2015

An unexplored diversity of reverse transcriptases in bacteria

Zimmerly, Steven; Wu, Li

ASMscience

Zimmerly, S., Wu, L. (2015) An unexplored diversity of reverse transcriptases in bacteria. Microbiology Spectrum. April 2015 3:2 doi:10.1128/microbiolspec.MDNA3-0058-2014
http://hdl.handle.net/1880/51148
journal article

Microbiology Spectrum
American Society for Microbiology Press

An Unexplored Diversity of Reverse Transcriptases in Bacteria

STEVEN ZIMMERLY and LI WU
Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

ABSTRACT Reverse transcriptases (RTs) are usually thought of as eukaryotic enzymes, but they are also present in bacteria and likely originated in bacteria and migrated to eukaryotes. Only three types of bacterial retroelements have been substantially characterized: group II introns, diversity-generating retroelements, and retrons. Recent work, however, has identified a myriad of uncharacterized RTs and RT-related sequences in bacterial genomes, which exhibit great sequence diversity and a range of domain structures. Apart from group II introns, none of these putative RTs show evidence of active retromobility. Instead, available information suggests that they are involved in useful processes, sometimes related to phages or phage resistance. This article reviews our knowledge of both characterized and uncharacterized RTs in bacteria. The range of their sequences and genomic contexts promises the discovery of new biochemical reactions and biological phenomena.

INTRODUCTION

Reverse transcriptase (RT) is generally considered a eukaryotic enzyme because it is prevalent in eukaryotes and was first characterized from eukaryotic sources. Discovered in 1970 in the Rous Sarcoma and murine leukemia viruses (1, 2), RT has since been studied for its central role in the replication of many eukaryotic genetic elements including retroviruses (e.g., HIV-1), pararetroviruses, hepadnaviruses, long terminal repeat (LTR), and non-LTR retroelements, Penelope-like elements, and telomerase (3, 4, 5, 6, 7, 8, 9, 10). Over the years, the accumulated studies of RT have painted a picture in which the enzyme functions primarily as the replicative enzyme of selfish DNAs (viruses, retrotransposons), while occasionally becoming domesticated to perform useful cellular functions. These functions include the maintenance of chromosomal ends (telomerase, *Drosophila* Het-A elements) (10, 11) and contributions to genomic change (both beneficial and deleterious) through pseudogene formation or other retroprocessing events (12, 13, 14, 15).

RT was discovered in bacteria in 1989 as a component of an element named a retron (16, 17, 18). Retrons produce small nucleic acid products in the cell called multicopy single-stranded DNAs (msDNAs), whose function remains elusive to this day. In 1993, a second class of bacterial RTs was discovered, group II introns, which had previously been found in the mitochondria and chloroplasts of fungi, algae, and plants (19). Group II introns were subsequently found to have features typical of mobile DNAs, in that they amplify themselves and spread in genomes without providing obvious benefits to their hosts (20, 21, 22). About a decade later, the third type of bacterial retroelement was discovered, the diversity-generating retroelement (DGR) (23). Unlike most retroelements, DGRs are not actively mobile but carry out a directed-mutagenesis reaction that is advantageous to its host genome.

More recently, it has become evident that bacteria, in fact, harbor a plethora of overlooked RTs and RT-related sequences. Most of these putative RTs do not show signs of being mobile DNAs because they do not accumulate to multiple copies in genomes. While the functions of these elements remain unclear, they do not appear to fit into the paradigm of eukaryotic selfish

Received: 2 October 2014, **Accepted:** 6 November 2014, **Published:** 12 March 2015
Editors: Alan Lambowitz, University of Texas, Austin, TX, and Nancy Craig, Johns Hopkins University, Baltimore, MD
Citation: Zimmerly S, Wu L. 2014. An unexplored diversity of reverse transcriptases in bacteria. *Microbiol Spectrum* 3(2):MDNA3-0058-2014. doi:10.1128/microbiolspec.MDNA3-0058-2014.
Correspondence: Steve Zimmerly, zimmerly@ucalgary.ca

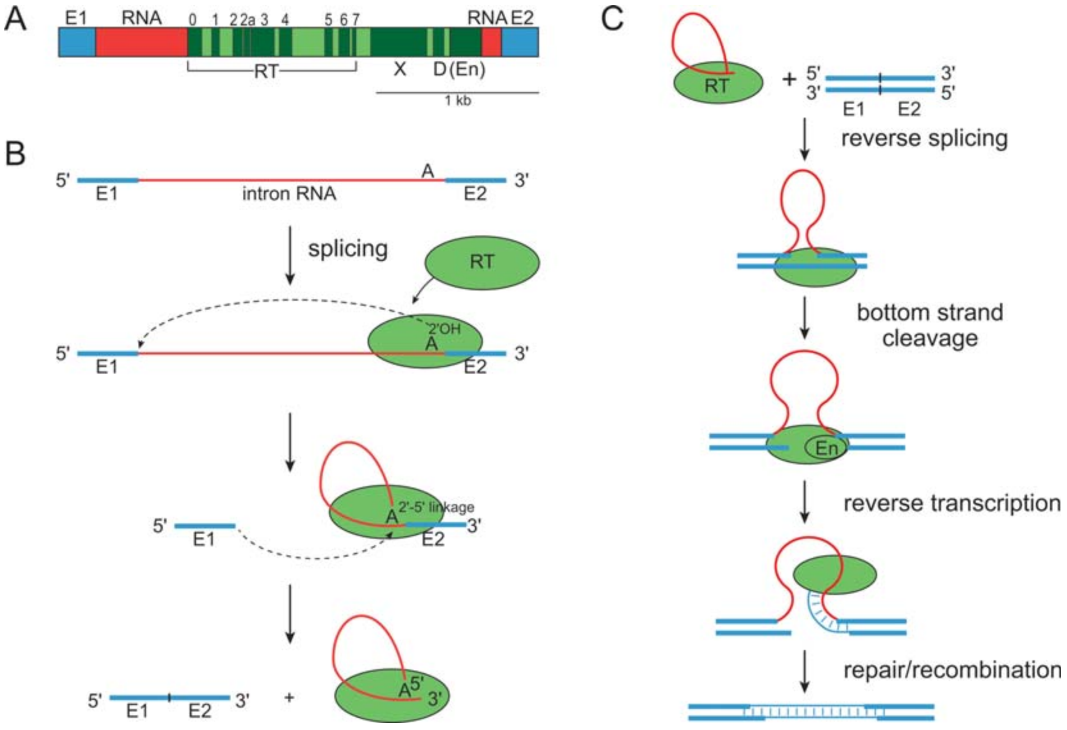
retroelements. The aim of this article is to summarize what is known about this enigmatic universe of retroelements in bacteria and to consider their relationships to characterized retroelements.

GROUP II INTRONS: THE PROTOTYPICAL BACTERIAL RETROELEMENT

Group II introns are the prototypical retroelements in bacteria because they are the most abundant and best understood (reviewed in references 20, 21, 22, 24, and 25). They consist of a ~500 to 800 bp sequence corresponding to a catalytic, self-splicing intron RNA (ribozyme) and an internally encoded ~1.0 to 1.5 kb ORF that is translated into an intron-encoded protein (IEP) (Fig. 1A). The IEP is an RT, containing seven conserved sequence blocks that are alignable across RT types, as well as an X domain that is structurally analogous to the polymerase's thumb domain but is not conserved in sequence. Downstream of the X domain is a DNA-binding domain (D), and sometimes, an endonuclease (En) domain (Fig. 1A). The RT and X domains together promote splicing of the intron RNA, while all four IEP domains participate in the mobility reaction (26, 27, 28, 29, 30, 31).

In bacterial cells, the intron is initially transcribed along with its exons as part of a precursor mRNA transcript, after which the intron RNA folds into a three-dimensional structure that catalyzes splicing (32, 33, 34, 35, 36, 37). Self-splicing of the intron occurs *in vitro* under conditions of elevated magnesium and salt; however, to efficiently splice *in vivo*, the IEP must first be

FIGURE 1 Group II introns. (A) The genomic structure of a group II intron consists of sequence for an RNA structure (~500 to 800 bp; red boxes) and an ORF for an intron-encoded protein (green). The protein contains a reverse transcriptase (RT) domain with motifs 0 to 7, an X/thumb domain, a DNA-binding domain (D), and sometimes, an endonuclease domain (En). The intron is flanked by exons E1 and E2 (blue). The structure is drawn to scale for the Ll.LtrB intron of *Lactococcus lactis*. (B) After transcription of the intron, the intron-encoded protein is translated from unspliced transcript and binds to the RNA structure to facilitate a two-step splicing reaction, yielding spliced exons and an RNP consisting of the RT and intron lariat RNA. (C) The RNP inserts intron sequence into new genomic targets. To do this, the RNP binds to the double-stranded DNA target, the intron lariat reverse splices into the top strand, and the En domain cleaves the bottom strand to produce a primer that is reverse transcribed by the RT. Cellular repair activities convert the insertion product to dsDNA. doi:10.1128/microbiolspec.MDNA3-0058-2014.f1



translated from the unspliced mRNA and bind to the intron RNA to help it achieve its catalytic conformation (27, 28, 30) (Fig. 1B). After splicing, the protein remains bound to the intron lariat, forming a stable ribonucleoprotein (RNP) particle.

Mobility of group II introns has been well studied. The overall process is called retrohoming, and it occurs through the mechanism of target-primed reverse transcription (TPRT) (20, 38, 39, 40). The TPRT mechanism is initiated by reverse splicing of the lariat RNA into the top strand of a double-stranded DNA target, followed by cleavage of the bottom strand by the En domain of the IEP to form a primer, and finally reverse transcription of the inserted intron (Fig. 1C). The final steps of the integration, which vary across organisms, are carried out by cellular repair mechanisms to generate double-stranded DNA (41, 42, 43). Some IEPs lack the En domain, and require an alternative primer for bottom strand synthesis (44). This has been shown to be a nascent DNA strand provided by a replication fork (45). Other variations of mobility have been described as well (46, 47, 48, 49, 50, 51).

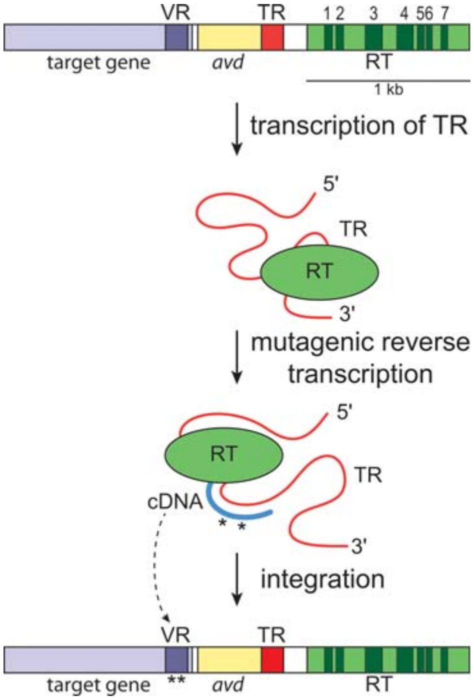
An important characteristic of group II introns in bacteria is that they behave primarily as retroelements rather than introns. Their selfish nature is evident in several ways. First, the introns are generally excluded from housekeeping or conserved genes, suggesting that they inhibit gene expression in some way (52, 53). Second, over half of intron copies in bacteria are truncated or nonfunctional, with the introns often located among or within other mobile DNAs, suggesting a migratory rather than stable lifestyle (52, 54, 55). These features contrast with group II introns in mitochondrial and chloroplast genomes, where the introns are located in housekeeping genes, and many are nonmobile splicing units (56). Third, the distribution of group II introns across bacterial species and strains is sporadic (44, 54, 57, 58). For example, related strains of *Escherichia coli* can harbor from one to 15 copies of group II introns (59), and some genomes contain over 20 identical copies of an intron (60).

There are limited exceptions to the introns' selfish character. Some bacterial introns are immobile (54, 61) or are located in housekeeping genes such as RecA, DNA polymerase III, helicases, and DNA repair enzymes, where they presumably do not impair gene expression (53). Still, the overall pattern of group II introns in bacterial genomes indicates that the introns survive by constant movement, spreading faster than they are lost, as opposed to taking up residence in conserved genes, as occurs in organellar genomes where splicing can potentially help to regulate gene expression. Notably, group II introns are by far the most numerous of RT types in bacteria (below), consistent with their robust retromobility.

DGRs: RETROELEMENTS THAT EVOLVED A USEFUL FUNCTION

DGRs do not spread selfishly but they have a clearly useful function in creating sequence diversity in a target gene (62). DGRs are comprised of multiple components that make up a functional cassette of genes: an RT gene, a 100 to 150 bp TR (template repeat) gene, a target gene ending in a VR (variable region) sequence that is ~90% identical to the TR sequence, and usually, the accessory gene *avd* (accessory variability determinant) (Fig. 2).

FIGURE 2 Diversity-generating retroelements (DGRs). A DGR consists of a reverse transcriptase (RT) gene with seven motifs, a target gene with a C-terminal variable region (VR), a template repeat (TR), and usually, an accessory variability determinant gene (*avd*) (drawn to scale for the *Bordetella* phage DGR [23]). The RT's thumb motif is not defined in sequence but presumably would be present downstream of motif 7. For the mutagenic homing reaction, the RT reverse transcribes the TR transcript and the resulting cDNA is integrated into the target gene to replace the previous VR sequence. During this process, each A in the TR sequence is mutagenized to any nucleotide, producing directed randomization of the VR sequence in the target gene. doi:10.1128/microbiolspec.MDNA3-0058-2014.f2



During the so-called mutagenic retrohoming reaction, the RT reverse transcribes the TR transcript, during which every A in the TR template is mutagenized to any nucleotide in the resulting cDNA. The cDNA is integrated into the target gene's DNA, replacing the VR sequence and creating randomization in the amino acid sequence at the C-terminus of the target protein (23, 63, 64, 65, 66). In the case of the *Bordetalla* phage, the target protein is the phage's tail protein, and specifically the region that adheres to the cell surface of *Bordetella* during infection (67, 68). This is beneficial to the phage because *Bordetella* has two growth phases, virulent and avirulent, each bearing a characteristic cellular surface (69). The DGR allows the phage to adapt its tropism to the changing surface of its host bacterium by creating sequence diversity in the region of the tail protein that binds to the bacterial surface. Other DGR examples are known that are encoded on a bacterial chromosome rather than by a phage. Among these is the chromosomally encoded DGR of *Legionella pneumophila*, which creates sequence diversity in the C-terminus of a variable cell surface lipoprotein (70). Thus, DGRs represent a general mechanism of adaptation in bacteria that can benefit either phages or bacteria.

The genomic components of DGRs can vary considerably. Some DGRs lack the *avd* gene or have another accessory gene belonging to the helicase and RNase D C-terminal (HRDC) family. DGRs can have either one or two target genes, and there is variation as well in the order and orientations of the RT, TR, *avd*, and target genes (62). Relevant to this article is the fact that even without experimental validation, DGRs can be identified bioinformatically in genomic sequences through the detection of the two repeat sequences, TR and VR (~90% identity), which are adjacent to an RT gene, with the VR sequence bearing A-to-N differences relative to the TR sequence (71). This illustrates how a retroelement and its potential properties can be identified and/or classified based on sequence alone, a pertinent point when considering the many uncharacterized RT and RT-like sequences (below).

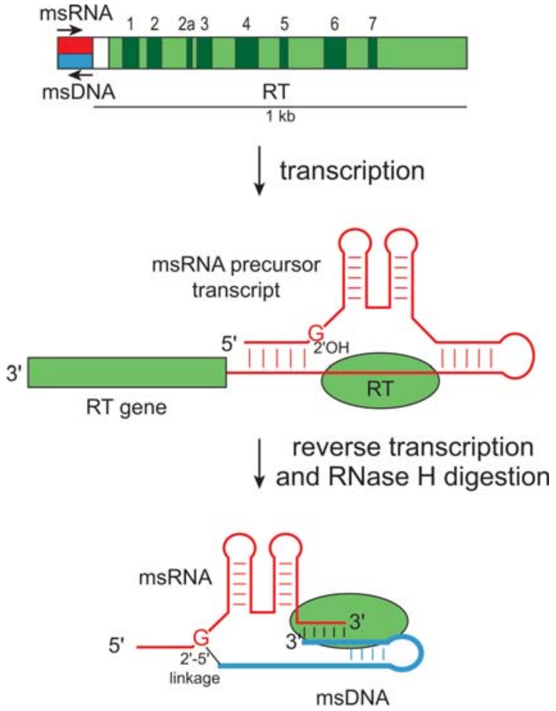


FIGURE 3 Retrons. A retron consists of an inverted repeat sequence corresponding to msRNA and msDNA genes, and a reverse transcriptase (RT) with seven conserved motifs (drawn to scale for retron Ec86 [77]). The thumb domain is presumably located directly downstream of motif 7. All three genes are transcribed in a single transcript and the RT binds to the RNA structure formed by the inverted repeat sequence. A specific G residue presents a 2′OH that acts as the primer for reverse transcription. After removal of the RNA template by cellular RNase H, the final msDNA consists of one RNA and one DNA linked by a 2′OH bond, and base paired at the 3′ ends of both. doi:10.1128/microbiolspec.MDNA3-0058-2014.f3

RETRONS: RETROELEMENTS IN SEARCH OF A FUNCTION

Retrons consist of an RT gene named *ret* and an adjacent inverted repeat sequence corresponding to the overlapping genes for the RNA and DNA segments of an msDNA (Fig. 3) (18, 72, 73, 74, 75). Both the RT and inverted repeat sequences are transcribed together as a single operon. The translated RT binds to the RNA's inverted repeat sequence. A specific G residue within the structure provides its 2′OH as a primer for reverse transcription (76, 77). The RNA is partially reverse transcribed, with most of the RNA template removed by a cellular RNase H activity. The resulting chimeric RNA-DNA molecule is covalently linked via the 2′OH priming bond and is also base paired at the 3′ ends of the RNA and DNA where the RNA is not digested by RNase H. The RT remains bound to this structure after its formation.

Retromobility of retrons was suspected from the start, but this has not been demonstrated. The only experimental evidence for mobility is for the retron Ec73, which resides within the cryptic prophage φR73 in *E. coli* (78). When φR73 was mobilized by coinfection with a helper phage, the retron spread to another strain of *E. coli* and formed msDNA in the new host. This observation is consistent with retrons not being independently mobile, but being dispersed passively among strains or species. In another mobility-related observation, a comparison of different *E. coli* genome sequences led to the conclusion that the Ec107 retron replaced a 34 bp palindromic sequence through a *de novo* insertion; however, the mechanism was not elucidated (79). Similarly, two *Vibrio* species contain distinct retrons substituted at the same chromosomal locus, but again the mechanism of integration, loss, and/or replacement is unclear (80).

While the function of retrons remains unknown, there are two distribution patterns that are relevant. The first is vertical inheritance, which is found in soil-dwelling myxobacteria. The retron, Mx162, was found in all *Myxococcus xanthus* strains tested and in nearly all other myxobacterial species, while Mx65 is additionally found in some strains (the numbers refer to the length of the DNA component of the msDNA) (81, 82). The conclusion of vertical inheritance is supported by similar codon usage for the RTs and the host genomes (75). Vertical inheritance of the retrons implies that they carry out an essential—or at least a useful—function in their bacterial hosts.

The second distribution pattern is exemplified by retrons in *E. coli*, where presence of retrons is patchy. Most strains do not carry retrons, but there are seven distinct retrons among isolates. As a rough measure of frequency among *E. coli* strains, only 11 strains of the 72-strain ECOR collection contain a retron (83). The *E. coli* retron sequences are not closely related to each other, and they appear to have spread via horizontal transfers, a hypothesis supported by their codon usage, which differs from that of cellular genes (75). Because they are not universally present in the strains, the *E. coli* retrons are unlikely to carry out an essential function, although they may still carry out a useful function. The dynamics of their horizontal spread may be analogous to the propagation of antibiotic resistance genes among bacteria.

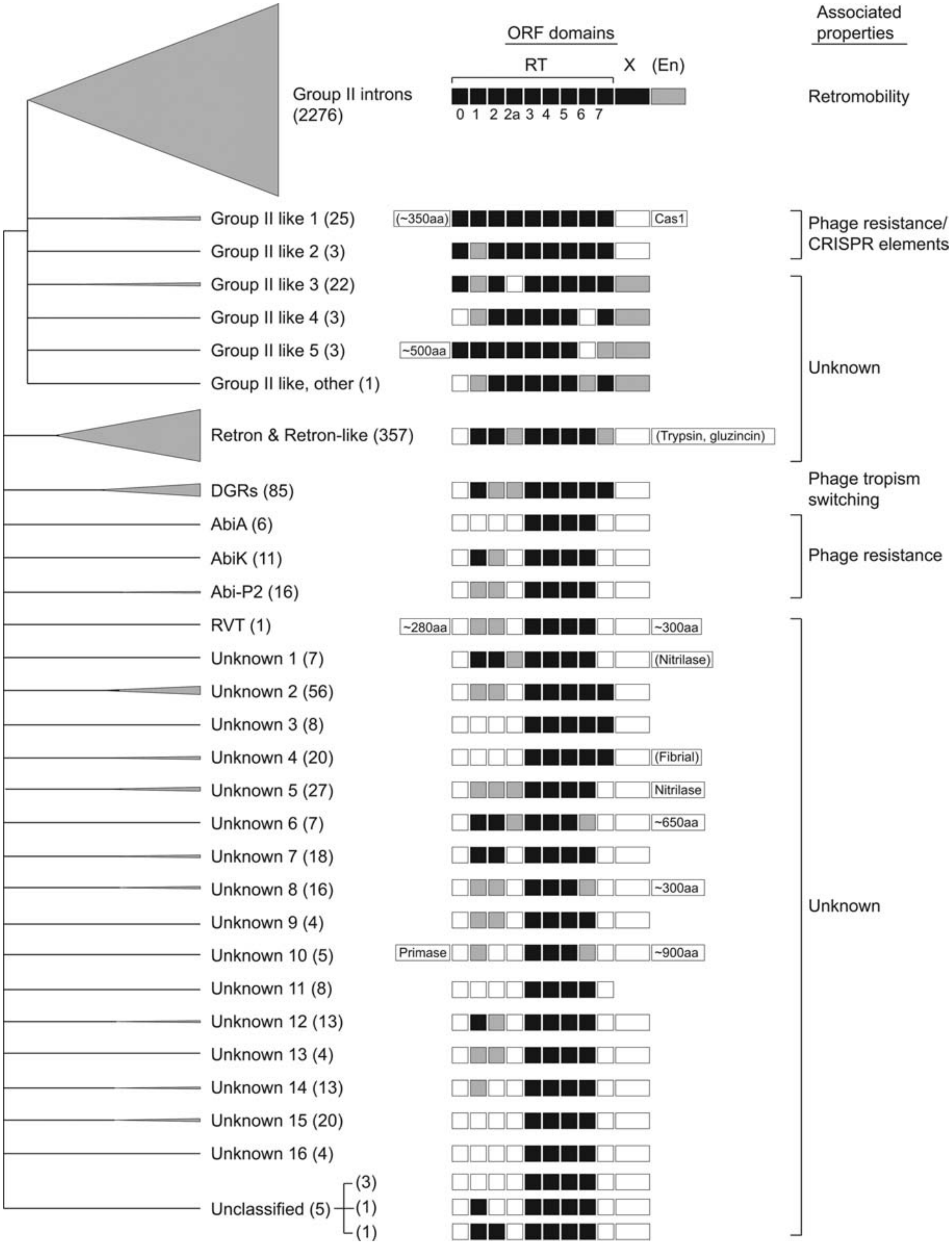
Two useful functions have been proposed for retrons, neither of which provides a satisfying general explanation for their existence. First, some retrons in *E. coli* have mutagenic effects on their host strains. The retrons, Ec83 and Ec86, have stem loops with mismatched base pairs that sequester the mismatch-repairing MutS protein, thereby leading to a higher mutation frequency. In certain contexts, an increased mutation rate might aid in bacterial adaptation (84, 85). Opposing this rationale is the fact that not all retrons contain mismatched bases in their stem-loops (86). A second potential function is based on the anecdotal observation that all 12 pathogenic *Vibrio* strains out of 21 isolates tested contain the retron Vp96, providing an apparent correlation between the retron and pathogenicity in this bacterium (86). However, the correlation does not hold for pathogenic and nonpathogenic strains of *E. coli*. Together, it would appear that retrons have an, as yet, unidentified biological function. Supporting this prediction is the presence of a putative protease domain at the C-terminus of some retron RTs (reference 87 and below), which suggests an undetected biochemical reaction of retrons.

A UNIVERSE OF UNCHARACTERIZED RTs AND RT-RELATED ENZYMES IN BACTERIA

Over the last decade, mounting sequence data have led to the identification of many new RTs and RT-like sequences in bacterial genomes. Two studies in 2008 reported numerous examples of novel bacterial RTs and retroelements (87, 88). An update of the compilation presented in reference 88 is summarized in Fig. 4, which is based on a set of 3,044 bacterial and archaebacterial RT sequences collected from GenBank (L. Wu and S. Zimmerly, unpublished). Despite their considerable sequence diversity, all putative RTs were identified as belonging to the superfamily of RTs (as opposed to other polymerases) by the NCBI Conserved Domain Database (CDD) (89). The classifications depicted in Fig. 4 are based on a combination of criteria, including the alignment of RT motif sequences, the presence of domain motifs or sequence extensions appended to the RT domain, and published data about characterized RTs (see reference 88 for a more detailed explanation).

As observed previously, approximately 90% of RT copies in bacterial genomes belong to characterized types of retroelements: group II introns (75%), retrons (12%), and DGRs (3%). Twenty-five additional groupings are now defined, which include five classes designated group II-like (G2L), which are related to group II introns but do not have an associated ribozyme structure. Due to the lack of experimental characterization, the classifications are not meant to imply unique functions for each class, but rather are an attempt to organize the diversity of uncharacterized RT-like sequences.

Five RTs remain unclassified, meaning that they have no close relatives among the set of 3044 RTs (E values to closest relatives range from e-6 to e-26 in BLASTP searches [Wu and Zimmerly, unpublished]). The five orphan sequences presumably represent classes that will become better defined when more DNA sequences are available. The existence of orphans indicates that


ORF domains
Associated properties
RT
X
(En)
0 1 2 2a 3 4 5 6 7
Group II introns (2276)
Retromobility
Group II like 1 (25)
(~350aa)
Cas1
Group II like 2 (3)
Phage resistance/ CRISPR elements
Group II like 3 (22)
Group II like 4 (3)
Group II like 5 (3)
~500aa
Group II like, other (1)
Retron & Retron-like (357)
(Trypsin, gluzincin)
Unknown
DGRs (85)
Phage tropism switching
AbiA (6)
AbiK (11)
Abi-P2 (16)
Phage resistance
RVT (1)
~280aa
~300aa
Unknown 1 (7)
(Nitrilase)
Unknown 2 (56)
Unknown 3 (8)
Unknown 4 (20)
(Fibrial)
Unknown 5 (27)
Nitrilase
Unknown 6 (7)
~650aa
Unknown 7 (18)
Unknown 8 (16)
~300aa
Unknown 9 (4)
Unknown 10 (5)
Primase
~900aa
Unknown 11 (8)
Unknown 12 (13)
Unknown 13 (4)
Unknown 14 (13)
Unknown 15 (20)
Unknown 16 (4)
Unclassified (5)
(3)
(1)
(1)
Unknown

sequence databases have not saturated coverage of the RTs, and we do not yet know all varieties of bacterial RTs and RT-like proteins that exist in nature. The sequences and alignments of all classified and unclassified RTs can be accessed in the Supplementary Data file.

All 3,044 putative RTs align clearly with other RTs across RT motifs 3, 4, and 5, which correspond to the palm and finger domains of the polymerase, and include the three aspartate residues that coordinate the two divalent metals at the active site (Fig. 5). In addition, most RTs have a recognizable motif 6, whose conserved lysine is also an active site residue that acts as a general acid to facilitate the pyrophosphate leaving group (90). The remaining motifs (0, 1, 2, 2a, and 7) are less conserved, with motifs 0 and 2a being the least conserved (Fig. 4 and Fig. 5). Apart from the core polymerase structure (motifs 3 to 6), the putative RTs can be predicted to have a ~50 to 150 aa thumb domain located directly downstream of the RT domain. Thumb domains have little or no sequence conservation across RT types, but their location in sequence and three-dimensional structure is generally conserved across polymerases (91). An exception is unknown class 11, which does not have sufficient C-terminal sequence to encode a thumb domain. Overall, all RT classes except for unknown class 11 are expected to have very similar three-dimensional RT core structures with a thumb domain, consistent with their common functions as polymerases (92).

About two-thirds of uncharacterized RT classes have no motifs or extensions in addition to the predicted RT structure (Fig. 4). These RTs range in size from ~200 aa (unknown class 11) to ~650 aa, with the variability in size due mainly to indels within motifs 1 to 7 and X. None of the indel sequences match domains defined in either the CDD database of NCBI or the Pfam database (89, 93; Wu and Zimmerly, unpublished). About a third of the RT classes have sizeable extensions, containing either conserved domain motifs identified by CDD or Pfam, or minimally conserved sequence extensions of >300 aa found among members of the class. In all there are 11 domain architectures of putative RT proteins, with six conserved domains appended to RT domains and five less conserved extensions of >300 amino acids (Fig. 4).

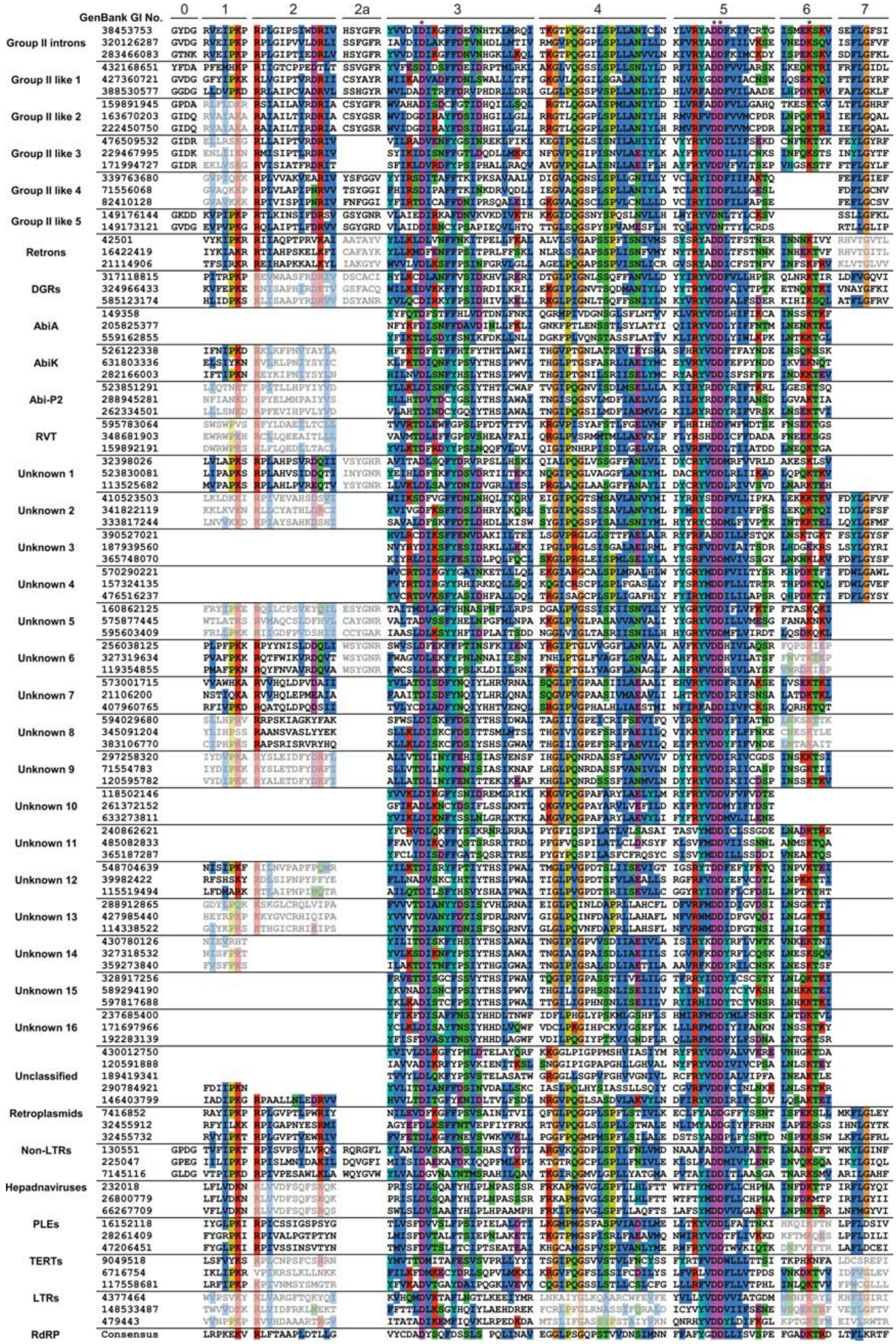
RETROELEMENTS ASSOCIATED WITH PHAGE IMMUNITY

While most RT classes in Fig. 4 are wholly uncharacterized, there is limited data for several proteins that give indications of their properties. Three bacterial RT-related proteins are implicated in phage resistance: AbiA, AbiK, and Abi-P2. AbiA and AbiK are encoded by *Lactococcus lactis* plasmids and provide phage immunity through abortive infection (Abi), a process in which phage DNA enters a bacterium but phage multiplication is subsequently blocked (Fig. 6A). A third RT-related protein, found in the P2 prophage of some *E. coli* strains and here called Abi-P2, operates through phage exclusion, which is a process that prevents new phage infections of cells that contain the protective mechanism. For AbiK, it has been demonstrated that the RT protein alone confers phage resistance. This issue is not clarified for AbiA or Abi-P2, but the genetic loci conferring resistance are small (<3 kb).

AbiK

The *abiK* gene is encoded by the natural plasmid pSRQ800, which was discovered in a *L. lactis* strain isolated from raw milk (94). The single copy, constitutively transcribed gene, confers resistance against the three major classes of lactococcal phages (936, c2, and P335), typically at a level of six orders of magnitude over sensitive cells (95). The conserved RT motifs are in the N-terminal half of the 599 aa AbiK protein, while a C-terminal region of ~275 aa contains the presumed thumb domain of the polymerase as well as additional sequence with no identified domain motifs (Fig. 6B).

FIGURE 4 Classes of reverse transcriptases (RTs) and RT-like sequences in bacterial genomes. The figure is an update of Fig. 1 in reference 88. A set of 3,044 RTs were collected from GenBank and classified according to alignability of RT motif sequences, phylogenetic analyses, and the presence of additional domains. The number of members in each class is indicated in parentheses and by the area of the gray triangles. RT motifs are denoted by boxes that are either black (clearly alignable with group II introns), gray (ambiguously alignable), or white (not alignable, although an analogous structure is expected to be present). Sizeable extensions to the RTs are indicated by amino acid sizes, and are in parentheses when the motif is present in fewer than half of the examples. Protein motifs identified by either CDD of NCBI or Pfam are Cas1, trypsin, gluzincin, nitrilase, fimbrial, and primase. Biological properties associated with the different classes are indicated to the right. doi:10.1128/microbiolspec.MDNA3-0058-2014.f4


GenBank GI No.
0
1
2
2a
3
4
5
6
7
Group II introns
Group II like 1
Group II like 2
Group II like 3
Group II like 4
Group II like 5
Retrons
DGRs
AbiA
AbiK
Abi-P2
RVT
Unknown 1
Unknown 2
Unknown 3
Unknown 4
Unknown 5
Unknown 6
Unknown 7
Unknown 8
Unknown 9
Unknown 10
Unknown 11
Unknown 12
Unknown 13
Unknown 14
Unknown 15
Unknown 16
Unclassified
Retroplasmids
Non-LTRs
Hepadnaviruses
PLEs
TERTs
LTRs
RdRP
Consensus

Nevertheless, the C-terminal region has an essential function because a 42 amino acid C-terminal deletion eliminates abortive infection (94). Point mutations in the RT motifs, including the predicted catalytic aspartate residues, disrupt or abolish abortive infection, consistent with AbiK being a polymerase (95, 96).

AbiK expressed in *E. coli* and purified as a GST fusion protein showed polymerase activity *in vitro*, but surprisingly, not the properties of a reverse transcriptase. Even without an RNA or DNA template, long DNAs were polymerized by AbiK having a "random" sequence, making AbiK's activity analogous to a terminal transferase (96). The newly synthesized DNA was covalently attached to the AbiK protein, because an amino acid of AbiK was the primer for polymerization. A similar self-priming reaction has been observed for the hepadnavirus RT (97), which uses the OH of a tyrosine residue as a primer. For the hepadnavirus enzyme, a priming domain is located in an N-terminal region upstream of the RT domains. By comparison, the priming site on the AbiK protein may lie in the 250 aa C-terminal domain, because the domain has an essential but undefined function.

Interestingly, phage mutants can be readily isolated in the laboratory that acquire resistance to the AbiK antiphage mechanism, giving clues as to how AbiK interacts with the phage replication process. For four different phages, AbiK-resistant mutations mapped to the *sak* genes (sensitivity to AbiK), *sak1*, *sak2*, *sak3*, and *sak4* (98). The four *sak* genes are not closely related to each other, but *sak1* and *sak2* have sequence similarity to single-strand annealing proteins (SSAPs) of the RAD52 and Erf families, respectively, while *sak4* is related to RAD51 recombinases (99, 100). Sak3 protein, having no known motifs, was shown experimentally to have SSAP activity, as was Sak1 (100, 101, 102). Taken together, the Sak proteins appear to participate in a step of phage genome replication.

The interaction between AbiK and Sak that blocks phage multiplication is predicted to be a functional interaction rather than direct binding. This conclusion is based on the diverse sequences of the four *sak* genes and the fact that the resistance mutations occur throughout the *sak* sequences (98). However, the precise nature of the interaction remains unclear, as does the overall mechanism by which AbiK mediates immunity to phages. Also unresolved is whether AbiK exhibits true RT activity (i.e., RNA-dependent DNA polymerization) under conditions not tested or whether AbiK is a highly derived enzyme that has lost its RT activity and evolved a new polymerase reaction.

AbiA and *Abi-P2*

AbiA shares many similarities with AbiK, although its sequence is distantly related. It is encoded by the lactococcal plasmid pTR2030, and at 628 amino acids, AbiA is comparable in size to AbiK, with a C-terminal extension lacking protein motifs. AbiA protects against the major classes of lactococcal phages with a sensitivity profile that is similar but not identical to the set of AbiK-sensitive phages (103, 104). AbiA is similarly inferred to act at the stage of phage DNA replication, because DNA products do not accumulate in infected bacteria (104, 105).

Phage mutations lending resistance to the AbiA mechanism mapped to a gene (ORF245) having motifs for RecA-like NTPases and phage P loops (nucleotide binding), which resemble features of the Sak proteins. A second group of resistance mutations mapped to an intergenic inverted repeat sequence lying 500 bp upstream of phage ORF245 (105), for which there is no reported parallel for AbiK.

The *abi-P2* gene is not plasmid-encoded but is found in two P2 prophages in *E. coli* strains ECOR30 and ECOR58 of the ECOR collection. The two *abi-P2* genes

FIGURE 5 Amino acid alignment of reverse transcriptase (RT) motifs 0 to 7 for different classes of RTs in bacteria and eukaryotes. Three example sequences are presented for each class for motifs 0 to 7. Sequences in black lettering and bold color shading are clearly alignable with group II introns, while sequences in gray and light color shading are ambiguously alignable. Sequences not shown indicate unalignability with group II RT sequence motifs, although similar structures are likely present in the proteins. Positions with >30% identity across the entire alignment are back-shaded in colors to highlight the most conserved residues across RT classes. For comparison, the sequences of major classes of eukaryotic RTs are listed, as well as a consensus sequence for the Pfam group, RNA-dependent RNA polymerase (RdRP) 1, which among RdRPs has the greatest alignability to group II RTs. Asterisks above the alignment mark the three catalytic aspartate residues in motifs 3 and 5 and the active site lysine in motif 6. DGRs, diversity-generating retroelements; LTR, long terminal repeat; PLEs, Penelope-like elements; TERT, telomerase reverse transcriptase. doi:10.1128/microbiolspec.MDNA3-0058-2014.f5

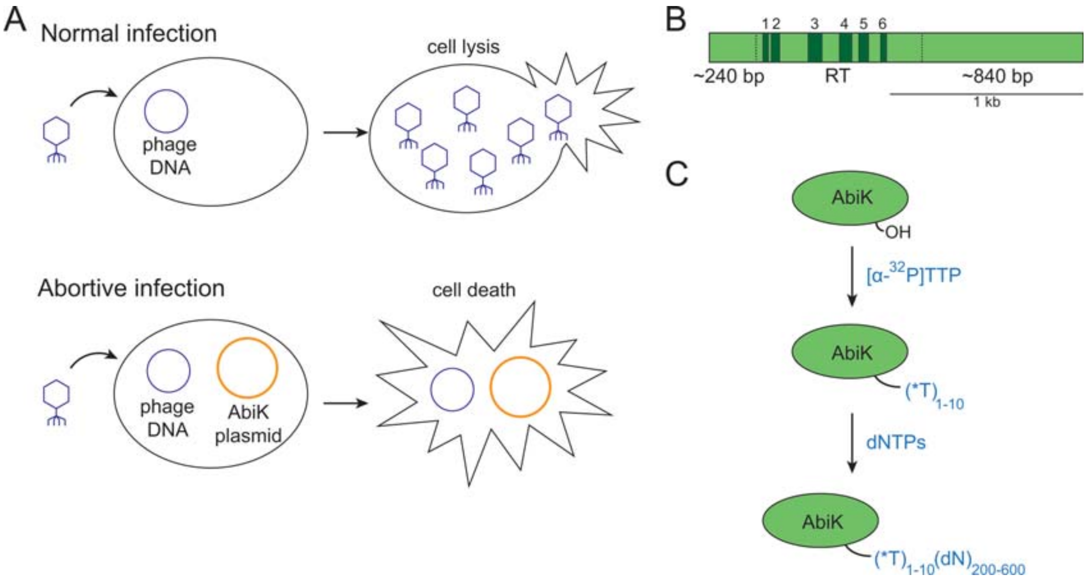


FIGURE 6 AbiK and abortive infection. (A) During abortive infection by AbiK, the phage injects its DNA into a *Lactococcus lactis* cell, but the multiplication cycle is blocked through an undefined mechanism by the AbiK protein. Although not necessarily a suicide mechanism, most cells still die and infective phages are not released. (B) The AbiK protein contains reverse transcriptase (RT) motifs 1 to 6 and 7 is essentially unalignable with group II introns. Estimates for the boundaries of the RT domain and thumb domain of the polymerase are indicated with dotted lines. In addition, the proteins contain a short N-terminal extension and a ~840 bp C-terminal extension (drawn to scale for AbiK of *Lactococcus lactis* [96]). (C) Purified AbiK protein has a terminal transferase activity, with the synthesized DNA becoming covalently linked to the AbiK protein. In the "label" reaction with low concentrations of [α-^{32}P]TTP, AbiK produce a short poly T DNA that is covalently linked to the AbiK protein. In the "chase" reaction, high concentrations of dNTPs cause polymerization of hundreds of nucleotides of heterogeneous sequence. doi:10.1128/microbiolspec.MDNA3-0058-2014.f6

share ~75% identity. When ORF570 of ECOR30 was expressed on a plasmid in *E. coli* strain BL21(DE3), the strain was rendered resistant at levels of 10^{-7} relative to infection by T5 phage, but there was no effect on infection by lambda, T2, T4, or T6 phages (106). The expressed ORF570 was reported to have RT activity in a biochemical assay, but the assays were with crude extracts under conditions where artifactual incorporation can occur (96).

Together, the AbiK, AbiA, and Abi-P2 all give similar properties of resistance against phage infections. It is tempting to suspect that they use similar mechanisms; however, this cannot be automatically assumed because their RT sequences are quite distantly related (Fig. 5). For example, only RT motifs 4, 5, and 6 align convincingly between AbiA and AbiK in a pairwise alignment (Wu and Zimmerly, unpublished). Interestingly, homologs of AbiK, AbiA, and Abi-P2 are present in a wide range of species across multiple eubacterial phyla, including in clinical isolates (107; Wu and Zimmerly, unpublished). At the very least, AbiA, AbiK, and Abi-P2 point to a family of mechanisms of phage resistance conferred by RT-related proteins, which are not yet understood.

rvt RETROELEMENTS

Discovered only in 2011, *rvt* elements are found primarily in eukaryotes, but there are a few examples in bacteria (108). Their distribution is sporadic, with most examples found in diverse fungi. Homologs are identified in animals (rotifers), plants (moss), stramenophiles (water molds), and a bacterium (*Herpetosiphon aurantiacus*). The RVT proteins are 800 to 1,000 aa long, and include a ~300 aa N-terminal extension and ~200 aa C-terminal extension, neither of which contains a discernible protein motif (Fig. 7A). In most cases, *rvt* genes are found as single copies in genomes, although sometimes there are two nonidentical copies (108).

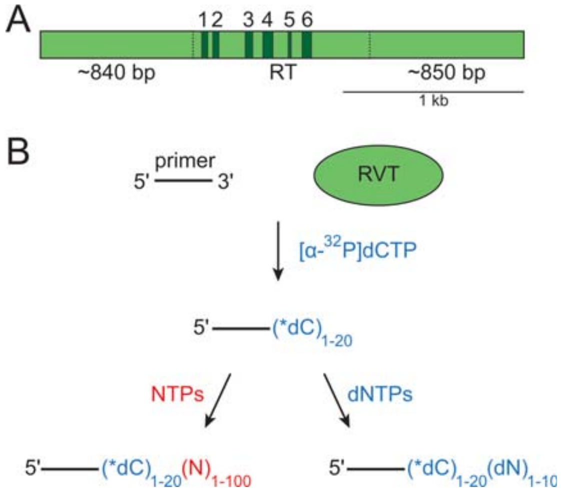


FIGURE 7 The *rvt* element. (A) The RVT ORF contains reverse transcriptase (RT) motifs 1 to 6, while motif 7 and thumb domains are unalignable with group II RTs but are presumably present in the polymerase structure. Estimates for the boundaries of the RT domain and thumb domain of the polymerase are noted with dotted lines. The large N-terminal and C-terminal extensions have no detectable protein motifs (drawn to scale for the *N. crassa* RVT [108]). (B) Purified RVT protein has terminal transferase activity that requires an RNA or DNA primer and has a preference for nucleoside triphosphates (NTPs) over deoxynucleoside triphosphates (dNTPs). When purified RVT protein is incubated with [α-^{32}P]dCTP, a short sequence is synthesized, which is extended by either NTPs or dNTPs in a chase reaction. doi:10.1128/microbiolspec.MDNA3-0058-2014.f7

The only full-length bacterial *rvt* gene is found in *H. aurantiacus*, a predatory, filamentous gliding bacterium of the phylum Chloroflexi. Additionally, two partial sequences are reported from environmental samples deemed to be bacterial. The closest relatives of the *H. aurantiacus rvt* gene are not bacterial RTs but eukaryotic RTs, suggesting that the *rvt* gene was transferred horizontally from a eukaryote to a bacterium rather than vice versa.

The *rvt* gene of *Neurospora crassa* has been studied genetically and biochemically. Interestingly, an *rvt* knockout in *N. crassa* exhibited no phenotype, revealing a nonessential role. Under normal laboratory growth conditions, the gene was transcribed at low levels, but transcription was elevated 50-fold when *N. crassa* was grown in the presence of drugs that disrupted translation or blocked histidine biosynthesis, implying that the element is responsive to stress conditions (108).

Like AbiK, purified *N. crassa* RVT showed terminal transferase activity and inability of template-dependent polymerization (Fig. 7B). Contrasting with AbiK's polymerization activity, RVT showed a strong preference for nucleoside triphosphates (NTPs) over deoxynucleoside triphosphates (dNTPs). A primer of either RNA or DNA was inferred to be used in the *in vitro* reactions but was not defined. Similar to AbiK, it remains possible that RVT possesses RNA-dependent DNA polymerization activity under conditions not tested. Also left unresolved is the biological function of the element. Paralleling retrons, the *rvt* elements presumably benefit their hosts through an unknown activity in order to account for their maintenance as single-copy genes over evolutionary time in the absence of retromobility.

RTs ASSOCIATED WITH CRISPR-Cas SYSTEMS

Although not experimentally investigated, a subset of bacterial RTs can be concluded to contribute to cellular defense against phages and foreign DNA through CRISPR/Cas systems, which are a large family of adaptive immunity systems in bacteria (109, 110, 111). The RTs classified as G2L1 and G2L2 are associated with *cas1* genes of CRISPR/Cas loci. Most G2L1 RTs are fused to the ORF of *cas1*, whereas G2L2 RTs are freestanding genes located adjacent to *cas1* genes (88). The *cas1* gene is universally present in all subtypes of CRISPR/Cas systems and encodes a metal-dependent nuclease that forms a complex with the Cas2 protein. The resulting nuclease activity is essential for integrating new spacer sequences into CRISPR arrays (112), giving the bacterium immunity against phages or plasmids containing those sequences. It seems likely that the CRISPR-associated RTs are involved in the insertion of new spacer sequences into CRISPR arrays using a mechanism similar to, and derived from, group II intron retromobility.

FUNCTIONS OF THE OTHER UNCHARACTERIZED RETROELEMENTS IN BACTERIA

Apart from group II introns, there is no evidence for active retromobility of bacterial RTs. None are present in genomes in more than one (identical) copy, nor are there easily discernible repeats flanking the RT genes that could be remnants of insertion events (Wu and Zimmerly, unpublished). If the RTs are, in fact, part of mobile DNAs, the mobility levels would have to be much lower than for group II introns or most other mobile DNAs. It is tempting to predict that most, if not all, of the putative RTs carry out useful functions for their host organisms and are retained and spread horizontally for the benefits they provide, rather than their retromobility.

A number of uncharacterized RTs have appended protein motifs that give hints as to their biochemical reactions. Peptidase motifs are found at the C-terminus of two subsets of retron RTs—either a trypsin-2 motif or a gluzincin (zinc-dependent) proteinase motif (87) (Fig. 4). Unknown classes 1 and 5 have an appended C-terminal nitrilase (C-N hydrolase) motif (87). Proteins with nitrilase domains break C-N bonds in nonpeptide cleavage reactions and they are implicated in processes of small molecule metabolism, detoxification, signaling, and posttranslational modification (89). The fimbrial domain found in unknown class 4 RTs is a pilus-related protein that serves structural roles in many processes including conjugation, intracellular trafficking, adhesion, and secretion (88). The primase domain in unknown class 10 suggests a concerted priming and reverse transcription reaction by the two polymerase domains of the protein.

A second source of information about the function of genomic elements can come from flanking genes or gene neighborhoods (87). Interestingly, RTs of unknown class 3 and unknown class 8 are found adjacent to each other in genomic sequences (87), suggesting that they operate together, perhaps forming a heterodimer and/or performing two distinct polymerization reactions in a process. More recent attempts to identify homologous flanking genes within the unknown classes did not identify additional conserved flanking genes as candidate cofactors for RT functions, although it is possible that such genes were overlooked (Wu and Zimmerly, unpublished).

Overall, despite the clues provided by the motifs appended to RT domains, there remains an exceedingly wide range of possibilities for the reactions and functions of the RTs. Given the range of properties uncovered for some of the example RTs, often involving phage resistance (Fig. 4), it seems highly likely that they are involved in novel biochemical reactions. We can look forward to interesting biochemistry and biological phenomena from these elements.

EVOLUTIONARY CONSIDERATIONS

The origin and evolution of RT has been the subject of considerable speculation (8, 113, 114, 115, 116, 117). The enzyme has often been postulated to date back to the transition from the RNA to the DNA world (118, 119, 120). According to this scenario, the first polymerase would have consisted of RNA molecule (ribozyme), which exhibited RNA-dependent RNA polymerase (RdRP) activity (121). The RNA-only enzyme would have evolved into an RNP enzyme and it would have eventually been replaced by a protein-only polymerase. One hypothesis holds that the catalytic region of contemporary polymerases is a remnant of that ancient, substrate-binding polypeptide, which then gained catalytic activity and supplanted the ribozyme polymerase (118). RT polymerases would have derived from this ancient protein RdRP, and helped to drive the transition from ancient RNA-based life forms to modern organisms with DNA as the dominant genetic material.

It is worth keeping in mind that all right hand polymerases had a common ancestor. These include, in addition to RTs, the A, B, and Y families of replicative DNA polymerases (e.g., *E. coli* pol I, *E. coli* pol II, and human pol I, respectively), single-subunit DNA-dependent RNA polymerases (e.g., T7 RNA polymerase), and RNA-dependent RNA polymerases (RdRPs) (91, 92). A recent study superimposed representative polymerase crystal structures and revealed that all six families had a common structural core of 57 superimposable amino acids, corresponding to the heart of the palm domain and including the aspartate residues that coordinate the catalytic metal ions (92). It can be reasonably assumed that the ancestral RT contained this 57 aa structural core. On the other hand, it may be grist for debate whether the first protein polymerase is best represented by RdRPs or RTs, or was a distinct enzyme that was less specialized than all contemporary polymerases.

Bacterial RTs are usually considered to be older than eukaryotic RTs. Specifically, bacterial group II intron RTs are thought to have given rise to eukaryotic non-LTR RTs, because of similarities in their sequences and mechanisms (8, 116, 117). In their sequences, group II and non-LTR RTs share motifs 0 and 2a (in addition to the conserved RT motifs 1 to 7), making them sister families of RTs (122) (Fig. 5). Mechanistically, group II introns and non-LTR elements share the retromobility mechanism of TPRT, in which the DNA target is cleaved by the RT's endonuclease domain to form a primer for reverse transcription (38, 123). Given that group II introns are widely considered to be the ancestors of nuclear spliceosomal introns (124), it is quite plausible that their invasion of the ancestral nuclear genome gave rise to nuclear non-LTR retroelements as well as spliceosomal introns (38). Other classes of eukaryotic retroelements (e.g., LTR elements, hepadnaviruses), which have more elaborate mechanisms and machineries, might have derived from non-LTR retroelements (8, 116, 125).

In any case, the many new types of RTs and RT-related proteins in bacteria raise new possibilities for the evolution of reverse transcriptase. The number of bacterial classes raises the possibility that RT might have migrated multiple times from bacteria to eukaryotes (in addition to the group II/non-LTR migration) because eukaryotic RTs themselves are not closely related (Fig. 5). The low sequence similarity of eukaryotic RTs might reflect different ancestors rather than a sequential pathway of eukaryotic diversification and evolution from a non-LTR retroelement. Another provocative possibility is suggested by the fact that among bacterial RTs, only group II introns appear to be actively mobile. This suggests that the oldest RTs may have been non-replicative polymerases that performed useful functions in their bacterial hosts. From these RTs emerged a selfish DNA (group II intron) that was capable of multiplying through retrotransposition, which eventually gave rise to eukaryotic retroelements. Of course, the precise history of RT may be ultimately unknowable. Still, these bacterial RT-related enzymes expand the biochemical and biological properties possible for an ancestral reverse transcriptase, and are fodder for consideration of the origin and evolution of RTs and polymerases.

REFERENCES

1. **Baltimore D.** 1970. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226:**1209–1211.

2. **Temin HM, Mizutani S.** 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226:**1211–1213.

3. **Le Grice SF.** 2012. Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise. *J Biol Chem* **287:**40850–40857.

4. **Hohn T, Rothnie H.** 2013. Plant pararetroviruses: replication and expression. *Curr Opin Virol* **3:**621–628.

5. **Glebe D, Bremer CM.** 2013. The molecular virology of hepatitis B virus. *Semin Liver Dis* **33:**103–112.

6. **Roy-Engel AM.** 2012. LINEs, SINEs and other retroelements: do birds of a feather flock together? *Front Biosci (Landmark Ed)* **17:**1345–1361.

7. **Eickbush TH, Jamburuthugoda VK.** 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res* **134:**221–234.

8. **Eickbush TH.** 1994. Origin and evolutionary relationships of retroelements, p 121–157. *In* Morse SS (ed), *The Evolutionary Biology of Viruses*. Raven Press, New York, NY.

9. **Evgen'ev MB, Arkhipova IR.** 2005. Penelope-like elements–a new class of retroelements: distribution, function and possible evolutionary significance. *Cytogenet Genome Res* **110:**510–521.

10. **Blackburn EH, Collins K.** 2011. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol* **3:**a003558.

11. **Pardue ML, DeBaryshe PG.** 2003. Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu Rev Genet* **37:**485–511.

12. **Cordaux R, Batzer MA.** 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10:**691–703.

13. **Hancks DC, Kazazian HH Jr.** 2012. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* **22:**191–203.

14. **Konkel MK, Batzer MA.** 2010. A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. *Semin Cancer Biol* **20:**211–221.

15. **Belfort M, Curcio MJ, Lue NF.** 2011. Telomerase and retrotransposons: reverse transcriptases that shaped genomes. *Proc Natl Acad Sci U S A* **108:**20304–20310.

16. **Lampson BC, Sun J, Hsu MY, Vallejo-Ramirez J, Inouye S, Inouye M.** 1989. Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science* **243:**1033–1038.

17. **Lim D, Maas WK.** 1989. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. *Cell* **56:**891–904.

18. **Lampson BC.** 2007. Prokaryotic reverse transcriptases, p 403–420. *In* Polaina J, MacCabe AP (eds), *Industrial Enzymes: Structure, Function and Applications*. Springer, The Netherlands.

19. **Ferat JL, Michel F.** 1993. Group II self-splicing introns in bacteria. *Nature* **364:**358–361.

20. **Lambowitz AM, Zimmerly S.** 2011. Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* **3:**a003616.

21. **Lambowitz AM, Zimmerly S.** 2004. Mobile group II introns. *Annu Rev Genet* **38:**1–35.

22. **Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM.** 2002. Mobile introns: pathways and proteins, p 761–783. *In* Craig NL, Craigie R, Gellert M, Lambowitz AM (eds), *Mobile DNA II*. ASM Press, Washington DC.

23. **Liu M, Deora R, Doulatov SR, Gingery M, Eiserling FA, Preston A, Maskell DJ, Simons RW, Cotter PA, Parkhill J, Miller JF.** 2002. Reverse transcriptase-mediated tropism switching in *Bordetella* bacteriophage. *Science* **295:**2091–2094.

24. **Pyle AM, Lambowitz AM.** 2006. Group II introns: ribozymes that splice RNA and invade DNA, p 469–506. *In* Gesteland RF, Cech TR, Atkins JF (eds), *The RNA World*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY.

25. **Lehmann K, Schmidt U.** 2003. Group II introns: structural and catalytic versatility of large natural ribozymes. *Crit Rev Biochem Mol Biol* **38:**249–303.

26. **Mohr G, Perlman PS, Lambowitz AM.** 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res* **21:**4991–4997.

27. **Cui X, Matsuura M, Wang Q, Ma H, Lambowitz AM.** 2004. A group II intron-encoded maturase functions preferentially in *cis* and requires both the reverse transcriptase and X domains to promote RNA splicing. *J Mol Biol* **340:**211–231.

28. **Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz AM.** 1999. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry* **38:**9069–9083.

29. **Wank H, San Filippo J, Singh RN, Matsuura M, Lambowitz AM.** 1999. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol Cell* **4:**239–250.

30. **Matsuura M, Noah JW, Lambowitz AM.** 2001. Mechanism of maturase-promoted group II intron splicing. *EMBO J* **20:**7259–7270.

31. **Singh RN, Saldanha RJ, D'Souza LM, Lambowitz AM.** 2002. Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J Mol Biol* **318:**287–303.

32. **Michel F, Ferat JL.** 1995. Structure and activities of group II introns. *Annu Rev Biochem* **64:**435–461.

33. **Pyle AM.** 2010. The tertiary structure of group II introns: implications for biological function and evolution. *Crit Rev Biochem Mol Biol* **45:**215–232.

34. **Fedorova O, Zingler N.** 2007. Group II introns: structure, folding and splicing mechanism. *Biol Chem* **388:**665–678.

35. **Marcia M, Pyle AM.** 2012. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151:**497–507.

36. **Marcia M, Somarowthu S, Pyle AM.** 2013. Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob DNA* **4:**14.

37. **Robart AR, Chan RT, Peters JK, Rajashankar KR, Toor N.** 2014. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514:**193–197.

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

39. **Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, Lambowitz AM.** 1995. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83:**529–538.

40. **Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M.** 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94:**451–462.

41. **Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M.** 2005. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev* **19:**2477–2487.

42. **Coros CJ, Landthaler M, Piazza CL, Beauregard A, Esposito D, Perutka J, Lambowitz AM, Belfort M.** 2005. Retrotransposition strategies of the *Lactococcus lactis* Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol Microbiol* **56:**509–524.

43. **Yao J, Truong DM, Lambowitz AM.** 2013. Genetic and biochemical assays reveal a key role for replication restart proteins in group II intron retrohoming. *PLoS Genet* **9:**e1003469.

44. **Toro N, Martinez-Abarca F.** 2013. Comprehensive phylogenetic analysis of bacterial group II intron-encoded ORFs lacking the DNA endonuclease domain reveals new varieties. *PLoS ONE* **8:**e55102.

45. **Zhong J, Lambowitz AM.** 2003. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J* **22:**4555–4565.

46. **Cousineau B, Lawrence S, Smith D, Belfort M.** 2000. Retrotransposition of a bacterial group II intron. *Nature* **404:**1018–1021.

47. **Robart AR, Seo W, Zimmerly S.** 2007. Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc Natl Acad Sci U S A* **104:**6620–6625.

48. **Eskes R, Liu L, Ma H, Chao MY, Dickson L, Lambowitz AM, Perlman PS.** 2000. Multiple homing pathways used by yeast mitochondrial group II introns. *Mol Cell Biol* **20:**8432–8446.

49. **Mastroianni M, Watanabe K, White TB, Zhuang F, Vernon J, Matsuura M, Wallingford J, Lambowitz AM.** 2008. Group II intron-based gene targeting reactions in eukaryotes. *PLoS ONE* **3:**e3121.

50. **White TB, Lambowitz AM.** 2012. The retrohoming of linear group II intron RNAs in *Drosophila melanogaster* occurs by both DNA ligase 4-dependent and -independent mechanisms. *PLoS Genet* **8:**e1002534.

51. **Muñoz-Adelantado E, San Filippo J, Martínez-Abarca F, García-Rodríguez FM, Lambowitz AM, Toro N.** 2003. Mobility of the *Sinorhizobium meliloti* group II intron RmInt1 occurs by reverse splicing into DNA, but requires an unknown reverse transcriptase priming mechanism. *J Mol Biol* **327:**931–943.

52. **Dai L, Zimmerly S.** 2002. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res* **30:**1091–1102.

53. **Candales MA, Duong A, Hood KS, Li T, Neufeld RA, Sun R, McNeil BA, Wu L, Jarding AM, Zimmerly S.** 2012. Database for bacterial group II introns. *Nucleic Acids Res* **40:**D187–D190.

54. **Simon DM, Clarke NA, McNeil BA, Johnson I, Pantuso D, Dai L, Chai D, Zimmerly S.** 2008. Group II introns in Eubacteria and Archaea: ORF-less introns and new varieties. *RNA* **14:**1704–1713.

55. **Toro N, Martinez-Rodriguez L, Martinez-Abarca F.** 2014. Insights into the history of a bacterial group II intron remnant from the genomes of the nitrogen-fixing symbionts *Sinorhizobium meliloti* and *Sinorhizobium medicae*. *Heredity* **113:**306–315.

56. **Michel F, Umesono K, Ozeki H.** 1989. Comparative and functional anatomy of group II catalytic introns–a review. *Gene* **82:**5–30.

57. **Toro N, Jiménez-Zurdo JI, García-Rodríguez FM.** 2007. Bacterial group II introns: not just splicing. *FEMS Microbiol Rev* **31:**342–358.

58. **Toro N, Martinez-Abarca F, Fernandez-Lopez M, Munoz-Adelantado E.** 2003. Diversity of group II introns in the genome of *Sinorhizobium meliloti* strain 1021: splicing and mobility of RmInt1. *Mol Genet Genomics* **268:**628–636.

59. **Dai L, Zimmerly S.** 2002. The dispersal of five group II introns among natural populations of *Escherichia coli*. *RNA* **8:**1294–1307.

60. **Ueda K, Yamashita A, Ishikawa J, Shimada M, Watsuji TO, Morimura K, Ikeda H, Hattori M, Beppu T.** 2004. Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalism. *Nucleic Acids Res* **32:**4937–4944.

61. **McNeil BA, Simon DM, Zimmerly S.** 2014. Alternative splicing of a group II intron in a surface layer protein gene in *Clostridium tetani*. *Nucleic Acids Res* **42:**1959–1969.

62. **Medhekar B, Miller JF.** 2007. Diversity-generating retroelements. *Curr Opin Microbiol* **10:**388–395.

63. **Doulatov S, Hodes A, Dai L, Mandhana N, Liu M, Deora R, Simons RW, Zimmerly S, Miller JF.** 2004. Tropism switching in *Bordetella* bacteriophage defines a family of diversity-generating retroelements. *Nature* **431:**476–481.

64. **Alayyoubi M, Guo H, Dey S, Golnazarian T, Brooks GA, Rong A, Miller JF, Ghosh P.** 2013. Structure of the essential diversity-generating retroelement protein bAvd and its functionally important interaction with reverse transcriptase. *Structure* **21:**266–276.

65. **Guo H, Tse LV, Barbalat R, Sivaamnuaiphorn S, Xu M, Doulatov S, Miller JF.** 2008. Diversity-generating retroelement homing regenerates target sequences for repeated rounds of codon rewriting and protein diversification. *Mol Cell* **31:**813–823.

66. **Guo H, Tse LV, Nieh AW, Czornyj E, Williams S, Oukil S, Liu VB, Miller JF.** 2011. Target site recognition by a diversity-generating retroelement. *PLoS Genet* **7:**e1002414.

67. **McMahon SA, Miller JL, Lawton JA, Kerkow DE, Hodes A, Marti-Renom MA, Doulatov S, Narayanan E, Sali A, Miller JF, Ghosh P.** 2005. The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat Struct Mol Biol* **12:**886–892.

68. **Miller JL, Le Coq J, Hodes A, Barbalat R, Miller JF, Ghosh P.** 2008. Selective ligand recognition by a diversity-generating retroelement variable protein. *PLoS Biol* **6:**e131.

69. **Cummings CA, Bootsma HJ, Relman DA, Miller JF.** 2006. Species- and strain-specific control of a complex, flexible regulon by *Bordetella* BvgAS. *J Bacteriol* **188:**1775–1785.

70. **Arambula D, Wong W, Medhekar BA, Guo H, Gingery M, Czornyj E, Liu M, Dey S, Ghosh P, Miller JF.** 2013. Surface display of a massively variable lipoprotein by a *Legionella* diversity-generating retroelement. *Proc Natl Acad Sci U S A* **110:**8212–8217.

71. **Schillinger T, Lisfi M, Chi J, Cullum J, Zingler N.** 2012. Analysis of a comprehensive dataset of diversity generating retroelements generated by the program DiGReF. *BMC Genomics* **13:**430.

72. **Lampson BC, Inouye M, Inouye S.** 2005. Retrons, msDNA, and the bacterial genome. *Cytogenet Genome Res* **110:**491–499.

73. **Lampson B, Inouye M, Inouye S.** 2001. The msDNAs of bacteria. *Prog Nucleic Acid Res Mol Biol* **67:**65–91.

74. **Inouye S, Inouye M.** 1993. The retron: a bacterial retroelement required for the synthesis of msDNA. *Curr Opin Genet Dev* **3:**713–718.

75. **Inouye M, Inouye S.** 1991. msDNA and bacterial reverse transcriptase. *Annu Rev Microbiol* **45:**163–186.

76. **Inouye M, Ke H, Yashio A, Yamanaka K, Nariya H, Shimamoto T, Inouye S.** 2004. Complex formation between a putative 66-residue thumb domain of bacterial reverse transcriptase RT-Ec86 and the primer recognition RNA. *J Biol Chem* **279:**50735–50742.

77. **Inouye S, Hsu MY, Xu A, Inouye M.** 1999. Highly specific recognition of primer RNA structures for 2′-OH priming reaction by bacterial reverse transcriptases. *J Biol Chem* **274:**31236–31244.

78. **Inouye S, Sunshine MG, Six EW, Inouye M.** 1991. Retronphage phi R73: an *E. coli* phage that contains a retroelement and integrates into a tRNA gene. *Science* **252:**969–971.

79. **Herzer PJ, Inouye S, Inouye M.** 1992. Retron-Ec107 is inserted into the *Escherichia coli* genome by replacing a palindromic 34bp intergenic sequence. *Mol Microbiol* **6:**345–354.

80. **Shimamoto T, Ahmed AM, Shimamoto T.** 2013. A novel retron of *Vibrio parahaemolyticus* is closely related to retron-Vc95 of *Vibrio cholerae*. *J Microbiol* **51:**323–328.

81. **Lampson BC, Inouye M, Inouye S.** 1991. Survey of multicopy single-stranded DNAs and reverse transcriptase genes among natural isolates of *Myxococcusxanthus*. *J Bacteriol* **173:**5363–5370.

82. **Rice SA, Lampson BC.** 1995. Phylogenetic comparison of retron elements among the myxobacteria: evidence for vertical inheritance. *J Bacteriol* **177:**37–45.

83. **Herzer PJ, Inouye S, Inouye M, Whittam TS.** 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol* **172:**6175–6181.

84. **Maas WK, Wang C, Lima T, Hach A, Lim D.** 1996. Multicopy single-stranded DNA of *Escherichia coli* enhances mutation and recombination frequencies by titrating MutS protein. *Mol Microbiol* **19:**505–509.

85. **Maas WK, Wang C, Lima T, Zubay G, Lim D.** 1994. Multicopy single-stranded DNAs with mismatched base pairs are mutagenic in *Escherichia coli*. *Mol Microbiol* **14:**437–441.

86. **Yamanaka K, Shimamoto T, Inouye S, Inouye M.** 2002. Retrons, p 784–795. *In* Craig NL, Craigie R, Gellert M, Lambowitz AM (ed), *Mobile DNA II*. ASM Press, Washington DC.

87. **Kojima KK, Kanehisa M.** 2008. Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. *Mol Biol Evol* **25:**1395–1404.

88. **Simon DM, Zimmerly S.** 2008. A diversity of uncharacterized reverse transcriptases in bacteria. *Nucleic Acids Res* **36:**7219–7229.

89. **Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH.** 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* **41:**D348–D352.

90. **Castro C, Smidansky ED, Arnold JJ, Maksimchuk KR, Moustafa I, Uchida A, Gotte M, Konigsberg W, Cameron CE.** 2009. Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol* **16:**212–218.

91. **Steitz TA.** 1999. DNA polymerases: structural diversity and common mechanisms. *J Biol Chem* **274:**17395–17398.

92. **Mönttinen HA, Ravantti JJ, Stuart DI, Poranen MM.** 2014. Automated structural comparisons clarify the phylogeny of the right-hand-shaped polymerases. *Mol Biol Evol* **31:**2741–2752.

93. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M.** 2014. Pfam: the protein families database. *Nucleic Acids Res* **42:** D222–D230.

94. **Emond E, Holler BJ, Boucher I, Vandenbergh PA, Vedamuthu ER, Kondo JK, Moineau S.** 1997. Phenotypic and genetic characterization of the bacteriophage abortive infection mechanism AbiK from *Lactococcus lactis*. *Appl Environ Microbiol* **63:**1274–1283.

95. **Fortier LC, Bouchard JD, Moineau S.** 2005. Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK. *J Bacteriol* **187:**3721–3730.

96. **Wang C, Villion M, Semper C, Coros C, Moineau S, Zimmerly S.** 2011. A reverse transcriptase-related protein mediates phage resistance and polymerizes untemplated DNA *in vitro*. *Nucleic Acids Res* **39:**7620–7629.

97. **Wang GH, Seeger C.** 1992. The reverse transcriptase of hepatitis B virus acts as a protein primer for viral DNA synthesis. *Cell* **71:**663–670.

98. **Bouchard JD, Moineau S.** 2004. Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. *J Bacteriol* **186:**3649–3652.

99. **Lopes A, Amarir-Bouhram J, Faure G, Petit MA, Guerois R.** 2010. Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res* **38:**3952–3962.

100. **Ploquin M, Bransi A, Paquet ER, Stasiak AZ, Stasiak A, Yu X, Cieslinska AM, Egelman EH, Moineau S, Masson JY.** 2008. Functional and structural basis for a bacteriophage homolog of human RAD52. *Curr Biol* **18:**1142–1146.

101. **Scaltriti E, Moineau S, Launay H, Masson JY, Rivetti C, Ramoni R, Campanacci V, Tegoni M, Cambillau C.** 2010. Deciphering the function of lactococcal phage ul36 Sak domains. *J Struct Biol* **170:**462–469.

102. **Scaltriti E, Launay H, Genois MM, Bron P, Rivetti C, Grolli S, Ploquin M, Campanacci V, Tegoni M, Cambillau C, Moineau S, Masson JY.** 2011. Lactococcal phage p2 ORF35-Sak3 is an ATPase involved in DNA recombination and AbiK mechanism. *Mol Microbiol* **80:**102–116.

103. **Hill C, Miller LA, Klaenhammer TR.** 1990. Nucleotide sequence and distribution of the pTR2030 resistance determinant (hsp) which aborts bacteriophage infection in lactococci. *Appl Environ Microbiol* **56:**2255–2258.

104. **Tangney M, Fitzgerald GF.** 2002. Effectiveness of the lactococcal abortive infection systems AbiA, AbiE, AbiF and AbiG against P335 type phages. *FEMS Microbiol Lett* **210:**67–72.

105. **Dinsmore PK, Klaenhammer TR.** 1997. Molecular characterization of a genomic region in a *Lactococcus* bacteriophage that is involved in its sensitivity to the phage defense mechanism AbiA. *J. Bacteriol* **179:**2949–2957.

106. **Odegrip R, Nilsson AS, Haggård-Ljungquist E.** 2006. Identification of a gene encoding a functional reverse transcriptase within a highly variable locus in the P2-like coliphages. *J Bacteriol* **188:**1643–1647.

107. **Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJ, Yoo HS, Zhang C, Zhang Y, Sobral BW.** 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42:**D581–D591.

108. **Gladyshev EA, Arkhipova IR.** 2011. A widespread class of reverse transcriptase-related cellular genes. *Proc Natl Acad Sci U S A* **108:**20311–20316.

109. **Barrangou R, Marraffini LA.** 2014. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol Cell* **54:**234–244.

110. **Chylinski K, Makarova KS, Charpentier E, Koonin EV.** 2014. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* **42:**6091–6105.

111. **van der Oost J, Westra ER, Jackson RN, Wiedenheft B.** 2014. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **12:**479–492.

112. Nuñez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. 2014. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* **21:**528–534.

113. Curcio MJ, Belfort M. 2007. The beginning of the end: links between ancient retroelements and modern telomerases. *Proc Natl Acad Sci U S A* **104:**9107–9108.

114. Inouye S, Inouye M. 1995. Structure, function, and evolution of bacterial reverse transcriptase. *Virus Genes* **11:**81–94.

115. Nakamura TM, Cech TR. 1998. Reversing time: origin of telomerase. *Cell* **92:**587–590.

116. Eickbush TH, Malik HS. 2002. Origins and evolution of retrotransposons, p 1111–1144. *In* Craig NL, Craigie R, Gellert M, Lambowitz AM (ed), *Mobile DNA II*. ASM Press, Washington DC.

117. Eickbush TH. 1997. Telomerase and retrotransposons: which came first? *Science* **277:**911–912.

118. Iyer LM, Koonin EV, Aravind L. 2003. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol* **3:**1.

119. Darnell JE, Doolittle WF. 1986. Speculations on the early course of evolution. *Proc Natl Acad Sci U S A* **83:**1271–1275.

120. Cech TR, Golden BL. 1999. Building a catalytic active site using only RNA, p 321–349. *In* Gesteland RF, Cech TR, Atkins JF (ed), *The RNA World*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

121. Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP. 2001. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292:**1319–1325.

122. Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* **16:**793–805.

123. Luan DD, Korman MH, Jakubczak JL, Eickbush 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.

124. Martin W, Koonin EV. 2006. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440:**41–45.

125. Eickbush TH. 1999. Mobile introns: retrohoming by complete reverse splicing. *Curr Biol* **9:**R11–R14.