expression analysis was determined through one-way analysis of variance (ANOVA). Primers used in the experiments above are shown in table S3.

For the verification of plasmid interference capability of HYER1 with different length of TRS, the TRS-coding region on the *HYER1* plasmid was changed to 0 to 16 nt through site-directed mutagenesis by PCR. For the single-mismatch plasmid interference assay of HYER1 with 10-nt TRS in bacteria, the single-nucleotide mutations were introduced to TRS-coding region on the *HYER1* plasmid through site-directed mutagenesis by PCR. The plasmid with mutated *HYER1* was electroporated into BW25141 competent cells containing the *ccdB* plasmid for activity analysis as described above.

Cell line and culture

We constructed the reporter cell line containing 3n+1 frame-shifted puromycin resistance gene (*puro**) with a target cassette inserted upstream of the *puro**. The target cassette consists of three target sites with 30-bp intervals. Each target site is composed of native HYER1's or HYER2's TRS-complementary sequences on both the top and bottom DNA strands. The target cassette and *puro** were cloned into a PiggyBac integration plasmid under a CMV promoter by Gibson assembly. 1×10^6 HEK293T cells were electroporated with a mixture of PiggyBac integration plasmid and transposase plasmid at a 1:1 mass ratio (2 μg each) in

Downloaded from https://www.science.org at Donghua University on February 01, 2024

20 μl Opti-MEM (Gibco). After 24 hours post electroporation, cells were selected by corresponding selection marker using 10 μg/ml blasticidin (MedChemExpress) for 10 days. The HEK293T-*puro** cells were cultured in DMEM (Hyclone) supplemented with 10% (v/v) FBS (Biological Industries) and 1% (v/v) penicillin-streptomycin (Hyclone) at 37°C in 5% CO_2. The PiggyBac system often results in multiple inserted copies of the target cassettes across the genome in a single cell (*79*). For each cell clone, at least one in-frame cassette generated by HYER editing is required to restore the puromycin resistance. For editing validation, all edited and unedited target cassettes in each survival clone were PCR amplified for analysis.

Transfection of HYERs and cell editing evaluation

Native HYERs and nontargeting HYERs with mutated TRS were cloned into the PiggyBac integration plasmid under a U6 promoter, and the selection marker was replaced by *hygBR* (hygromycin B resistance gene). 1×10^6 HEK293T-*puro** cells were electroporated with a mixture of HYER plasmid and PiggyBac transposase plasmid at a 1:1 mass ratio (2 μg each) in 20 μl Opti-MEM (Gibco). After 24 hours postelectroporation, cells were selected with 50 μg/ml hygromycin B (MedChemExpress) for 2 weeks, which was followed by the puromycin selection using 5 μg/ml puromycin (MedChemExpress) for 2 weeks. 10 days into the puromycin selection, survival cells were photographed. Before and after puromycin selection, we collected the cells and extracted the genome using the Universal Genomic DNA Kit (CWBIO).

For NGS, the 254-bp regions containing the target cassette were amplified through PCR with 2× Phanta Max Master Mix (Vazyme) and primers containing adaptor sequences. Amplicons were purified by VAHTS DNA Clean Beads (Vazyme) and sequenced using the MiSeq platform. Sequencing reads containing the target sites were analyzed for editing events by CRISPResso2 with the following parameters: quantification window size of 20 bp for each target site individually and 60 bp for the whole target cassette, and plot window size of 20 bp for each target site individually and 60 bp for the whole target cassette. INDEL percentage of each target site (not including substitutions) was presented by the ratio of modified reads from the CRISPResso2 output.

For T7E1 assay (*47*), the 717-bp regions containing the target cassette were amplified (the forward primer on the CMV promoter and the reverse primer on the *puro**) through PCR with 2× Phanta Max Master Mix (Vazyme) and purified by VAHTS DNA Clean Beads (Vazyme). About 200 ng purified DNA was re-annealed for T7E1 assay according to the manufacturer's protocol (Vazyme). Reaction products were analyzed by electrophoresis using 2% agarose gels.

Primers used in the experiments above are shown in table S3.

Purification of HYERs using size exclusion chromatography

The RNA samples were in vitro transcribed and purified as described above. The further purification by size exclusion chromatography column (Superose 6 Increase 10/300 or Superose 6 Increase 5/150 GL, GE Healthcare) was performed in SEC buffer [10 mM $MgCl_2$, 50 mM KCl, 20 mM HEPES (pH 7.5)] at 4°C. The RNA fractions within the elution peak were collected and analyzed by electrophoresis using 8% Urea-PAGE. The SEC-purified RNAs were concentrated using 100 kDa molecular weight cut-off centrifugal filters (Merck Millipore) and stocked at −80°C.

Cryo-EM sample preparation and data collection

To prepare the cryo-EM samples, 4 μL purified intron RNA (~1.5 μM) was applied to the graphene oxide grid (Quantifoil Au 1.2/1.3, 300 mesh), which was glow-discharged (in a HARRICK PLASMA) for 10 s at middle level after 2 min evacuation. The grid was then blotted by a pair of 55 mm filter papers (Ted Pella) for 2.5 s at 8°C with 100% humidity, and flash-frozen in liquid ethane using FEI Vitrobot Marke IV. Cryo-EM data were collected on a Titan Krios electron microscope operated at 300 kV equipped with a Gatan K3 direct electron detector. Micrographs were recorded in counting mode at a nominal magnification of 81,000 times, resulting in a physical pixel size of 1.0825 Å per pixel. The defocus was set between -1.0 μm to -2.5 μm. The total exposure time of each movie stack led to a total accumulated dose of 50 electrons per $Å^2$, which fractionated into 32 frames. More parameters for data collection are shown in table S4.

Image processing and model building

The raw dose-fractionated image stacks were two times Fourier binned, aligned, dose-weighted, and summed using MotionCor2. CTF-estimation, blob particle picking, 2D reference-free classification, initial model generation, and final 3D refinement and local resolution estimation were performed in cryoSPARC using raw stacks (*80*). Several rounds of 3D reference-based classification were performed in cryoSPARC. The details of data processing were summarized in figs. S10, S16, S17, and table S4.

The RNA and DNA models were manually built in Coot based on the refined cryo-EM maps (*81*, *82*). The complete models were then refined against the EM maps by PHENIX in real space with secondary structure and geometry restraints (*83*). The final models were validated by PHENIX software package. The structural validation details for the final models are summarized in table S4.

Palindrome identification in DIII

We predicted the secondary structures of the DI to DIII region within 469 introns using RNAfold. By comparing these secondary structures, we found that DIII is always the last stem-loop structure on the 3′ end. We searched for 6-nt windows within the DI to DIII segments that met the following criteria: (1) The sequence should be palindromic; (2) It should be located in an unpaired region according to the secondary structure; (3) It should be located at the loop region at the terminal stem-loop structure in the DI to DIII. Combined with a manual final check, we identified 45 introns that contain the palindrome in DIII.

In vitro single-mismatch cleavage assays of HYER1

The 3′-NH_2 labeled ssDNA with single-nucleotide mismatches to the TRS of native HYER1 was ordered from Sangon Biotech and labeled with Cy5 as described above. 0.1 μM ssDNA was incubated with 1.5 μM HYER1 with 6-nt TRS at 37°C in a buffer containing 50 mM $MgCl_2$, 10 mM KCl, 40 mM MOPS (pH 7.5) and 100 ng/μL BSA. The reactions were terminated as described above and the reaction samples were analyzed on 20% Urea-PAGE. The gels were visualized and analyzed as described above. All experiments were performed with at least three replicates. Oligonucleotides used in this study and related descriptions are listed in table S3.

In vitro cleavage assays for TRS and RS engineered HYERs

DNA templates encoding HYERs with TRS in different lengths or sequences and RS-inserted HYER were constructed through site-directed mutagenesis by PCR as described above, and HYERs were prepared through IVT. 0.1 μM ssDNA with one target site was incubated with 1.5 μM reprogrammed HYERs at 37°C for 8 hours in a buffer containing 50 mM $MgCl_2$, 10 mM KCl, 40 mM MOPS (pH 7.5) and 100 ng/μL BSA. After the addition of EDTA, RNase digestion and proteinase K treatment described above, the reactions were analyzed by electrophoresis using 20% Urea-PAGE. The gels were visualized and analyzed as described above. All experiments were performed with at least three replicates. Oligonucleotides used in this study and related descriptions are listed in table S3.

Cleavage of bubbled dsDNA by HYER1 heterodimer

HYER1s with mutated palindromes and distinct TRS regions were constructed through site-directed mutagenesis by PCR. The 3′-NH_2 and 3′-FAM labeled ssDNA (corresponding to top and bottom strands respectively) was

Downloaded from https://www.science.org at Donghua University on February 01, 2024

ordered from Sangon Biotech, and the 3′-NH_2 labeled ssDNA was further labeled with Cy5 as described above. Top and bottom strands were annealed to form bubbled dsDNA. Cleavage reactions were initiated by mixing 1.5 μM HYER1 heterodimer with 0.1 μM dsDNA at 37°C in a buffer containing 50 mM $MgCl_2$, 10 mM KCl, 40 mM MOPS (pH 7.5) and 100 ng/μL BSA. The reactions were terminated as described above and the reactions were analyzed using 10% Urea-PAGE. The gels were visualized as described above. Fluorescence channels (Cy5, red; FAM, green; merge, yellow) were merged through ImageJ. Oligonucleotides used in this study and related descriptions are listed in table S3.

REFERENCES AND NOTES

1. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325 (2001). doi: 10.1126/science.1060786; pmid: 11358999
2. K. Kruger *et al*., Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–157 (1982). doi: 10.1016/0092-8674(82)90414-7; pmid: 6297745
3. M. Baer, S. Altman, A catalytic RNA and its gene from Salmonella typhimurium. *Science* **228**, 999–1002 (1985). doi: 10.1126/science.2408335; pmid: 2408335
4. C. L. Peebles *et al*., A self-splicing RNA excises an intron lariat. *Cell* **44**, 213–223 (1986). doi: 10.1016/0092-8674(86)90755-5; pmid: 3510741
5. T. A. Lincoln, G. F. Joyce, Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232 (2009). doi: 10.1126/science.1167856; pmid: 19131595
6. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983). doi: 10.1016/0092-8674(83)90117-4; pmid: 6197186
7. D. B. Haack *et al*., Cryo-EM Structures of a Group II Intron Reverse Splicing into DNA. *Cell* **178**, 612–623.e12 (2019). doi: 10.1016/j.cell.2019.06.035; pmid: 31348888
8. K. Chung *et al*., Structures of a mobile intron retroelement poised to attack its structured DNA target. *Science* **378**, 627–634 (2022). doi: 10.1126/science.abq2844; pmid: 36356138
9. D. M. Simon *et al*., Group II introns in eubacteria and archaea: ORF-less introns and new varieties. *RNA* **14**, 1704–1713 (2008). doi: 10.1261/rna.1056108; pmid: 18676618
10. A. M. Lambowitz, S. Zimmerly, Mobile group II introns. *Annu. Rev. Genet.* **38**, 1–35 (2004). doi: 10.1146/annurev.genet.38.072902.091600; pmid: 15568970
11. J. S. Rest, D. P. Mindell, Retroids in archaea: Phylogeny and lateral origins. *Mol. Biol. Evol.* **20**, 1134–1142 (2003). doi: 10.1093/molbev/msg135; pmid: 12777534
12. N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle, Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008). doi: 10.1126/science.1153803; pmid: 18388288
13. M. Marcia, A. M. Pyle, Visualizing group II intron catalysis through the stages of splicing. *Cell* **151**, 497–507 (2012). doi: 10.1016/j.cell.2012.09.033; pmid: 23101623
14. M. Karberg *et al*., Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat. Biotechnol.* **19**, 1162–1167 (2001). doi: 10.1038/nbt1201-1162; pmid: 11731786
15. J. Zhong, M. Karberg, A. M. Lambowitz, Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res.* **31**, 1656–1664 (2003). doi: 10.1093/nar/gkg248; pmid: 12626707
16. J. Yao, J. Zhong, A. M. Lambowitz, Gene targeting using randomly inserted group II introns (targetrons) recovered from an Escherichia coli gene disruption library. *Nucleic Acids Res.* **33**, 3351–3362 (2005). doi: 10.1093/nar/gki649; pmid: 15947133
17. P. J. Enyeart, G. Mohr, A. D. Ellington, A. M. Lambowitz, Biotechnological applications of mobile group II introns and their reverse transcriptases: Gene targeting, RNA-seq, and non-coding RNA analysis. *Mob. DNA* **5**, 2 (2014). doi: 10.1186/1759-8753-5-2; pmid: 24410776
18. L. Dai, S. Zimmerly, ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *RNA* **9**, 14–19 (2003). doi: 10.1261/rna.2126203; pmid: 12554871
19. M. A. Candales *et al*., Database for bacterial group II introns. *Nucleic Acids Res.* **40** (D1), D187–D190 (2012). doi: 10.1093/nar/gkr1043; pmid: 22080509
20. L. Dai, N. Toor, R. Olson, A. Keeping, S. Zimmerly, Database for mobile group II introns. *Nucleic Acids Res.* **31**, 424–426 (2003). doi: 10.1093/nar/gkg049; pmid: 12520040
21. M. Granlund, F. Michel, M. Norgren, Mutually exclusive distribution of IS1548 and GBSi1, an active group II intron identified in human isolates of group B streptococci. *J. Bacteriol.* **183**, 2560–2569 (2001). doi: 10.1128/JB.183.8.2560-2569.2001; pmid: 11274116
22. N. Toor, A. R. Robart, J. Christianson, S. Zimmerly, Self-splicing of a group IIC intron: 5′ exon recognition and alternative 5′ splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.* **34**, 6461–6471 (2006). doi: 10.1093/nar/gkl820; pmid: 17130159
23. E. A. Griffin Jr., Z. Qin, W. J. Michels Jr., A. M. Pyle, Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2′-hydroxyl groups. *Chem. Biol.* **2**, 761–770 (1995). doi: 10.1016/1074-5521(95)90104-3; pmid: 9383483
24. M. Skilandat, R. K. Sigel, The role of Mg(II) in DNA cleavage site recognition in group II intron ribozymes: Solution structure and metal ion binding sites of the RNA-DNA complex. *J. Biol. Chem.* **289**, 20650–20663 (2014). doi: 10.1074/jbc.M113.542381; pmid: 24895129
25. A. M. Pyle, The tertiary structure of group II introns: Implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.* **45**, 215–232 (2010). doi: 10.3109/10409231003796523; pmid: 20446804
26. S. Nayfach *et al*., A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* **39**, 499–509 (2021). doi: 10.1038/s41587-020-0718-6; pmid: 33169036
27. E. P. Nawrocki, S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013). doi: 10.1093/bioinformatics/btt509; pmid: 24008419
28. G. Mohr *et al*., A Highly Proliferative Group IIC Intron from Geobacillus stearothermophilus Reveals New Features of Group II Intron Mobility and Splicing. *J. Mol. Biol.* **430**, 2760–2783 (2018). doi: 10.1016/j.jmb.2018.06.019; pmid: 29913158
29. A. R. Robart, W. Seo, S. Zimmerly, Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6620–6625 (2007). doi: 10.1073/pnas.0700561104; pmid: 17420455
30. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29-37 (2011). doi: 10.1093/nar/gkr367; pmid: 21593126
31. L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426 [stat.ML] (2018).
32. S. Mohr, M. Matsuura, P. S. Perlman, A. M. Lambowitz, A DEAD-box protein alone promotes group II intron splicing and reverse splicing by acting as an RNA chaperone. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3569–3574 (2006). doi: 10.1073/pnas.0600332103; pmid: 16505350
33. P. Rangan, S. A. Woodson, Structural requirement for Mg2+ binding in the group I intron core. *J. Mol. Biol.* **329**, 229–238 (2003). doi: 10.1016/S0022-2836(03)00430-3; pmid: 12758072
34. A. Ren *et al*., In-line alignment and Mg^{2+} coordination at the cleavage site of the env22 twister ribozyme. *Nat. Commun.* **5**, 5534 (2014). doi: 10.1038/ncomms6534; pmid: 25410397
35. M. N. Ucisik, P. C. Bevilacqua, S. Hammes-Schiffer, Molecular Dynamics Study of Twister Ribozyme: Role of Mg(2+) Ions and the Hydrogen-Bonding Network in the Active Site. *Biochemistry* **55**, 3834–3846 (2016). doi: 10.1021/acs.biochem.6b00203; pmid: 27295275
36. D. M. Truong, D. J. Sidote, R. Russell, A. M. Lambowitz, Enhanced group II intron retrohoming in magnesium-deficient Escherichia coli via selection of mutations in the ribozyme core. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3800–E3809 (2013). doi: 10.1073/pnas.1315742110; pmid: 24043808
37. A. M. Lambowitz, S. Zimmerly, Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* **3**, a003616 (2011). doi: 10.1101/cshperspect.a003616; pmid: 20463000
38. M. Marcia, A. M. Pyle, Principles of ion recognition in RNA: Insights from the group II intron structures. *RNA* **20**, 516–527 (2014). doi: 10.1261/rna.043414.113; pmid: 24570483
39. A. Sun *et al*., The compact Casπ (Cas12l) 'bracelet' provides a unique structural platform for DNA manipulation. *Cell Res.* **33**, 229–244 (2023). doi: 10.1038/s41422-022-00771-2; pmid: 36650285
40. C. A. Tsuchida *et al*., Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity. *Mol. Cell* **82**, 1199–1209.e6 (2022). doi: 10.1016/j.molcel.2022.02.002; pmid: 35219382
41. M. Jinek *et al*., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012). doi: 10.1126/science.1225829; pmid: 22745249
42. J. H. Jeon, J. Adamcik, G. Dietler, R. Metzler, Supercoiling induces denaturation bubbles in circular DNA. *Phys. Rev. Lett.* **105**, 208101 (2010). doi: 10.1103/PhysRevLett.105.208101; pmid: 21231267
43. E. M. Bahassi *et al*., Interactions of CcdB with DNA gyrase. Inactivation of Gyra, poisoning of the gyrase-DNA complex, and the antidote action of CcdA. *J. Biol. Chem.* **274**, 10936–10944 (1999). doi: 10.1074/jbc.274.16.10936; pmid: 10196173
44. Z. Chen, H. Zhao, A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.* **33**, e154 (2005). doi: 10.1093/nar/gni148; pmid: 16214805
45. C. N. Takacs, Y. Nakajima, J. E. Haber, C. Jacobs-Wagner, Cas9-mediated endogenous plasmid loss in Borrelia burgdorferi. *PLOS ONE* **17**, e0278151 (2022). doi: 10.1371/journal.pone.0278151; pmid: 36441794
46. W. Y. Wu *et al*., The miniature CRISPR-Cas12m effector binds DNA to block transcription. *Mol. Cell* **82**, 4487–4502.e7 (2022). doi: 10.1016/j.molcel.2022.11.003; pmid: 36427491
47. R. D. Mashal, J. Koontz, J. Sklar, Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* **9**, 177–183 (1995). doi: 10.1038/ng0295-177; pmid: 7719346
48. A. M. Pyle, I. I. Group, Intron Self-Splicing. *Annu. Rev. Biophys.* **45**, 183–205 (2016). doi: 10.1146/annurev-biophys-062215-011149; pmid: 27391926
49. C. Zhao, A. M. Pyle, Structural Insights into the Mechanism of Group II Intron Splicing. *Trends Biochem. Sci.* **42**, 470–482 (2017). doi: 10.1016/j.tibs.2017.03.007; pmid: 28438387
50. S. M. Fica *et al*., RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013). doi: 10.1038/nature12734; pmid: 24196718
51. M. Landthaler, N. C. Lau, D. A. Shub, Group I intron homing in Bacillus phages SPO1 and SP82: A gene conversion event initiated by a nicking homing endonuclease. *J. Bacteriol.* **186**, 4307–4314 (2004). doi: 10.1128/JB.186.13.4307-4314.2004; pmid: 15205433
52. G. Santoyo, D. Romero, Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol. Rev.* **29**, 169–183 (2005). doi: 10.1016/j.femsre.2004.10.004; pmid: 15808740
53. R. Holliday, A mechanism for gene conversion in fungi. *Genet. Res.* **89**, 285–307 (2007). doi: 10.1017/S0016672308009476; pmid: 18976517
54. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018). doi: 10.1126/science.aat5011; pmid: 30166482
55. L. Cong *et al*., Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013). doi: 10.1126/science.1231143; pmid: 23287718
56. P. Mali *et al*., RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013). doi: 10.1126/science.1232033; pmid: 23287722
57. B. Zetsche *et al*., Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017). doi: 10.1038/nbt.3737; pmid: 27918548
58. D. Y. Kim *et al*., Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* **40**, 94–102 (2022). doi: 10.1038/s41587-021-01009-z; pmid: 34475560
59. S. Lin, B. T. Staahl, R. K. Alla, J. A. Doudna, Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014). doi: 10.7554/eLife.04766; pmid: 25497837
60. G. Qu *et al*., Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.* **23**, 549–557 (2016). doi: 10.1038/nsmb.3220; pmid: 27136327
61. M. P. Robertson, G. F. Joyce, The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2012). doi: 10.1101/cshperspect.a003608; pmid: 20739415

Downloaded from https://www.science.org at Donghua University on February 01, 2024

62. P. G. Higgs, N. Lehman, The RNA World: Molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**, 7–17 (2015). doi: 10.1038/nrg3841; pmid: 25385129
63. T. R. Cech, The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* **4**, a006742 (2012). doi: 10.1101/cshperspect.a006742; pmid: 21441585
64. P. Deng *et al*., Structural RNA components supervise the sequential DNA cleavage in R2 retrotransposon. *Cell* **186**, 2865–2879.e20 (2023). doi: 10.1016/j.cell.2023.05.032; pmid: 37301196
65. S. Zimmerly, C. Semper, Evolution of group II introns. *Mob. DNA* **6**, 7 (2015). doi: 10.1186/s13100-015-0037-5; pmid: 25960782
66. P. Forterre, The origin of DNA genomes and DNA replication proteins. *Curr. Opin. Microbiol.* **5**, 525–532 (2002). doi: 10.1016/S1369-5274(02)00360-0; pmid: 12354562
67. F. Sievers *et al*., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011). doi: 10.1038/msb.2011.75; pmid: 21988835
68. D. Hyatt *et al*., Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010). doi: 10.1186/1471-2105-11-119; pmid: 20211023
69. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990). doi: 10.1016/S0022-2836(05)80360-2; pmid: 2231712
70. R. Lorenz *et al*., ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011). doi: 10.1186/1748-7188-6-26; pmid: 22115189
71. B. 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). doi: 10.1093/molbev/msaa015; pmid: 32011700
72. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007). doi: 10.1093/bioinformatics/btl529; pmid: 17050570
73. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011). doi: 10.1093/bioinformatics/btr011; pmid: 21217122
74. Y. Hao *et al*., Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021). doi: 10.1016/j.cell.2021.04.048; pmid: 34062119
75. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004). doi: 10.1101/gr.849004; pmid: 15173120
76. T. D. Schneider, R. M. Stephens, Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990). doi: 10.1093/nar/18.20.6097; pmid: 2172928
77. R. Rivas *et al*., An effective, rapid and simple method for total RNA extraction from bacteria and yeast. *J. Microbiol. Methods* **47**, 59–63 (2001). doi: 10.1016/S0167-7012(01)00292-5; pmid: 11566228
78. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001). doi: 10.1006/meth.2001.1262; pmid: 11846609
79. S. Ding *et al*., Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005). doi: 10.1016/j.cell.2005.07.013; pmid: 16096065
80. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: 10.1038/nmeth.4169; pmid: 28165473
81. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
82. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
83. D. Liebschner *et al*., Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D.* **75**, 861–877 (2019). doi: 10.1107/S2059798319011471; pmid: 31588918

ACKNOWLEDGMENTS

The brainstorm about HYER was inspired by A.M. Pyle's iBiology talks: RNA Structure & Inside an RNA Splicing Machine. EM data were collected at the Tsinghua Cryo-EM facility. The data were analyzed using the Bio-Computation platform at the Tsinghua University Branch of the Chinese National Center for Protein Sciences (Beijing). We thank J.L. Lei, X.M. Li, and F. Yang for expert electron microscopy assistance. We thank T. Yang, Y.K. Wang, and A.B. Jia for their computational support. We thank D. Chia for his kind revision of this manuscript. **Funding:** This work was funded by the following grants: National Natural Science Foundation of China 32150018 (to J.J.G.L.); Ministry of Agriculture and Rural Affairs of China (to J.J.G.L. and C.G.); Start-up funds from Tsinghua University, Beijing (to J.J.G.L.). **Author contributions:** Project design: J.J.G.L., C.G., Z.X.L., S.Z., H.Z.Z., Z.H.C., Y.Y., and L.Q.L. Bioinformatics: S.Z. and Y.L. Biochemical analysis: Z.X.L., H.Z.Z., D.Y.L., A.S., G.L.W., J.Y.S., and S.Q.T. Structure analysis: H.Z.Z. Mammalian cell editing: Y.Y., L.Q.L., Y.L., and S.J. Plasmid interference: Z.H.C. and L.Q.L. Writing: J.J.G.L., Z.X.L., S.Z., H.Z.Z., Z.H.C., Y.Y., L.Q.L., and C.P.L. Supervision: J.J.G.L. and C.G. **Competing interests:** Tsinghua University has filed a patent that includes the work described in this manuscript. **Data and materials availability:** Plasmids generated in this study will be deposited to Addgene or are available upon request. The electron density maps have been deposited to the Electron Microscopy Data Bank (EMDB) under the accession numbers EMD-36776, EMD-36777, EMD-36778, EMD-36786 and EMD-36779. The atomic coordinates and structure factors have been deposited to the Protein Data Bank (PDB) under the accession numbers 8K0P, 8K0Q, 8K0R, 8K15 and 8K0S. **License information:** Copyright © 2024 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh4859
Figs. S1 to S24
Tables S1 to S4
MDAR Reproducibility Checklist
Movies S1 and S2
Data S1 to S4

Submitted 8 March 2023; resubmitted 23 August 2023
Accepted 27 December 2023
10.1126/science.adh4859

Downloaded from https://www.science.org at Donghua University on February 01, 2024