Title
DGR mutagenic transposition occurs via hypermutagenic reverse transcription primed by nicked template RNA.

Permalink
https://escholarship.org/uc/item/9mz842k4

Journal
Proceedings of the National Academy of Sciences of the United States of America, 114(47)

ISSN
1091-6490

Authors
Naorem, Santa S
Han, Jin
Wang, Shufang
et al.

Publication Date
2017-11-06

DOI
10.1073/pnas.1715952114

Peer reviewed
DGR mutagenic transposition occurs via hypermutagenic reverse transcription primed by nicked template RNA

Santa S. Naorem a, Jin Han b, Shufang Wang c, William R. Lee b, Xiao Heng b, Jeff F. Miller c,d,1, and Huatao Guo a,1

aDepartment of Molecular Microbiology and Immunology, University of Missouri School of Medicine, Columbia, MO 65212; bDepartment of Biochemistry, University of Missouri, Columbia, MO 65212; cDepartment of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095; dThe California NanoSystems Institute, University of California, Los Angeles, CA 90095

Contributed by Jeff F. Miller, October 17, 2017 (sent for review September 11, 2017; reviewed by Haig H. Kazazian Jr. and Ry Young)

Diversity-generating retroelements (DGRs) are molecular evolution machines that facilitate microbial adaptation to environmental changes. Hypervariation occurs via a mutagenic retrotransposition process from a template repeat (TR) to a variable repeat (VR) that results in adenine-to-random nucleotide conversions. Here we show that reverse transcription of the Bordetella phage DGR is primed by an adenine residue in TR RNA and is dependent on the DGR-encoded reverse transcriptase (bRT) and accessory variability determinant (Aved), but is VR-independent. We also find that the catalytic center of bRT plays an essential role in site-specific cleavage of TR RNA for cDNA priming. Adenine-specific mutagenesis occurs during reverse transcription and does not involve dUTP incorporation, indicating it results from bRT-catalyzed misincorporation of standard deoxyribonucleotides. In vivo assays show that this hybrid RNA-cDNA molecule is required for mutagenic transposition, revealing a unique mechanism of DNA hypervariation for microbial adaptation.

Diversity-generating retroelements (DGRs) mediate accelerated evolution in thousands of microorganisms and are evolutionarily related to diverse retroelements that include group II introns, retrons, reverse transcriptase (RT)-encoding CRISPRs, non-long terminal repeat (non-LTR) retrotransposons, and retroversines (1–4). The prototypical DGR was discovered in Bordetella bacteriophage BPP-1, where it facilitates phage tropism switching by diversifying a gene (mtd) encoding its receptor recognition function (5). RNA sequence hypervariation occurs through mutagenic homing, a unidirectional transposition process from a template repeat (TR) to a variable repeat (VR) with concomitant VR mutations corresponding to TR adenine residues (1, 5) (Fig. L4 and Fig. S1). Homing requires a TR RNA intermediate and two DGR-encoded proteins, Aved (accessory variability determinant) and bRT (Bordetella reverse transcriptase) (1, 5–7). It has been hypothesized that homing occurs through a target-primed reverse transcription mechanism, similar to the mobility mechanisms of related retroelements (8–11); however, this model deserves further scrutiny. The directionality of BPP-1 DGR homing is controlled by its target (VR) recognition requirements, which include a (GC) n element, a 21-bp initiation of mutagenic homing (IMH) sequence, and a DNA stem-loop structure at the 3′ end (1, 2, 6, 12) (Fig. 1A and B). Target recognition at the 5′ end of VR is determined by homology with TR, but is otherwise sequence-independent (6). Guided by these rules, the BPP-1 DGR was engineered to target a heterologous gene encoding kanamycin resistance (12).

Adenine-specific mutagenesis (A-mutagenesis) is a hallmark of DGR activity. As RT is the only protein-encoding gene found in all DGRs, DGR RTs have been predicted to be responsible for A-mutagenesis (2). Consistent with this, adenines are found to be unaltered in the TR RNA intermediate, and A-mutagenesis has been proposed to result from random nucleotide incorporation when DGR RTs reverse transcribe TR adenine residues (2, 6).

However, an alternative hypothesis, which involves adenine-templated dUTP incorporation by the DGR RT, was also proposed (2, 6). Accordingly, uracil bases would be subsequently removed by uracil DNA glycosylase (UDG) and random nucleotides inserted when DNA polymerases copy abasic sites.

Here, we determined the mechanism of DGR mutagenic homing. By tagging TR with a self-splicing group I intron, we found that cDNA synthesis occurs in the absence of VR target sequences. 5′ RACE (rapid amplification of cDNA ends) and mutagenesis studies showed that reverse transcription of TR RNA is primed by an adenine residue at position 56 downstream of TR. Remarkably, we found that the catalytic center of bRT is essential for site-specific cleavage of TR RNA at A56 for cDNA initiation. A-mutagenesis occurs during minusstrand cDNA synthesis and does not require adenine-templated polymerization of dUT residues, indicating that A-mutagenesis is likely caused by bRT-mediated random nucleotide addition. Most importantly, we observed a strong correlation between the effects of different mutations on cDNA synthesis and their effects on mutagenic homing, suggesting the BPP-1 DGR functions through a target-independent reverse transcription mechanism.

Significance

Diversity-generating retroelements (DGRs) are in vivo sequence diversification machines that are widely distributed in bacteria, archaea, and their viruses. DGRs use a reverse transcriptase (RT)-mediated mechanism to diversify protein-encoding genes to facilitate adaptation of their hosts to changing environments. Here, we demonstrate that the Bordetella phage DGR-encoded RT uses the 3′-OH of a nicked template RNA to initiate reverse transcription, during which random nucleotides are incorporated when adenine residues in the template are copied into complementary DNA (cDNA). We further show that this mutated, covalently linked RNA-cDNA molecule is required for DGR-mediated sequence diversification, revealing a mechanism of accelerated evolution with broad practical applications.

Author contributions: S.S.N., X.H., J.F.M., and H.G. designed research; S.S.N., J.H., S.W., W.R.L., and H.G. performed research; S.S.N., X.H., J.F.M., and H.G. analyzed data; and S.S.N., J.F.M., and H.G. wrote the paper.

Reviewers: H.H.K., McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine; and R.Y., Texas A&M University.

Conflict of interest statement: J.F.M. is a founder of AvidBiotics Corporation and a member of its scientific advisory board. H.G. is a former consultant for AvidBiotics Corporation. The remaining authors declare no conflict of interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence may be addressed. Email: jfmiller@ucla.edu or guohua@missouri.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1715952114/-/DCSupplemental.
As an initial test of the target-cell by PCR. The 148-bp band was not detected in mutants that are RT-VR DNA in the plasmids (pMX- of pMX- end of RACE assay was next performed to and RNA intermediate had been induction. A 5 RNA (blue). If cDNA synthesis occurs, the products (pink) can be detected by PCR with primers P1 and E2a. ?, unknown cDNA priming deletion mutant, or mutants that have a ′ RNA. TR after P is a stem–loop RNA. TR promoter (P TR intron-containing, from the promoter (P TR avd, bRT, and RB50, which contains the YMDD to SMAA mutation in the bRT catalytic center, nor in the avd deletion mutant, or mutants that have a defective (ΔP7.1–2a) or inverted (td-) intron (Fig. 1 D and E). Together, these results showed that VR-independent reverse transcription had occurred in vivo and required both bRT and Avd, and that spliced products of the TR RNA intermediate had been used as templates. A-mutagenesis was detected in the smaller products derived from pMX-td (frequency = ~39.8%), demonstrating that A-mutagenesis occurs in a step before cDNA integration into VR, most likely as a result of error-prone nucleotide incorporation by bRT (Fig. S2).

cDNA Priming by TR RNA. A 5′ RACE assay was next performed to identify the primer used for VR-independent cDNA synthesis. Total nucleic acids were isolated from RB50 cells transformed with pMX-td after Pfha induction. A 5′-biotin-labeled sense-strand td exon oligo (EJBio) was annealed to (−cDNA for primer extension in the presence of SuperScript III RT (Fig. 2 A). The resulting biotinylated products were purified using streptavidin magnetic beads, and subsequently poly(dA) tails were

**Results**

**VR-Independent cDNA Synthesis.** As an initial test of the target-primed reverse transcription control for DGR homing, we determined whether cDNA synthesis can occur in the absence of VR target sequences. *Bordetella bronchiseptica* RB50, which contains no BPP-1 sequences, was transformed with pMX-td, which expresses DGR components Avd, bRT, and TR RNA tagged with a self-splicing td intron, under transcriptional control by an inducible fhaB promoter (Pfha) (6) (Fig. 1C). After transcription and intron splicing, the mature TR RNA retains a 36-nt ligated exon sequence (E1E2). Plasmid pMX-E1E2 is a positive control that contains the same ligated exons. Induced cells were used as templates for PCR analysis, using primers P1 and E2a. P1 is a sense-strand primer annealing to the 5′ end of TR, and E2a is an antisense-strand primer annealing to td exon 2 (Fig. 1C). In addition to a larger band (541 bp) amplified from the intron-containing TR of pMX-td, we observed a smaller band (148 bp) that has the same size as products generated from pMX-E1E2 (Fig. 1D). The smaller band was not detected in the RT-deficient mutant (pMX-td/SMAA), which contains the YMDD to SMAA mutation in the bRT catalytic center, nor in the avd deletion mutant, or mutants that have a defective (ΔP7.1–2a) or inverted (td-) intron (Fig. 1 D and E). Together, these results showed that VR-independent reverse transcription had occurred in vivo and required both bRT and Avd, and that spliced products of the TR RNA intermediate had been used as templates. A-mutagenesis was detected in the smaller products derived from pMX-td (frequency = ~39.8%), demonstrating that A-mutagenesis occurs in a step before cDNA integration into VR, most likely as a result of error-prone nucleotide incorporation by bRT (Fig. S2).
added to the 3’ ends of these cDNA products using terminal deoxynucleotidyl transferase (TdT). The products were amplified using nested PCR, first with primers EJ+9 (td exon junction primer with 9 nt from E2) and dT18 [annealing to the poly(dA) tail], and then with primers EJ+16 (containing 16 nt from E2) and dT18. We detected a ~170-bp band for pMX-td and a significantly weaker band for the SMAA mutant, which was also of slower mobility (Fig. 2B). Both products were cloned and sequenced. In all clones derived from pMX-td, we found that the TR-derived sequence was selectively mutated at adenine residues and fused to a sequence from the spacer (sp) region between TR and brt, but in the reverse orientation, suggesting this sequence had folded back onto TR to initiate cDNA synthesis (Fig. 2C and D). Because of complementarities between TRU118 (uracil at TR position 118) and spA56 (adenine at sp position 56), and between TRG119 and spC55, it was not clear whether spU54, spC55, or spA56 was the priming nucleotide. The four SMAA-derived clones analyzed contain TR (1/4) or both TR and sp (3/4) sequences, and none have A-mutagenesis in the TR region, suggesting they were not derived from bRT synthesized cDNA (Fig. S3).

To determine the exact cDNA initiation site, we introduced targeted mutations into plasmid pMX-td and determined their effects on cDNA synthesis. Two substitutions (spC55U and spA56G) were introduced into the primer region, and the other two (TRU118C and TRG119A) change complementary nucleotides on the template side (residues were named according to the TR RNA sequence; U = T; Fig. 3A). Nucleic acids were isolated from transformed RB50 cells. Covalently linked RNA-cDNA molecules were reverse transcribed with primer EJ+2 (containing 2 nt from E2), and products were amplified by PCR, using primers EJ+10 (containing 10 nt from E2) and TRsp, which anneals to the TR-sp junction (Fig. 3B). All mutations affected TR-independent cDNA synthesis. Three mutants, TRU118C, TRG119A, and spC55U, remained functional and generated A-mutagenized cDNA products, whereas spA56G had no detectable cDNA synthesis (Fig. 3C and D). In agreement, mutant spM4, which has a 5’-5’CAGC to 5’-5’GUCG mutation, had no detectable cDNA synthesis (Fig. 3A and C). Sequence analysis also yielded information on cDNA initiation (Fig. 3E). Mutations TRU118C and TRG119A were not copied into cDNA products, whereas spC55U was detected at the primer-cDNA junction, suggesting spA56 was the priming residue and TRG117 was the first nucleotide being reverse transcribed. Complementary mutations TRU118C-spA56G and TRG119A-spC55U partially rescued defects in cDNA synthesis resulting from corresponding single-base mutations (Fig. 3D). Sequence analysis detected mutations spA56G and spC55U at the primer-cDNA junction (Fig. 3F). As mutations TRU118C and TRG119A alone were not copied into cDNA products, this observation again indicated that sp56 is the priming site and that a guanine substitution can be tolerated if paired with a
cytosine at TR118. Interestingly, sequence analysis of cDNA products also showed nontemplated nucleotide addition in mutants TRU118C, TRG119A, and spC55U, but not in the complementary mutants (Fig. 3E and F). Bases were inserted between position spA56 and the nucleotide copied from TRG117 and ranged from 2 to 7 nt, with no obvious pattern observed. It appears that primer-template mispairing induces nontemplated polymerization during cDNA initiation.

**TR RNA Is Cleaved at the Priming Site.** To determine whether TR RNA is cleaved at spA56 during homing, we performed an RNase protection assay with a 32P-labeled antisense RNA probe that is complementary to TR RNA from position sp5 to position 6 of brt (Fig. 4A). We compared RNase protection patterns of RNA samples isolated from RB50 harboring plasmid pMX-Km1 or its RT-deficient mutant SMAA (12). An additional mutant, pMX-Km1/spM4, was used as a positive control, as it contains the 5′-GUCG mutation in the priming region. This results in mispairing between TR RNA and the labeled probe, thus allowing TRRNA cleavage (Fig. 4A). Indeed, we detected two protected bands in spM4, which were ~52 and ~94 nt, respectively, corresponding to 5′ and 3′ cleavage products (Fig. 4A). Interestingly, we observed two similar products in the wild-type construct, but not in the SMAA mutant, predicating the hypothesis that the TR RNA template was cleaved in a bRT-dependent manner to generate a 3′-OH primer for cDNA synthesis. An alternative possibility, however, is that cDNA synthesis is initiated at the 2′-OH of spA56, resulting in a branched structure that impairs base pairing of the probe with TR RNA and allows RNase cleavage. Interestingly, deletion of either avd or bRT, which are known to interact, bind the TR RNA substrate and protect it from degradation (7) (Fig. 4A).

We reasoned that if TR RNA cleavage had occurred, it should be possible to map the cleavage site by primer extension (Fig. 4B). Using a 32P-labeled primer annealing downstream of spA56, two products that correspond to termination at positions spG57 or spA56 were detected in the wild-type construct, but not in mutants containing substitutions in the YMDD box of the bRT catalytic center (SMAA, YMDA,YMAD, or YMAM). A likely explanation is that the TR RNA intermediate was cleaved between spG57 and spA56, with the slower-migrating band resulting from nontemplated nucleotide addition by SuperScript III RT (13) (Fig. 4B). Using a primer annealing to the cDNA (Fig. 4C), multiple bRT-dependent

---

**Fig. 3.** The priming residue and effects of TR RNA mutations on cDNA synthesis. (A) Diagram to show introduced mutations in TR and sp sequences. (B) Diagram to show primers used for RT-PCR assay. Primers EJ+2 and EJ+10 are exon junction oligos with 2 and 10 nt of E2, respectively. TRsp anneals to the junction of TR and sp. (C and D) RT-PCR assays to analyze cDNA synthesis in single and complementary mutants. (C, Left) Results of 25 PCR cycles. (C, Right) Results of 30 cycles for samples 2, 3, 6, and 7. Only products from reactions 1, 3, 4, and 5 showed A-mutagenesis. (D) 27 cycles. *Expected products. (E) Sequences at the junction of the RNA primer and cDNAs of the three functional single-base mutants. The sense-strand sequence was aligned. Sequences to the left and right of the vertical lines were derived from cDNA and the primer, respectively. Open and filled arrows show absence and presence of the introduced mutations in the RT-PCR products, respectively. Nontemplated polymerization was observed between the priming residue and the nucleotide copied from TRG117 in some products. (F) Sequences at the junction of the RNA primer and cDNAs of complementary mutants TRG119A-spC55U and TRU118C-spA56G.
Fig. 4. bRT-dependent TR RNA cleavage at the cDNA priming site. (A, Left) Diagram of the RNase protection assay. TR, blue; sp, black; cDNA, pink; uniformly labeled antisense RNA probe, red. Arrowheads indicate RNase cleavage sites (Right). spM4, which has four mutated residues at the priming area, served as a positive control. The two protected fragments are indicated by green and red triangles. Km1 and SMAA are wild-type and catalytically inactive bRT constructs, respectively. (A, Right) RNase protection analysis. The 194-nt band is probe, the 146-nt band is probe cleaved by RNase at the unpaired ends, and the 94- and 52-nt bands are RNase protection products resulting from cleavage at the priming region and the two ends. Δavd and Δbrt are deletion mutants of avd and brt, respectively. (B) Primer extension assay of wild type (Km1) and bRT-deficient mutants (SMAA, YMID, YMAD and YMAA) with a $^{32}$P-labeled primer (P2) annealing downstream of spA56. Two bRT-dependent bands (spA56 and spG57) are observed, suggesting that TR RNA is either cleaved or branched at spA56. Sequencing ladders were generated from pMX-Km1 with the same labeled primer. Shown below is a diagram of TR RNA cleavage at spA56. (C) Primer extension assay of Km1 and the SMAA mutant with a $^{32}$P-labeled primer (P3) annealing to (−)cDNA. Km1-specific terminations were detected in the IMH* region (blue dot) after passing the cDNA-RNA primer junction (red dot). The junction did not block primer extension. Sequencing ladders were generated from pMX-Km1 with the same labeled primer. (D) Primer extension of splint ligation products with $^{32}$P-labeled primer P2. Ligation of adaptor DNA oligos to spG57 of cleaved spRNA occurred in the presence of a Bridge DNA oligo (3′-biotin-tagged) and T4 DNA ligase. 16-mer, 18-mer, and 20-mer are 16-nt, 18-nt, and 20-nt adaptors. Oligo 2nt-gap is 2 nt away from spG57 after annealing. −, a negative control with no adaptor oligo added. Sequencing ladders were generated as in B. Primer extension products generated from those of splint ligation are marked by green dots. The ligation product of the 20-mer adaptor could not be detected because of a comigrating background band.
terminations were observed within the IMH* region of TR RNA, matching the major region of poly(dA) addition in the 5′ RACE assay (Fig. 2C). In contrast, we did not observe bRT-dependent termination at the RNA-cDNA junction, which would be expected if there were a branched intermediate resulting from cDNA priming by the 2′-OH of spA56 (14–16).

To further test the TR RNA cleavage hypothesis, we performed a splint ligation reaction followed by primer extension to detect free 5′-phosphoryl groups on spG57, the predicted products of cleavage at spA56 (Fig. 4D). Nucleic acids isolated from Bordetella cultures were combined with adaptor DNA oligos of different lengths and a bridge oligo that base pairs with adaptors and spRNA from spG57 to spUT74. After annealing, 16-mer and 18-mer adaptor oligos, which place 3′-OH groups directly adjacent to spG57, were ligated by T4 DNA ligase, and reaction products of expected sizes were detected by primer extension (Fig. 4D). A 20-mer adaptor with the same 3′ terminus was also expected to serve as a ligation partner, but the predicted product was obscured by a background band. Oligo 2nt-gap, which anneals 2 nt away from spG57, failed to generate ligation products, as expected. Most important, the ability to detect splint ligation products required enzymatically active bRT. Together, these observations support a model in which TR RNA is cleaved at spA56 in a bRT-dependent manner, leaving a 5′ phosphate on spG57 and a 3′-OH on spA56 to initiate cDNA synthesis.

Mutagenic Synthesis of (−)cDNA. The dUTP hypothesis suggests that A-mutagenesis occurs during (+)cDNA synthesis after uracil removal from (−)cDNA by UDG. If correct, then the results in Fig. 1 showing A-mutagenized PCR products in the absence of VR sequences would suggest that mutagenesis occurred during PCR amplification of (−)cDNA templates containing abasic sites, or that error-prone (+)cDNA synthesis had occurred in vivo.

To determine whether abasic sites in template DNA are intrinsically mutagenic during PCR, we generated dU-containing DNA corresponding to the BPP-1 VR and flanking sequences using PCR in the presence of dUTP instead of dTTP (Fig. 5A). PCR products were then treated with different amounts of Escherichia coli UDG, which removes uracils from both single- and double-stranded DNA substrates. Template DNA treated with 5 U of UDG led to almost no PCR product with Taq and standard dNTPs, whereas decreased amounts of UDG resulted in increased levels of amplification, indicating that abasic sites are inhibitory to PCR (Fig. 5B). No A-mutagenesis was detected in Taq polymerization products derived from dU-DNA treated or untreated with UDG, indicating that abasic sites in the template are not mutagenic during Taq polymerization, or that such templates are refractory to PCR amplification by Taq (Figs. S4 and S5).

To evaluate the possibility that (+)cDNA syntheses can occur in vivo in the absence of VR target sequences, we isolated nucleic acids from Bordetella cultures transformed with pMX-td and performed RT-PCR using primer EJ+2 for reverse transcription and primer pair EJ+4/Trsp for PCR (Fig. 5 C and D). A product of the expected size was generated in the complete reaction, but not in the absence of SuperScript III RT, and sequence analysis showed A-mutagenized cDNA products that were covalently linked to the TR RNA primer (Fig. S6). Importantly, no RT-PCR products were observed in the absence of sense-strand primer EJ+2 in the RT reaction, showing that (+)cDNA synthesis, which could produce PCR templates in the absence of EJ+2, did not occur in vivo. Thus, our results demonstrate that A-mutagenesis occurs during (−)cDNA synthesis, thereby excluding the dUTP incorporation hypothesis and strongly suggesting that A-mutagenesis is a result of adenine-specific misincorporation by bRT.

Hybrid RNA-cDNA Is Essential for Homing. To determine whether hybrid TR RNA-cDNA products are required intermediates for homing, we introduced mutations characterized above into donor plasmid pMX-Km1, which encodes a TR derivative with the 3′ end of a KanR ORF (12) (Fig. 6A). Mutant constructs, or positive (pMX-Km1) or negative (pMX-Km1/SMAA) controls, were transformed into Bordetella lysogen RB50/BPP-1ΔTR* KanR, which contains a 3′ truncated, defective KanR gene upstream of (GC)14 IMH, and the stem-loop structure (Fig. 6A). DGR homing repairs the defective KanR gene, and homing efficiencies were determined as percentages of KanR colonies. The wild-type plasmid had a homing efficiency of ∼2.2 × 10−7, whereas the RT-deficient control had no detectable homing activity (<6.8 × 10−11; Fig. 6B). Mutants TRG119A and spC55U, which had reduced
Fig. 6. Synthesis of hybrid RNA-cDNA products is required for DGR homing. (A) Diagram of the KanR reporter assay. As the recipient (BPP-1ΔATR*KanS prophage), a defective KanR gene, which has a 3′ truncation of six amino acids, was inserted upstream of (GC)14 and IMH and the stem–loop structure. As the donor (pMX-Km1), a 36-bp KanR sequence, which encodes the last 12 amino acids, was inserted upstream of (GC)14 and IMH* in TR. DGR homing, or TR-to-VR transfer, repairs the truncated KanR gene, which can be selected based on kanamycin resistance. (B) Results of the KanR reporter assay. Km1 and SMAA were positive and negative controls, respectively. Desired mutations were introduced into plasmid pMX-Km1. Results were the average of three independent experiments, and bars represent mean ± SD. Constructs SMAA, TRU118C, spA56G, and spM4 had no detectable homing activity, and bars for these mutants represent the detection limits. *P = 0.0019; **P = 0.0038; ***P < 0.0001; ****P < 0.0001 by unpaired Student’s t test. (C) A model for DGR mutagenic homing. Mutagenic homing by the BPP-1 DGR occurs through a TR RNA intermediate, which includes TR and flanking sequences (6, 7). After transcription, TR RNA is cleaved at spA56 in a manner that appears to use bRT as the catalytic component and Avd as a cofactor. Subsequently, the 3′-OH of spA56 is used by bRT as a primer to reverse transcribe the TR RNA sequence in an error-prone fashion, resulting in a nascent (−)cDNA copy containing dense substitutions corresponding to adenines in TR. The template sequence may then be cleaved by a host activity, and cDNA integration is predicted to involve base-pairing interactions between the RNA primer and VR, as well as cDNA-VR interactions (blue box). We speculate that cleavage of the RNA primer base-paired with VR DNA by a host-encoded RNase H-like activity creates what is essentially an “Okazaki fragment.” DNA repair by DNA polymerase I would likely include nick-translation, which may remove ~5–10 bases of DNA in addition to the residual RNA primer, leading to our previous observation of a 3′ marker coconversion boundary (TR107-TR112) shortly upstream of TR117, the first nucleotide copied into cDNA (6). The nick would then be sealed by a host DNA ligase.
cDNA synthesis activities, also had reduced DGR homing (~1.1 × 10^{-9} and ~3.5 × 10^{-11}, respectively). Complementary mutations TRG119A-gpC55U partially rescued single-base mutations in DGR homing (~4.9 × 10^{-10}), similar to their effect on cDNA synthesis. No homing activity was observed for mutants TRU118C and spA56G, but the complementary mutant TRU118C-spA56G regained a low but detectable level of homing activity (~4.4 × 10^{-10}) (Fig. 6B). In a different experiment using a lower kanamycin concentration, we did detect one Kan^{R} colony in mutant TRU118C, suggesting a low level of residual activity. Finally, mutant spM4 had no detectable DGR homing, in agreement with cDNA synthesis data. For all constructs with homing activity, sequence analysis showed A-mutagenesis in repaired Kan^{R} genes. Taken together, we conclude that target-independent cDNA synthesis primed by spA56 in the TR RNA intermediate is a key step in DGR mutagenic homing.

Discussion
In this study, we have determined the mechanisms of TR-to-VR transposition and A-mutagenesis, two key activities of DGR mutagenic homing. Our results show that BPP-1 DGR homing occurs through a target-independent reverse transcription mechanism mediated by an RNP complex consisting of Avd, bRT, and TR RNA. After transcription, the TR RNA intermediate is cleaved at position spA56 by a mechanism that requires catalytically active bRT (Fig. 6C). This raises the intriguing possibility that bRT itself enzymatically cleaves TR RNA at spA56. To our knowledge, no RNase H-less RT, natural or engineered, has ever been shown to have endoribonuclease activity. After TR RNA cleavage, bRT uses the 3'-OH of the cleaved RNA as a primer for reverse transcription.

How the cDNA portion of the hybrid TR RNA-cDNA intermediate integrates into VR remains an unanswered question. Comparing the hybrid RNA-cDNA molecule with VR and TR DNA sequences, however, may be relevant in this regard. As shown in Fig. 6C (blue box), the RNA primer sequence forms more extensive base pairing interactions with VR IMH than with TR IMH. Our previous observation that IMH dictates the directionality of mutagenic homing is likely explained by cDNA integration at the 3' end of VR resulting from base pairing interactions between the cDNA-linked RNA primer and IMH DNA at the 5'-stem-loop structure downstream of VR, which increases homing efficiency, could mediate transient DNA unwinding or replication pausing to facilitate strand invasion of the hybrid RNA-cDNA molecule (12). The RNA primer, once base-paired with VR DNA, would likely be cleaved by host-encoded RNase H, leading to formation of an Okazaki-like fragment base-paired with VR DNA (17). Subsequent repair may involve nick-translation catalyzed by DNA polymerase I (or a Pol I-like enzyme), which also has a 5' → 3' exonuclease activity that removes nucleotides from DNA-DNA primers during Okazaki fragment maturation. Variability at the 5' end of the integrated cDNA is consistent with our previous observation of a marker coconversion boundary spanning several nucleotides (TR107-TR112) shortly upstream of TR117, the first residue reverse transcribed (6). Eventually, nicks would be connected by host DNA ligase.

A-mutagenesis is perhaps the most distinctive feature of DGRs and is considered to be an intrinsic property of DGR-encoded RTs (2). Our results clearly show that A-mutagenesis occurs during (−)cDNA synthesis, which precludes models that invoke dUTP incorporation and UDG-mediated base removal followed by DNA replication. In support of this, Taq does not insert random nucleotides when amplifying DNA templates containing abasic sites, and there is no detectable (−)cDNA synthesis in vivo in the absence of VR. Thus, our results provide convincing evidence that A-mutagenesis does not involve dUTP incorporation, indicating it is a consequence of DGR RT-mediated incorporation of random nucleotides. To our knowledge, DGR RTs are the most error-prone RTs ever discovered. The error rate of the BPP-1 RT, which stands at ~40% for TR adenine residues, is more than 13,000 times higher than that of the HIV type 1 (HIV-1) RT, which misincorporates nucleotides at a rate of ~1.4–3.0 × 10^{-6} per cycle per base (18, 19). The low fidelity of HIV-1 replication plays a critical role in virus immune evasion and drug resistance. DGR RTs are related to HIV-1 RT and have conserved finger and palm domains (2). It will be of interest to determine whether the unique residues and motifs conserved within the DGR RT clade play a role in determining A-specific infidelity during polymerization. We predict that the fundamental mechanism of mutagenic homing will be conserved among the thousands of DGRs identified in bacteria, archaea, and their viruses, as they are structurally similar and have related RTs and signatures of A-mutagenesis (4). A better understanding of DGR mutagenic homing will be important for understanding their physiological roles in host organisms, and for exploiting DGRs for diverse bioengineering applications.

Materials and Methods
Bacterial Strains, DNA Oligonucleotides, and Plasmid Constructs. B. bronchiseptica strains RB50 and RB50-BPP-1ΔATP*Kan^{R} enzymes have been described (12, 20). E. coli DH5α-pIR was used for routine cloning (21). DNA oligonucleotides are listed in Table S1. Plasmid constructs are described in SI Materials and Methods.

PCR Assay of VR-Independent cDNA Synthesis. Plasmids pMX-E1E2, pMX-td and derivatives of pMX-td were electroporated into Bordetella RB50 cells. Transformed cells were then cultured and used for VR-independent cDNA synthesis assay as described in SI Materials and Methods.

cDNA Primer Determination by 5′-RACE. Nucleic acids were isolated from Bordetella RB50 cells transformed with either plasmid pMX-td or pMX-tdSMAA. 5′ RACE assay was performed as described in SI Materials and Methods.

RT-PCR Assay of VR-Independent cDNA Synthesis. Plasmid pMX-td and its derivatives were electroporated into Bordetella RB50 cells. Nucleic acids were then isolated, and VR-independent cDNA synthesis assay was performed as described in SI Materials and Methods.

RNase Protection Assay. RNase protection assay was performed using the RPA III kit (Ambion) according to manufacturer’s instructions (see SI Materials and Methods for details).

Primer Extension and Splint Ligation Assays. To determine whether TR RNA is cleaved, primer extension and splint ligation assays were performed as described in SI Materials and Methods. To check whether cDNA is branched at spA56, a similar primer extension assay was performed with 32P-labeled primer P3, which anneals to (−)cDNA.

Fidelity of Taq in Copying Abasic Sites. Deoxyuracil (dU) residues were introduced into BPP-1 VR and its flanking sequences by PCR in the presence of dUTP. The dU-containing DNA was then treated with UDG to generate DNA containing abasic sites (see SI Materials and Methods for more details). Subsequently, UDG-treated DNA samples were used as templates to determine the fidelity of Taq polymerase in copying abasic sites (see SI Materials and Methods for more details).

Kan^{R} Reporter Assay. The Kan^{R}-based DGR homing assay was performed as described previously, with minor modifications (see SI Materials and Methods for more details) (12).

ACKNOWLEDGMENTS. We thank Dr. Parto Ghosh, Dr. Sumit Handa, Dr. Alan M. Lambowitz, Dr. Xiaoxia Cui, Dr. Steve Zimmerly, Dr. Jeff D. Boeke, Dr. Mark McIntosh, Dr. David Pintel, and members of the laboratories of H.G. and J.F.M. for constructive input. This work was supported by the University of Missouri Startup Fund (to H.G.), NIH Grant R01 AI068838 (to J.F.M.), and NIH Grant P50 GM103297 (to X.H.).
1. Doulatov S, et al. (2004) Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. *Nature* 431:476–481.

2. Guo H, Arambula D, Ghosh P, Miller JF (2015) Diversity-generating retroelements in phage and bacterial genomes. *Mobile DNA III*, eds Craig NL, et al. (American Society for Microbiology, Washington, DC), pp 1237–1252.

3. Silas S, et al. (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* 351:aad4234.

4. Paul BG, et al. (2017) Retroelement-guided protein diversification abounds in vast lineages of bacteria and archaea. *Nat Microbiol* 2:17045.

5. Liu M, et al. (2002) Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science* 295:2091–2094.

6. Guo H, et al. (2008) Diversity-generating retroelement homing regenerates target sequences for repeated rounds of codon rewriting and protein diversification. *Mol Cell* 31:813–823.

7. Alayyoubi M, et al. (2013) Structure of the essential diversity-generating retroelement protein bAvd and its functionally important interaction with reverse transcriptase. *Structure* 21:266–276.

8. Zimmerly S, Guo H, Perlman PS, Lambowitz AM (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 82:545–554.

9. Zimmerly S, et al. (1995) A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* 83:529–538.

10. Luan DD, Korman MH, Jakubczak JL, Eckbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell* 72:595–605.

11. Cost GI, Feng Q, Jacquier A, Boeke JD (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J* 21:5899–5910.

12. Guo H, et al. (2011) Target site recognition by a diversity-generating retroelement. *PLoS Genet* 7:e1002414.

13. Potter J, Zheng W, Lee J (2003) Thermal stability and cDNA synthesis capability of Superscript III reverse transcriptase. *Focus* 25:19–24.

14. Cousineau B, et al. (1998) Retrosnoing of a bacterial group II intron: Mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* 94:451–462.

15. Molina-Sánchez MD, Barrientos-Duran A, Toro N (2011) Relevance of the branch point adenosine, coordination loop, and 3′ exon binding site for in vivo excision of the Sinorhizobium meliloti group II intron RmInt1. *J Biol Chem* 286:21154–21163.

16. Döring J, Hurek T (2017) Arm-specific cleavage and mutation during reverse transcription of 2′,5′ branched RNA by Moloney murine leukemia virus reverse transcriptase. *Nucleic Acids Res* 45:3967–3984.

17. Okazaki T (2017) Days weaving the lagging strand synthesis of DNA: A personal recollection of the discovery of Okazaki fragments and studies on discontinuous replication mechanism. *Proc Jpn Acad Ser B Phys Biol Sci* 93:322–338.

18. Mansky LM, Temin HM (1995) Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* 69:5087–5094.

19. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH (2010) Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* 84:9864–9878.

20. Liu M, et al. (2004) Genomic and genetic analysis of Bordetella bacteriophages encoding reverse transcriptase-mediated tropism-switching cassettes. *J Bacteriol* 186:1503–1517.

21. Kolter R, Inuzuka M, Helinski DR (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* 15:1199–1208.