

SURVEY AND SUMMARY

Retrons and their applications in genome engineering

Anna J. Simon¹*, Andrew D. Ellington and Ilya J. Finkelstein¹*

Center for Systems and Synthetic Biology and Department of Molecular Biosciences, University of Texas at Austin, Austin, Texas 78712, USA

Received April 23, 2019; Revised September 19, 2019; Editorial Decision September 23, 2019; Accepted October 02, 2019

ABSTRACT

Precision genome editing technologies have transformed modern biology. These technologies have arisen from the redirection of natural biological machinery, such as bacteriophage lambda proteins for recombineering and CRISPR nucleases for eliciting site-specific double-strand breaks. Less well-known is a widely distributed class of bacterial retroelements, retrons, that employ specialized reverse transcriptases to produce noncoding intracellular DNAs. Retrons' natural function and mechanism of genetic transmission have remained enigmatic. However, recent studies have harnessed their ability to produce DNA *in situ* for genome editing and evolution. This review describes retron biology and function in both natural and synthetic contexts. We also highlight areas that require further study to advance retron-based precision genome editing platforms.

INTRODUCTION

The ability to precisely edit genomic loci has been a long-standing goal of biotechnology. Targeted genome editing is being exploited to probe genotype-phenotype relationships, repair disease-causing alleles, develop designer organisms with modified genomes, and to biologically record intra- and extracellular conditions. Classically, strain improvement has been achieved either via non-specific mutagenesis of genomes (1,2) or via targeted modifications of specific genes expressed on plasmids (3,4). Recombineering (5) and CRISPR-based genome editing (6,7) have dramatically increased the ease and throughput of targeting individual genomic loci. Multiplexing and iterating recombineering- and CRISPR-based editing technologies have enabled high-throughput exploration of genotype-phenotype relationships and continuous evolution of synthetic genomes with novel properties, for example in Multi-

plex Automated Genome Engineering ('MAGE') (8), Yeast Oligo-mediated Genome Engineering ('YOGEE') (9), and iterative CRISPR EnAbleD Trackable genome Engineering ('iCREATE') (10).

Retrons are a potentially useful tool for targeted genomic engineering because they produce intracellular DNAs via reverse transcription of noncoding structural RNAs (11,12). While the natural biology of these widespread retroelements is largely unknown, retrons have recently been repurposed as *in situ* sources of donor DNA for both recombineering- and CRISPR-based genome editing applications (13–15). This production of intracellular DNA has so far enabled recording of cellular states (13), high-throughput genetic editing (14), and targeted genome mutation (15). Here, we review retron biology and retron-enabled genome engineering applications. We conclude with a discussion of future prospects and challenges in retron-based genome editing.

RETRON BIOLOGY

Retron discovery and biochemistry

Retrons were initially discovered in 1984 from the presence of short satellite RNA–DNA molecules in bacterial DNA preparations (16). Gel electrophoresis of phenol-extracted chromosomal DNA from *Myxobacteria xanthus* and *Stigmatella aurantica* indicated a secondary satellite band with a mobility of ~120–190 base pairs (bp) (16). Subsequent studies revealed that these bands, termed multicopy single-stranded DNAs (msDNAs), are comprised of one strand of structured RNA, the 'msr,' connected to one strand of DNA, the 'msd.' The msr and msd molecules are joined by a 2'-5' phosphodiester bond between a priming guanosine within a conserved AGC sequence in the msr and the phosphate of the 5' end of the msd that covalently links the RNA and DNA strands into a single branched molecule (Figure 1). The msr and msd are encoded in a compact, contiguous transcriptional cassette that also includes a specialized reverse transcriptase (RT); we refer to this cassette as a whole

*To whom correspondence should be addressed. Tel: +1 512 475 6172; Email: ilya@finkelsteinlab.org
Correspondence may also be addressed to Anna J. Simon. Email: annajsimon@gmail.com

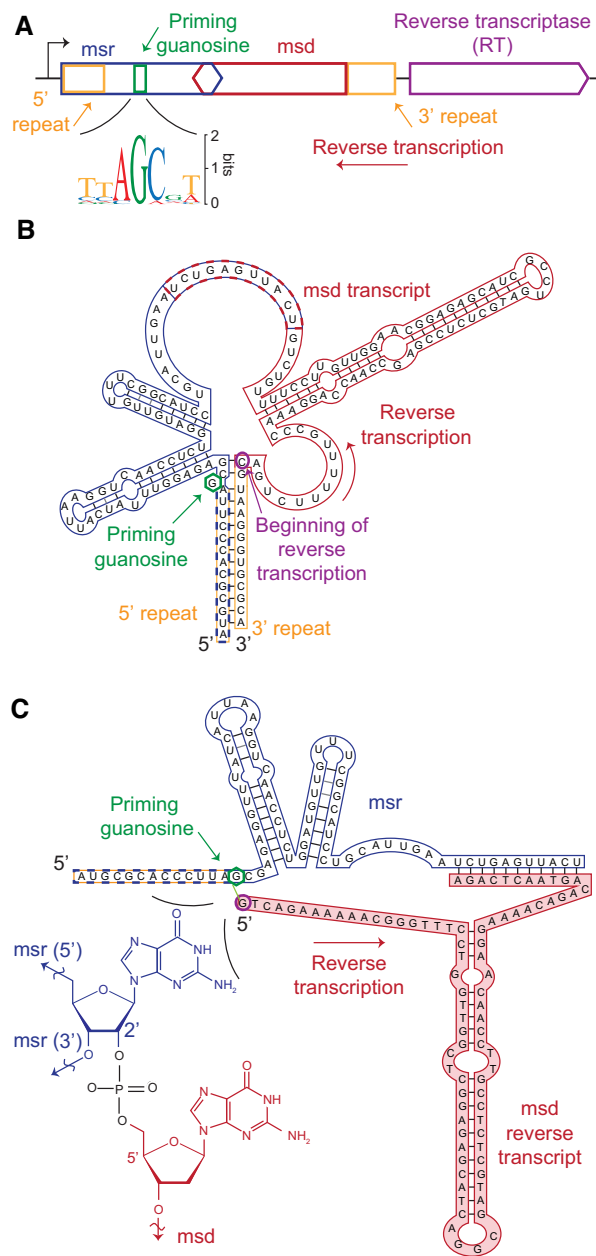


Figure 1. Retron structure and organization. (A) Retrons are encoded as a single polycistronic transcriptional cassette containing a promoter, an msr (blue) including a conserved priming guanosine residue (green), an msd (red), self-complementary regions (yellow), and a reverse transcriptase (purple, not to scale) (12). The black arrow shows the direction of the coding strand for the retron cassette; the red arrow shows the direction of reverse transcription. Inset: sequence logo of nucleotides adjacent to the priming guanosine of 15 msr-msd sequences. Retron-Sen1 (Se72) was excluded because its msr lacks an apparent priming guanosine (46). (B) Structure of a representative msr-msd transcript encoded by Retron-Eco1 (Ec86) (23). The transcript encoding the reverse transcriptase is cleaved from that of the msr-msd before folding (12). The cytosine at which reverse transcription begins is circled in purple. (C) Structure of the mature Retron-Eco1 (Ec86) msDNA. Inset: the msr and the 5' end of the reverse-transcribed msd are covalently bonded to the priming guanosine via a 2'-5' linkage.

as the retron. The discovery of retron RTs was particularly notable because they were the first discovered bacterially-encoded RTs (17). *In vitro* studies demonstrated that *Escherichia coli* retron RTs and their paired msrs were both necessary and sufficient to produce msDNA with the characteristic 2'-5' linkage (18,19). To date, 38 distinct retrons have been identified experimentally by their production of msDNA; 16 of these, which we subsequently refer to as the 'experimentally validated retrons', contain fully annotated and experimentally-validated msr-msd-RT cassettes. Table 1 summarizes all experimentally-characterized retrons reported to date, along with their host organisms. Bioinformatic approaches have also identified hundreds of putative retrons that have not been characterized experimentally (20–22).

The rapid discovery of hundreds of putative retrons has prompted us to develop a systematic naming convention for these retroelements. Currently, retrons are named by the first letters of their genus name, species name, and the length of reverse-transcribed DNA in their native msDNA sequence. This convention was initially developed by Inouye *et al.* in 1989, who termed a retron isolated from *E. coli* that contained 86 bases of reverse-transcribed DNA 'Ec86' (23). However, this nomenclature is not easy to adapt for large-scale, systematic retron discovery for three reasons: (i) the length of the msd of most experimentally validated and putative retrons is unknown; (ii) two letters are insufficient to uniquely identify the genus and species of many retron-encoding bacteria; and (iii) the convention does not distinguish the entire retron from the RT. Therefore, we propose a more general naming convention that is adapted from the restriction enzyme literature (Table 1) (24).

- I. Retrons are named as the first letter of their genus, first two letters of their species, and an Arabic numeral corresponding to the order with which they were discovered relative to other retrons from the same species or a different species with the same first genus and first two species letters, respectively.
- II. The prefix 'Retron-' is added to signify that this object is a retron, and the prefixes 'RT-', 'msr', 'msd' and 'msDNA' are added when referring to the reverse transcriptase enzyme, msr, msd and msDNA, respectively.

Under the new convention, 'Ec86' is re-named 'Retron-Eco1' and the corresponding RT is 'RT-Eco1.' For clarity and continuity, the rest of this review will use the new convention and will also include the historical names in parenthesis (e.g. Retron-Eco1 (Ec86)).

Msr, msd and msDNA structure and production

In all experimentally validated retrons, the msr and msd are encoded within a single transcriptional cassette, with the msr to the 5' of the msd (Figure 1A). The retron msr-msd transcriptional cassettes are divergent at the sequence level but share substantial structural homology (Supplementary Figure S1). The structural elements are likely necessary for function. Prior to reverse transcription, the msr and msd form a single highly structured transcript (Figure 1B). The msr structure contains 1–3 stable hairpins with ~7–10 bp-

Table 1. Summary of all retrons whose production of msDNA has been experimentally validated. The new systematic nomenclature is described in the main text. In parenthesis, we also include the names initially developed in (23). The 16 shaded entries, referred to as the ‘experimentally validated retrons’, contain fully annotated and experimentally-validated msr-msd-RT cassettes listed in the National Center for Biotechnology Information (NCBI) databank (accessed August 2019). The 22 remaining entries contain elements that have been experimentally demonstrated to produce msDNA but do not correspond to fully annotated or validated entries in the NCBI databank

	Name	Species	Bacterial taxon	Reference
1	Retron-Eco1 (Ec86)	<i>Escherichia coli</i>	<i>Gammaproteobacteria</i>	Lim & Maas 1989 (86)
2	Retron-Eco2 (Ec67)	<i>Escherichia coli</i>	<i>Gammaproteobacteria</i>	Lampson et al. 1989 (23)
3	Retron-Eco3 (Ec73)	<i>Escherichia coli</i>	<i>Gammaproteobacteria</i>	Sun et al. 1991 (58)
4	Retron-Eco4 (Ec83)	<i>Escherichia coli</i>	<i>Gammaproteobacteria</i>	Lim 1992 (30)
5	Retron-Eco5 (Ec107)	<i>Escherichia coli</i>	<i>Gammaproteobacteria</i>	Herzer et al. 1992 (87)
6	Retron-Eco6 (Ec48)	<i>Escherichia coli</i>	<i>Gammaproteobacteria</i>	Mao et al. 1996 (88)
7	Retron-Eco7 (Ec78)	<i>Escherichia coli</i>	<i>Gammaproteobacteria</i>	Lima & Lim, 1997 (34)
8	Retron-Mxa1 (Mx162)	<i>Myxococcus xanthus</i>	<i>Deltaproteobacteria</i>	Yee et al. 1984 (16)
9	Retron-Mxa2 (Mx65)	<i>Myxococcus xanthus</i>	<i>Deltaproteobacteria</i>	Inouye et al. 1990 (31)
10	Retron-Sau1 (Sa163)	<i>Stigmatella aurantica</i>	<i>Deltaproteobacteria</i>	Yee et al. 1984 (16)
11	Retron-Nex1 (Ne160)	<i>Nannocystis exedens</i>	<i>Deltaproteobacteria</i>	Lampson & Rice 1997 (60)
12	Retron-Nex2 (Ne144)	<i>Nannocystis exedens</i>	<i>Deltaproteobacteria</i>	Lampson et al. 2002 (89)
13	Retron-Sen1 (Se72)	<i>Salmonella enterica</i> Enteridis	<i>Gammaproteobacteria</i>	Rychlik et al. 2001 (44)
14	Retron-Sen2 (St85)	<i>Salmonella enterica</i> Typhimurium	<i>Gammaproteobacteria</i>	Ahmed & Shimamoto 2003 (90)
15	Retron-Vch1 (Vc95)	<i>Vibrio cholerae</i>	<i>Gammaproteobacteria</i>	Shimamoto et al. 1999 (66)
16	Retron-Vch2 (Vc81)	<i>Vibrio cholerae</i>	<i>Gammaproteobacteria</i>	Inouye et al. 2011 (64)
17	Retron-Vch3 (Vc137)	<i>Vibrio cholerae</i>	<i>Gammaproteobacteria</i>	Inouye et al. 2011 (64)
18	Retron-Vpa1 (Vp96)	<i>Vibrio parahaemolyticus</i>	<i>Gammaproteobacteria</i>	Ahmed & Shimamoto 2003 (90)
19	Retron-Kpn1	<i>Klebsiella pneumoniae</i>	<i>Gammaproteobacteria</i>	Rice et al. 1993 (91)
20	Retron-Pmi1	<i>Proteus mirabilis</i>	<i>Gammaproteobacteria</i>	Rice et al. 1993 (91)
21	Retron-Rxx1	<i>Rhizobium</i> sp.	<i>Alphaproteobacteria</i>	Rice et al. 1993 (91)
22	Retron-Bxx1	<i>Bradyrhizobium</i> sp.	<i>Alphaproteobacteria</i>	Rice et al. 1993 (91)
23	Retron-Fel1	<i>Flexibacter elegans</i>	<i>Bacteriodes</i>	Dhundale et al. 1985 (92)
24	Retron-Cco1	<i>Coralococcus coralloides</i>	<i>Deltaproteobacteria</i>	Dhundale et al. 1985 (92)
25	Retron-Adi1	<i>Angiococcus disciformis</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
26	Retron-Age1	<i>Archangium gephyra</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
27	Retron-Cfe1	<i>Cystobacter ferrugineus</i>	<i>Deltaproteobacteria</i>	Dhundale et al. 1985 (92)
28	Retron-Cfu1	<i>Cystobacter fuscus</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
29	Retron-Cvi1	<i>Cystobacter violaceus</i>	<i>Deltaproteobacteria</i>	Dhundale et al. 1985 (92)
30	Retron-Mli1	<i>Melittangium lichenicola</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
31	Retron-Mco1	<i>Myxococcus coralloides</i>	<i>Deltaproteobacteria</i>	Dhundale et al. 1985 (92)
32	Retron-Mfu1	<i>Myxococcus fulvus</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
33	Retron-Mma1	<i>Myxococcus macrosporus</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
34	Retron-Mst1	<i>Myxococcus stipitatus</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
35	Retron-Mvi1	<i>Myxococcus virescens</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
36	Retron-Cap1	<i>Chondromyces apiculatus</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
37	Retron-Cpe1	<i>Chondromyces pedicatus</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)
38	Retron-Sce1	<i>Sorangium cellulosum</i>	<i>Deltaproteobacteria</i>	Rice & Lampson 1995 (50)

long stems and ~3–10 nt loops (12,25). The msd-encoding segment of the transcript is also highly structured, folding into a single hairpin with a long (~10–50 bp) stem that generally contains mismatched or wobble base pairs. In addition to this internal structure of the msr and msd, the ~8–30 bases at the 5' end of the msr and the 3' end of the msd transcript are complementary and hybridize (Figure 1B). The internal (e.g. 3') and of the msr portion of this stem contains or ends immediately 5' to the conserved guanosine residue that primes reverse transcription of the satellite DNA. Most commonly, this guanosine is found within a conserved TAGC sequence, though occasionally this sequence is CAGA (as in Retron-Eco2 (Ec67)), CAGC (as in Retron-Sau1 (Sa163) and Retron-Mxa1 (Mx162)), GAGC (as in Retron-Mxa2 (Mx65)) or ACGC (as in Retron-Vch2 (Vc81)) (Figure 1A, inset).

The RT initiates reverse transcription using the base-paired 5' and 3' stem of the msr-msd as a primer (12). Early experiments demonstrated that several *E. coli* retron RTs and their paired msrs are both necessary and sufficient to produce msDNA with the characteristic 2'-5' linkage (18,19), and that the 2'-5' linkage between the priming guanosine in the msr and first base of the msd DNA is catalyzed directly by the RT (26). The RT extends the reverse transcript until a defined position at which reverse transcription stops *in vitro* and *in vivo* via an unknown mechanism. The observation that the RT terminates at similar sites *in vitro* and in cells strongly suggests that RNA structure directs RT termination. Concurrently with reverse transcription, cellular RNase H activity degrades the template of the msd, excluding ~5–10 RNA bases at its 5' end. This segment of the RNA remains hybridized to the complementary reverse transcript and is thus be considered to be part of both the msr and msd (27) (Figure 1C). The mechanism limiting RNase H activity at terminal msd residues also remains unexplored.

Retron msDNA is further nucleolytically processed after reverse transcription (12,17,28,29). Often, some or all of the msr upstream of the conserved priming guanosine is degraded. For example, the Retron-Mxa1 (Mx162) msDNA is initially transcribed as a 241 nucleotide (nt) transcript. In the mature msDNA, 16 residues are cleaved from the 5' end, leaving a 71-nt msr and a 162-nt msd msDNA (the msr and msd overlap at 8 bases) (17). Other retrons undergo much more dramatic processing that results in a mature msDNA that is primarily reverse transcribed DNA. For example, msDNA-Eco4 (Ec83) is endonucleolytically cleaved between the fourth and fifth nucleotides from the priming guanosine residue, removing that linkage to the msr (12,28). This cleavage likely results in the dissociation of the base pairs at the 3' ends of the msd and msr, since 90% of experimentally-observed msDNA-Eco4 (Ec83) is not joined to RNA (30). While the enzyme responsible for this cleavage is not known, it may arise from a putative ATP-binding protein encoded just downstream of RT-Eco4 (Ec83) in the native cassette.

Reverse transcription requires both an msr-msd sequence and its cognate RT. Expressing the RT of one retron and the msr of another does not produce msDNA (31–33), with some exceptions for highly homologous retrons (34,35). Mutational and Systematic Evolution of Ligands by Ex-

ponential Enrichment (SELEX) -based studies of Retron-Eco1 (Ec86) and Retron-Eco3 (Ec73) indicated that both msr sequence and secondary structure are important for RT recognition. While RT-Eco1 (Ec86) can bind moderately truncated or modified variants of its cognate msr, removing or destabilizing a core section containing an 8 bp stem and a three base loop abrogates binding. Similarly, RT-Eco3 (Ec73) binds a specific 9 bp stem, three base loop section of its cognate msr but fails to bind other msr truncations (32,36).

Reverse transcriptase structure and function

Retron RTs form a diverse class within the broader family of bacterial RTs (Figure 2, Supplementary Figure S2). Early estimates suggested that retrons encode 11–14% of all bacterial RTs, second only to Group II intron maturases (20,21). More recently, a metagenomic survey suggested that diversity-generating retroelements (DGRs) may be more common than retrons in uncultivated bacteria (37). Nonetheless, retron RTs are quite common in diverse bacteria (see next section). Experimentally validated retron RTs share just ~14% average amino acids (aa) homology with the *Roseburia intestinalis* Group II intron maturase RT domain, ~12% with the *Bordetella* bacteriophage DGR *bRT* RT, and ~25% homology with each other (Supplementary Table S1).

Phylogenetic analysis indicates several distinct groups of retron RT sequences, which together form a group distinct from RTs associated with both Group II intron maturases and DGRs (Figure 2A). To identify the relationship between retron RTs and other bacterial RTs we first aligned the sequences of all 16 fully annotated retron RTs, eight previously putatively identified retron RTs from taxonomic groups lacking experimentally validated retrons, three Group II intron maturase RTs, and the *Bordetella* bacteriophage DGR *bRT* using NCBI's Constraint-based Multiple Alignment Tool (COBALT) (38). We then built a phylogenetic tree from these alignments using the Geneious Tree Builder tool with the Jukes-Cantor genetic distance model and the neighbor-joining tree build method. In the resulting tree, retron RTs branch from the Group II intron maturase and DGR RTs. Notably, the putative retron RT sequences, which originated from genomes of bacteria in the *Bacteroidetes*, *Betaproteobacteria*, *Cyanobacteria*, *Fusobacteria*, and *Firmicutes*, and the archaeon *Methanosarcina acetivorans* (39), and the experimentally validated retron RT sequences, which originated only from *Gammaproteobacteria* and *Deltaproteobacteria*, are clearly within this group separate from other bacterial RTs. There are several clear groups within the retron RTs. One group contains the myxococcal RT-Mxa2 (Mx65), RT-Mxa1 (Mx162), RT-Nex2 (Ne144) and RT-Sau1 (Sa163) and putative cyanobacterial retron RTs from *Nostoc punctiforme* and *Trichodesmium erythraeum*. Another group contains RT-Vch1 (Vc95), RT-Eco4 (Ec83), RT-Vpa3 (Vp96), and RT-Eco7 (Ec78). The remaining RTs in this analysis do not fall into closely related groups.

Examination of the aligned retron RT sequences highlights the seven conserved RT motifs and two retron-specific inter-motif regions (Figure 2B, Supplementary Figure S2)

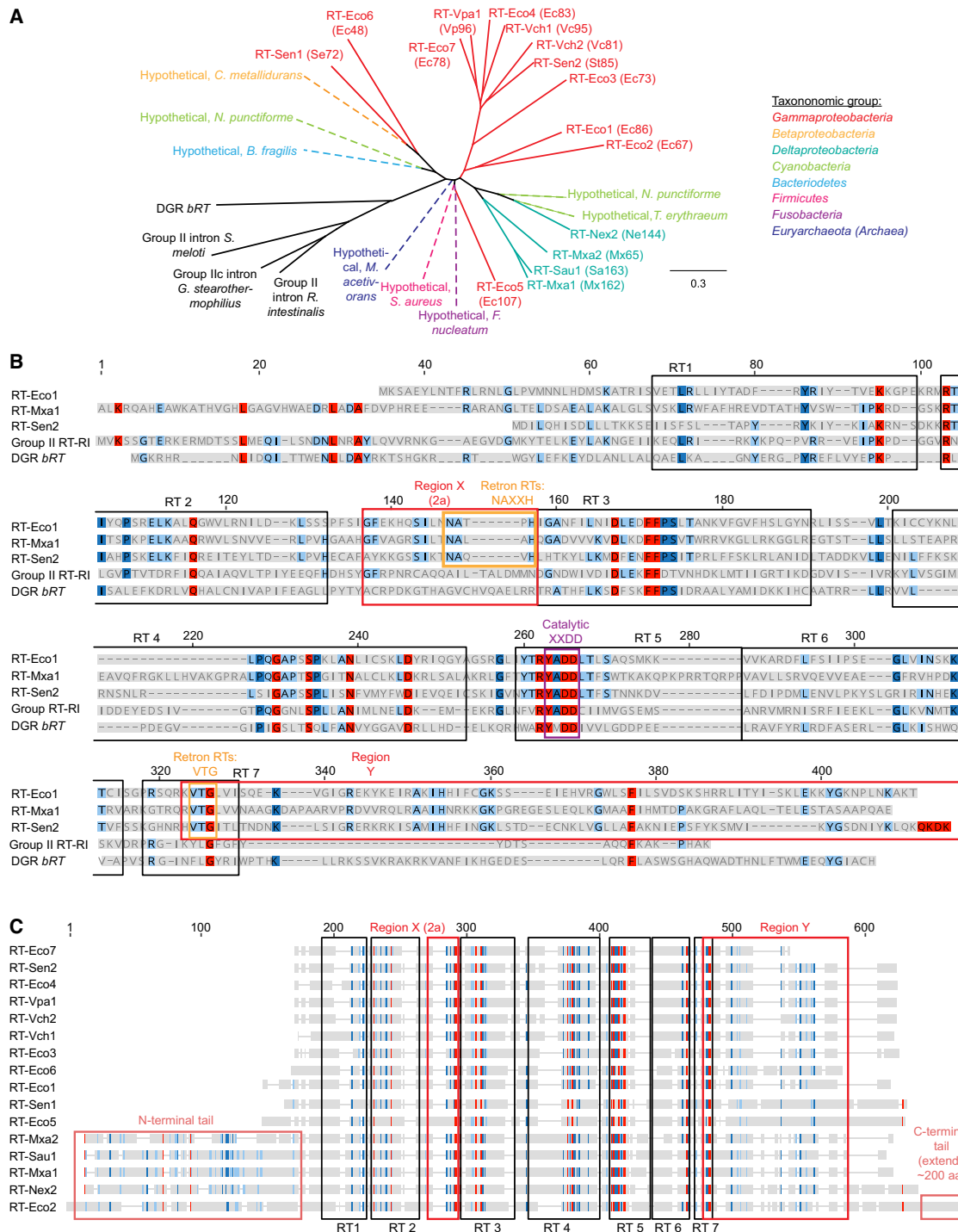


Figure 2. Retrons form a distinct class within the larger group of bacterial RTs. (A) Phylogenetic tree showing relationships between Group II intron maturase reverse transcriptases, *bRT* RT from the *Bordetella* bacteriophage Diversity-Generating Retroelement (DGR), and retron RTs. Group II intron maturase and DGR RTs are colored black; retron RTs are colored by phylum. Putative retons (dashed) identified from *Bacteroidetes*, *Betaproteobacteria*, *Cyanobacteria*, *Fusobacteria*, *Firmicutes*, and *Euryarchaeota* genomes were employed as these groups lack experimentally validated retons (39). (B) Sequence alignments of RT-Eco1 (Ec86), RT-Mxa1 (Mx162), RT-Sen2 (Se72), *Roseburia intestinalis* Group II intron maturase reverse transcriptase domain ('Group II RT-RI'), and the *Bordetella* bacteriophage DGR reverse transcriptase *bRT*. Each retron RT sequence aligns well to the seven universal reverse transcriptase regions (boxed in black) and includes several interdomain regions, including the retron-specific regions X and Y (boxed in red) (12,40). Conserved, retron-specific residues in regions X and Y are boxed in orange; the catalytic XXDD core common to all known reverse transcriptases is boxed in purple. (C) Distant-level view of the alignment of RT sequences from all 16 fully characterized retons. Sequences were aligned with the NCBI online tool COBALT (38); the alignment was then employed to construct a phylogenetic tree using the neighbor-joining tree build method in Geneious software under a Jukes-Cantor genetic distance model. Residues shaded in red, dark blue, light blue, and grey correspond to residues that are 100%, 80–100%, 60–80% and <60% homologous across the aligned sequences. RT-Mxa1 (Mx162), RT-Mxa2 (Mx65), RT-Sau1 (Sa163), and RT-Nex2 (Ne144) contain an extended N-terminus; RT-Eco2 (Ec67) contains both N- and C-terminal extensions (boxed in light red).

(40). Most retron RTs are ~300–400 aa long and lack bioinformatically identifiable endonuclease, maturase, or RNase domains (Figure 2C) (12,23). Several retron RTs also encode additional N- and C-terminal extensions of unknown function. For example, myxococcal retron RTs (RT-Mxa1 (Mx162), RT-Mxa2 (Mx65), RT-Sau1 (Sa163) and RT-Nex2 (Ne144)) contain an extended N-terminal region. Similarly, RT-Eco2 (Ec67) is 586 residues long and contains both N- and C-terminal extensions that lack homology to non-retron protein domains.

The conserved RT motifs include the ‘palm’ and ‘fingers’ domains and the conserved double aspartic acid-containing catalytic core. Retron RTs encode the YADD catalytic tetrad, as seen in Group II intron maturase RTs and other bacterial RTs. As with other aspects of the retron RT sequence, several variants encode altered catalytic cores (e.g., YCDD in RT-Eco5 (Ec107), Supplementary Figure S2). Of the two conserved retron-specific motifs, region X is a ~16 aa segment that is between RT motifs 2 and 3 (21,22). Fourteen of the 16 experimentally-validated retron RTs contain an NAXXH aa sequence in region X. The two outliers, RT-Eco6 (Ec48) and RT-Mxa2 (Mx65), contain substitutions at the asparagine residue (Supplementary Figure S2). Notably, while both Group II intron and DGR RTs likewise contain residues between motifs 2 and 3, they lack the characteristic NAXXH motif seen in retron RTs. Deletion of region X from RT-Eco1 (Ec86) and RT-Eco3 (Ec73) does not block RT-cognate msr binding, but abrogates synthesis of the msDNA product, suggesting that it may be essential in reverse transcription initiation or catalysis (32). Notably, structures of Group II intron maturase RTs indicate that the equivalent region 2a forms an alpha helix that may stabilize the β hairpin containing the active site, influencing catalysis (41,42). We conjecture that region X in retron RTs may likewise stabilize the active site to promote catalysis.

Region Y is a ~90 aa segment at the C-terminus of retron RTs. Region Y is defined as beginning within motif 7 at a highly conserved VTG triplet and extending to the C-terminal end of the protein (12,32). Fourteen experimentally verified retron RTs contain the VTG triplet; RT-Vch2 (Vc81) and RT-Eco6 (Ec48) contain a V→I and T→H substitutions, respectively. Beyond the conserved VTG, region Y is variable. Biochemical studies indicate that region Y directs the RT to its cognate msr. Deleting region Y from RT-Eco1 (Ec86) and RT-Eco3 (Ec73) abrogated their ability to bind their msrs (32). Moreover, a purified C-terminal fragment of region Y of RT-Eco1 (Ec86) (residues 255–320) bound the stem-loop of msr-Eco1 (Ec86) with a K_D of $\sim 5 \times 10^{-8}$ M, as well as a shorter fragment containing an eight base-pair stem of the msr-Eco1 (Ec86) with a three base UUU loop, but not the msr-Eco3 (Ec73). Likewise, a purified C-terminal fragment of RT-Eco3 (Ec73) containing its region Y (residues 251–311) bound the cognate msr-Eco3 (Ec73) msr tightly (though the specific K_D was not measured), but not msr-Eco1 (Ec86) msr (36). Swapping region Ys between RT-Eco1 (Ec86) and RT-Eco3 (Ec73) produced chimeric proteins with ‘swapped’ msr recognition (32). However, all region Y studies to date have focused on Retron-Eco1 (Ec86) and Retron-Eco3 (Ec73). Additional biochemical experiments will be required to confirm that

region Y broadly recognizes the stem of its cognate msr in diverse retrons.

Accessory open reading frames in retron cassettes

Approximately a third of annotated retrons encode accessory open reading frames (ORFs) of unknown function (Figure 3A). These ORFs may be between the msr-msd and the RT (e.g., in Retron-Eco3 (Ec73), Retron-Sen2 (St85), and Retron-Vch2 (Vc81)) or immediately 3' of the RT (e.g. in Retron-Eco4 (Ec83)). Some of these ORFs are putatively annotated as adenosine-binding proteins, (e.g., in Retron-Eco4 (Ec83) (43)) or cold-shock proteins (e.g. in the *S. enterica* Enteridis Retron-Sen1 (Se72)) (44). Notably, some of the retrons with accessory ORFs also produce unusual msDNAs. For example, Retron-Eco4 (Ec83) produces single-stranded DNA via site-specific cleavage of its msDNA (45). Retron-Sen1 (Se72) shows an even more dramatic departure from the retron paradigm; the 2'-5' linkage is not experimentally observed (44,46). Here, the RNA template is degraded after the first round of RT, and the single-stranded DNA is used as a template to produce double-stranded DNA (dsDNA). RT-Sen1 uses the free 3'OH end of the msr to initiate the second round of DNA synthesis. Despite the unusual dsDNA product, RT-Sen1 (Se72) is structurally homologous with other retron RT sequences. Although it is tempting to speculate that the presence of accessory ORFs is important for producing unusual msDNA variants, additional biochemical studies are required to define their functions. These enigmatic ORFs may also provide opportunities for basic discovery and biotechnological applications.

Biological distribution and transmission

Experimental and bioinformatic studies indicate that retrons are sporadically distributed throughout the bacterial kingdom. Experimentally, msDNA has been isolated from a diverse subset of *Alpha*-, *Delta*- and *Gammaproteobacteria* (Table 1) (47). Screens of msDNA production, however, have commonly indicated that msDNA is not consistently produced by all strains within these taxonomic groups. Screens of *E. coli* identified msDNA in four of 89 and seven of 113 clinical and environmental isolates (48,49). Another study identified msDNA from 4 of 23 *Proteus mirabilis* isolates; 1 of 21 *Klebsiella pneumoniae* isolates, and 10 of 63 *Rhizobia* isolates (49). In contrast, msDNA production appears very common within the myxobacteria, with msDNA found in 27 of 28 tested myxobacterial strains (50).

In addition to these experimentally-identified retrons, bioinformatic surveys have identified putative retrons in *Cyanobacteria*, *Fusobacteria*, *Bacteroidetes*, *Firmicutes* (39), *Actinobacteria* (21,51) and even in the archaeon *Methanosarcina acetivorans* (39). Typically, these studies first identify retron-specific RTs in gene or protein databanks (20,21), metagenomic data (52,53), or in newly-sequenced bacterial genomes (51,54). Recent studies have additionally examined the regions upstream of the RT for palindromic and hairpin-containing motifs similar to msr-msd regions of experimentally validated retrons (55,56).

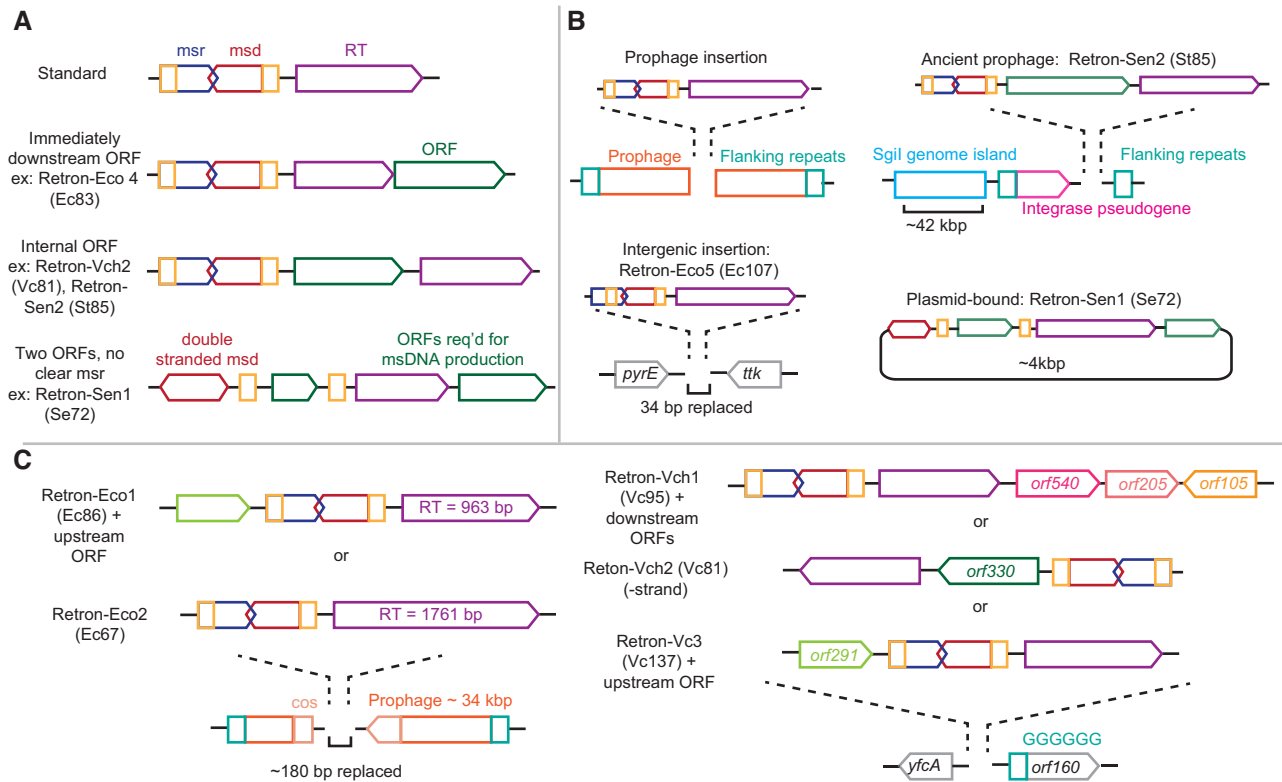


Figure 3. Organization and genomic context of select retron cassettes. (A) Structures of representative validated retron cassette genes. The msr (blue), msd (red), complementary repeat regions (yellow), and RT (purple), and associated unknown ORFs (green) are indicated. Many retron cassettes ('standard') consist only of an msr and msd flanked by complementary regions upstream of an RT. Variations on this theme include ORFs either immediately downstream of the RT or between the msr-msd and the RT. Retron-Sen1 (Se72) contains an msd that is transcribed in both directions to produce double-stranded DNA. A hairpin-containing ORF flanked by repeat regions may function as an alternative msr. The cassette also includes an RT and a second unknown ORF with homology to cold shock proteins (46). (B) Genomic contexts of retron cassettes. Several validated retrons occur in putative prophage sites flanked by repeat regions (12). Retron-Sen2 (St85) occurs in what is thought to be an ancient prophage site downstream of the SgiI genomic island associated with pathogenesis. Retron-Eco5 (Ec107) replaces 34 bases of an intergenic region with no apparent mobile element nearby. Retron-Sen1 (Se72) occurs on a plasmid. (C) Several different retrons, along with additional orfs, may occupy the same locus in different bacterial strains or populations. For example, the Retron Eco1 (Ec86) and an upstream ORF occupies the same genetic site as Retron-Eco2 (Ec67) in different *E. coli* strains (59,85). The *Vibrio cholerae* retrons Retron-Vch1 (Vc95), Retron-Vch2 (Vc81) and Retron-Vch3 (Vc137) also occupy the same genetic site.

The sporadic distribution of retrons within prokaryotic taxonomic groups and the mismatch in GC% content from their host strains suggest lateral transmission (12). While retrons do not appear to encode elements that support genetic mobility (e.g., RT endonuclease domains such as those required for the mobility of Group II introns), many are associated with mobile elements and prophage insertion sites (Figure 3B) (12). For example, Retron-Eco3 (Ec73) is inserted into a bacteriophage-like element termed ϕ R73 integrated into the chromosomal selenocystyl-tRNA locus. Infection of ϕ R73-containing *E. coli* cells with P2 phage produced infectious virions that lysogenized new host strains, resulting in the integration of ϕ R73 at the selenocystyl-tRNA locus and production of msDNA-Eco3 (Ec73) (57,58). Retron-Eco1 (Ec86) and Retron-Eco2 (Ec67) both replace an identical ~180 bp between a phage cohesive end site and the terminator of a DNA packaging gene (59) (Figure 3A). Retron-Sen1 (Se72) even exists on a bacterial plasmid (44). Retron-Nex1 (Ne160) contains several apparently duplicated repeats of msr and msd sequences without an RT within a kilobase, suggesting that this element may have been repeatedly du-

plicated and inserted by a distant RT or alternative element (60). Other retrons, however, are not associated with obvious prophage or other mobile elements. Retron-Eco5 (Ec107), for example, replaces a palindromic 34 bp in an intergenic region that lacks any further characteristics of mobility or prophage insertion. In addition to horizontal transfer, many myxococcal retrons, including Retron-Mxa1 (Mx162) and Retron-Sau1 (Sa163), cluster in closely related bacterial species and show a codon bias that is similar to that of the host. These retrons likely arose or were acquired in a common ancestor of the myxobacteria and were vertically inherited (50).

Hypothesized biological functions

Retrons' functional role—if any—remains poorly understood. Some retrons could potentially be purely selfish genetic elements that are neutral or weakly deleterious to the host. However, the observation that retrons encode functional proteins suggests that they may be beneficial to the host, possibly under specialized conditions (61). In the following paragraphs, we discuss the retrons with known func-

tions and consider potential roles for those whose function remains unknown.

Circumstantial evidence suggests that some retrons may play key roles in pathogenesis. Disabling Retron-Sen2 (St85) or Retron-Eco4 (Ec83) by erasing their ability to produce msDNA reduced the ability of their enteropathogenic *S. enterica* Typhimurium and *E. coli* hosts to persistently colonize mouse intestine (62). Closer studies of the msDNA-null *S. enterica* Typhimurium cells indicated severe growth defects in anaerobic conditions. Growth was rescued by the addition of nitrate, a terminal electron acceptor typically present in aerobic environments. Proteome analysis indicated differential expression of 238 proteins between wild-type and msDNA-lacking *S. enterica* Typhimurium under anaerobic conditions, most involved in the reduction of electron acceptors present in anaerobic environments. While the patterns of protein regulation did not follow those associated with the major known regulatory proteins in *S. enterica* Typhimurium, the authors concluded that in some genetic or epigenetic way Retron-Sen2 (St85) controls protein abundance that in turn enables the use of the electron acceptors present in anaerobic environments. Conversely, other studies found that deletion of RT-Sen2 (Se85) from other, multidrug-resistant strains of *S. enterica* Typhimurium increased virulence, suggesting an opposing role in the regulation of virulence for this retron (63).

The apparent phylogenetic subgroup obtained from aligning RTs that includes Retron-Sen2 (St85) and Retron-Eco4 (Ec83) also contains retrons from other pathogenic bacterial serovars, including Retron-Eco7 (Ec78), Retron-Vch1 (Vc95), and Retron-Vpa3 (Vp96). The RTs associated with these retrons have relatively similar sequences, with aa identity of ~40–58% (Supplementary Table S1). Several also contain accessory ORFs between the msd and RT or downstream of the RT (62,64–66) (Figure 3A). Interestingly, the msrs of two retron pairs in this group, msr-Vch1 (Vc95)/msr-Vpa1 (Vp96) and msr-Eco4 (Ec83)/msr-Eco7 (Ec78), share ~75% and ~60% sequence identity (63). Correspondingly, RT-Vch1 (Vc95) and RT-Eco4 (Ec83) recognize and reverse transcribe msr-msd-Vpa3 (Vp96) and msr-msd-Eco7 (Ec78), respectively (35). These are the only two examples where an RT can function on a non-cognate msr-msd. Moreover, Retron-Vch1 (Vc95) is almost exclusively found in the epidemic cholera-causing serotypes (O1 and its derivative O139); this retron is very rare in non-pathogenic serotypes (66). In contrast, two non-pathogenic serotypes contain alternate retrons in the same genomic locus—Retron-Vch2 (Vc81) and Retron-Vch3 (Vc137). Retron-Vch2 (Vc81) is also oriented in the opposite direction from Retron-Vch1 (Vc95) (64). The presence of certain retron variants only in pathogenic species, and the similarities between such retrons, has led to speculation that retrons may play a mechanistic role in pathogenesis (35). Alternatively, the ORFs of unknown function could themselves be pathogenicity determinants that associate with retrons to hitchhike between genomes (Figure 3C).

Beyond potential roles in pathogenesis, retrons may play adaptive roles in the utilization of non-traditional energy sources. For example, expression of Retron-Eco5 (Ec107) is strongly enhanced in cells that are starved for phos-

phate or entering stationary phase in rich media (67). Similarly, Retron-Eco1 (Ec86) expression resulted in significant downregulation of eight genes and upregulation of one gene. The eight downregulated genes were important in carbon utilization; the remaining gene upregulated the production of the ribosomal protein S2 (65). Beyond non-pathogenic *E. coli*, retrons are especially common in myxococcal bacteria that undergo significant morphological changes, forming multicellular-like, ‘fruiting’ bodies in response to nutrient insufficiency, and show complex intercommunity behavior and predation (68); perhaps retrons in these species help enable these phenotypic changes. Additional studies are required to link retron expression and cell specialization.

While there are several hypotheses about retrons’ biological roles, there is little known about their mechanism of action. One possibility is that retrons modulate host DNA-binding proteins via binding to the msDNA. Overexpression of some *E. coli* retrons producing DNA with stems containing mismatched base pairs increases genomic mutation rates by ~40-fold (69) and recombination rates in high-frequency recombination (Hfr) crosses with a competent *S. enterica* Typhimurium donor strain ~5–10-fold (70). This increase in genomic mutation rate likely occurs because the mismatch repair protein MutS is sequestered onto the mismatch-containing msDNA stem, reducing protein availability for genomic replication-associated mismatch repair. Notably, overexpression of retrons that produce msDNA with fully complementary stems does not increase mutation or Hfr rates (70,71). While these increased mutation frequencies only occurred after ~100-fold overexpression of the msDNA relative to standard lab growth conditions, this study may mimic intracellular or environmental conditions that increase retron expression to transiently increase genomic mutation rates (69). msDNA may sequester other DNA-binding proteins to regulate their activity and, in turn, gene expression patterns (72,73). Additionally, msDNA may directly act as a post-transcriptional gene regulator, similarly to microRNAs or RNA-binding proteins (74).

Finally, retrons might play a role in targeted genome diversification or adaptive immunity. DGRs are found in bacteria, viruses, and archaea and selectively diversify targeted loci known as ‘variable’ regions (75). These variable regions typically encode surface-displayed peptides that bind a changing set of ligands in their environments (76,77). To target and diversify these regions, DGRs employ specialized RTs to reverse transcribe homologous regions known as ‘template’ regions, incorporating random nucleotides opposite from adenine residues on the RNA transcript intermediate. The mutated reverse transcript then specifically retrotransposes to the variable region. While the details of reverse transcription and transposition have yet to be determined, recent studies have found that, like retron-produced msDNA, the prototypical *Bordetella* bacteriophage *brt* RT produces a branched RNA-DNA structure most likely joined by 2’-5’ phosphodiester bonds (78,79). These similarities between retron msDNA and DGR cDNA suggest that perhaps retron msDNA may home to genetic sequences via a mechanism that is analogous to DGR retrohoming. Alternatively, a recent study suggests that certain

CRISPR-related RTs may have arisen from retron-like RTs, suggesting that retrons could potentially play a similar role in adaptive immunity or integration of specific RNA or DNA sequences into defined genomic positions (80).

RETRONS IN BIOTECHNOLOGY

Retrons are potentially powerful tools for genome editing because they can produce high copy number intracellular DNAs in orthogonal hosts. Early experiments indicated that a Retron-Eco2 (Ec67) msr and RT could successfully reverse transcribe a Retron-Eco3 (Ec73) msd, and vice-versa (33). This critical experiment defined that while a retron's msr and RT are paired, the msd is variable and can encode an intracellular DNA with an arbitrary sequence. This critical insight enabled the repurposing of retrons for biotechnological applications. To date, four studies have employed retrons for gene silencing or genome editing (13–15,81). Below, we discuss the context, central findings, and potential further development of these studies.

Gene targeting in bacteria

An early study demonstrated that retrons can be engineered to produce antisense cDNAs that hybridized to and ultimately silenced, mRNAs of the targeted genes. Shimada and colleagues replaced a portion of the msd of Retron-Eco5 (Ec107) with a sequence targeting the translation initiation sites of an mRNA encoding the *E. coli* outer membrane lipoprotein (81). The authors tested two different targeting sequences, as well as a targeting sequence containing an EcoRI digestion site, potentially enabling cleavage of the sequence targeting lipoprotein. Each targeting sequence inhibited lipoprotein production ~75%.

More recently, a landmark study used retrons as bacterial genome-editing tools (13). In this work, small molecule-inducible retrons acted as a population-level genomic analog 'tape-recorder' of the small molecules' presence in *E. coli* (Figure 4A, B). Termed Synthetic Cellular Recorders Integrating Biological Events ('SCRIBE'), this approach used engineered variants of Retron-Eco1 (Ec86) with msds programmed to correct two premature stop codons in reporter genes, ultimately editing a total of 5–6 contiguous bases. Co-expressing the retron msr-msd with the bacteriophage lambda red protein Beta (a single-stranded DNA-binding protein commonly employed in recombineering (5)) increased editing frequency of the retron transcript a thousand-fold (from $\sim 10^{-7}$ /cell to $\sim 10^{-4}$ /cell) after overnight expression. This improved genome editing efficiency was appropriate for detection of modified cells within a heterogeneous population. SCRIBE could (i) quantitatively record inducers' concentration and duration by measuring the proportion of cells containing retron-introduced edits; (ii) record exposure to complex combinations of inducer molecules and (iii) record cellular exposure to light via the use of a light-responsive promoter.

Retron-based editing in bacteria has been further adapted to support the possibility of continuous evolution of target loci (15) (Figure 4C). Knocking out MutS, previously shown to bind retron msDNA, further increased editing efficiency 21-fold compared to wild-type cells (70).

Removing exonuclease ExoX, which degrades double and single-stranded DNAs from the 3' to 5' end, increased editing efficiency 5-fold (82). An *E. coli* that harbored the double $\Delta mutS \Delta exoX$ deletion increased retron-based editing efficiency 78-fold to ~ 0.05 /cell after overnight expression. Measurements with a similar kanamycin resistance-correcting model system to that employed in the bacterial tape recorder indicated that retrons could simultaneously edit up to 13 bases spanning a 31-base region, and insert and delete three-base regions. Expressing the retron in this optimized system under an error-prone RNA polymerase led to the introduction of unprogrammed, random mutations in the targeted gene, setting the stage for continuous evolution schemes. Overall, these modifications increased the mutation rate of the targeted region to $\sim 10^{-6}$ base $^{-1}$ generation $^{-1}$, ~ 200 -fold above background.

Heterologous expression of retrons

Biotechnological retron applications were bolstered by early studies that retrons can produce satellite DNAs in eukaryotic hosts. For example, Retron-Eco2 (Ec67) expressed under the control of a Gal10 promoter produced msDNA in *S. cerevisiae* (83). Robust msDNA production required the concurrent expression of the RT and the msr-msd. The resulting msDNA was the same length as seen in the native *E. coli* host. In yeast, moving the RT to the 5' side of the msr-msd region increased msDNA production seven-fold over that produced in the native orientation where the RT is to the 3' side of the msr-msd region (83). Retron-Eco2 (Ec67) could also produce msDNA in mouse embryonic fibroblasts by expressing this 5' RT-containing variant with T7 RNA polymerase (84). However, msDNA copy number was ~ 25 copies per cell, ~ 25 -fold lower than seen in the native bacterial host. Nonetheless, these early studies established that retrons can be expressed in orthogonal hosts and that the host's RNase H activity was sufficient to process the reverse-transcribed msd to generate msDNA.

More recently, retrons have been used for aiding homology-directed repair of Cas9-targeted breaks in yeast in a platform termed Cas9 Retron precISE Parallel Editing via homologY ('CRISPEY') (14) (Figure 4D). The msr-msd of Retron-Eco1 (Ec86) was appended to a Cas9 single guide RNA (sgRNA) that targeted a genomic region of interest. Co-expression of this chimeric sgRNA-retron RNA, RT-Eco1 (Ec86) and Cas9 led to the co-localization of two functional units: (i) a Cas9 nuclease that generated a defined DNA break in the vicinity of the edit site; (ii) a DNA template derived from the msd segment of the sgRNA-retron RNA fusion. After Cas9 and RT induction, the reverse-transcribed msd could carry out homology-directed DNA break repair at high efficiency. Single nucleotide polymorphisms were incorporated in $>96\%$ of all clones and a 765 base pair insertion was observed in 87% of edited yeast cells. Ultimately, this technique was employed for the high-efficiency profiling of the effect of $\sim 16,000$ single-nucleotide polymorphisms in yeast, enabling identification of 572 variants that impacted fitness during growth on glucose. This approach can now potentially be applied to the combinatorial diversification of targeted regions, similar to combina-

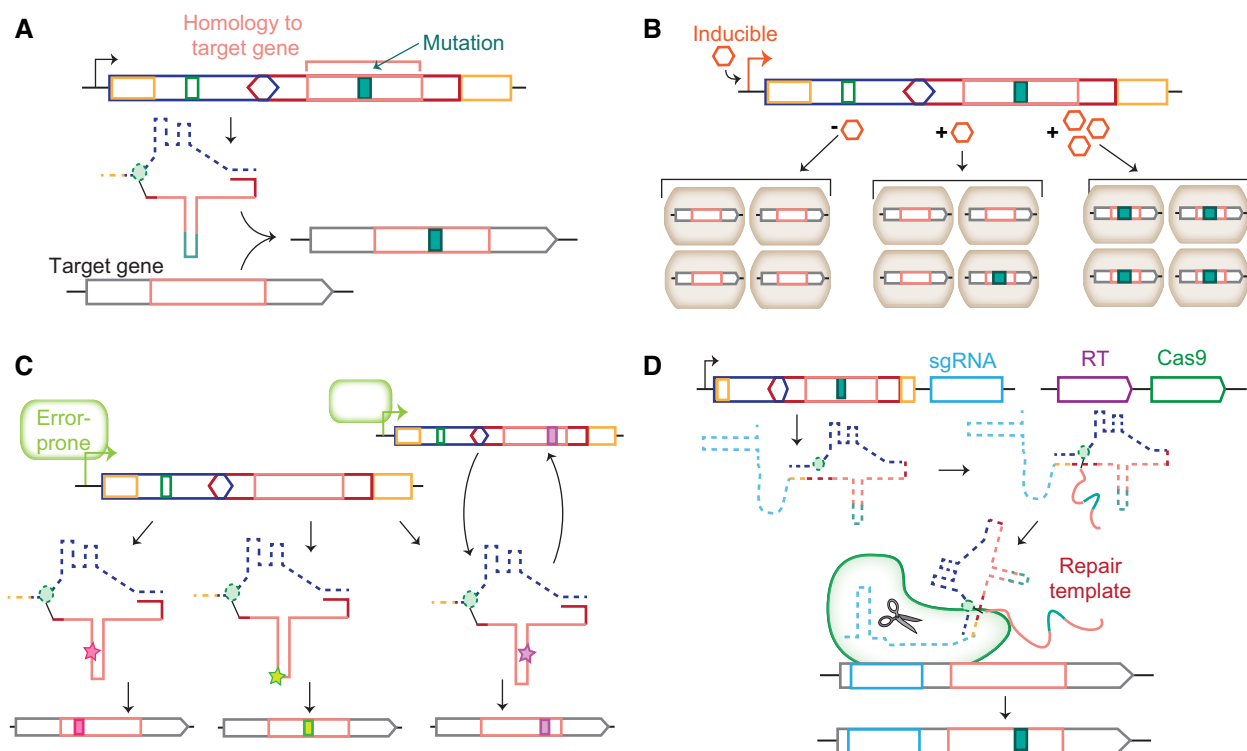


Figure 4. Biotechnological applications of retrons. (A) In retron-based bacterial gene editing, the msd is homologous to a segment of the target gene (pink). The desired mutation is also introduced into the msd (turquoise). msDNA (thin dashed/solid line) is produced and edits the targeted sequence via an undetermined mechanism. (B) Schematic of retron-mediated genomic recording via Synthetic Cellular Recorders Integrating Biological Events, e.g., ‘SCRIBE’ (13). A small-molecule inducible retron is designed to turn on or off a selectable marker. Switching of the selectable marker occurs only in the presence of the small molecule inducer (orange hexagon). The desired edit is produced in proportion to the duration and concentration of inducer, enabling ‘tape recording’ of the inducer’s presence. (C) Retron-mediated mutagenesis and evolution (15). In this application, a retron homologous to the desired target gene is expressed under an error-prone T7 RNA polymerase (green). The resulting msDNAs have random mutations (stars) that are subsequently introduced into their target gene as in (A). Alternatively, these mutation-containing msDNAs can edit the parent retron sequence, enabling continuous accumulation of mutations and evolution. (D) Schematic for Cas9 Retron preCISE Parallel Editing via homology, e.g., ‘CRISPEY’ (14). A retron msr/msd sequence containing homology and a desired edit to a targeted gene is expressed as a fusion to a sgRNA (light blue) targeting the gene of interest. The retron RT (purple) and Cas9 (green) are expressed *in trans*. Cas9 generates a DNA break and the RT-generated DNA is used as a template by the host’s DNA repair machinery. Genes are not drawn to scale.

torial recombineering-based platforms such as MAGE (8), YOGRE (9) or iCREATE (10).

CONCLUSIONS AND OUTLOOK

A recent renaissance in retron research is fuelled by their broad phylogenetic distribution, intriguing potential linkages to pathogenesis, and unique biotechnological capabilities. Even with their newfound popularity, retron function, mechanism of action, and evolutionary origins remain obscure. One critical gap is the lack of a retron RT structure, either alone or as part of a priming complex. Thus, we can only conjecture about the unique priming mechanism and the functions of accessory domains in retron RTs. The functions of other retron-associated ORFs are also unexplored. As with many natural biomolecules that have proven indispensable to modern biotechnologies, an improved understanding of natural retron function and biology should greatly accelerate biotechnology applications.

Retrons are unique in their ability to produce satellite DNAs with arbitrary sequences *in vivo*. This is a feature that suddenly thrusts retrons to the fore of genome editing techniques, as they can be harnessed to enable genome edit-

ing without the need to introduce external oligonucleotides via electroporation or transfection. On the horizon are opportunities to harness retrons as dynamic, self-diversifying templates for editing targeted genes, thereby enabling targeted continuous evolution. Further optimizing retron expression and function in eukaryotic cells and organisms will likely enable diverse applications, such as the ability to mark cells during embryogenesis and carry out lineage analyses via continuous evolution of barcodes. In another vein, retron-produced DNA may also prove useful for generating highly programmable DNA nanostructures inside cells. Finally, retron-derived RTs may have unique applications on their own, just as Group II intron RTs have found applications for NGS library preparation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Y.G. Garcia Huante, L. Lopez, and members of the Finkelstein lab for helpful discussion and

assistance with illustrations. We also thank the broader retroelement community for invaluable feedback relating to the revised retron nomenclature described herein.

FUNDING

Welch Foundation [F-1808 to I.J.F., F-1654 to A.D.E.]; National Institutes of Health [GM124141 to I.J.F.]; Arnold O. Beckman Postdoctoral Fellowship held by A.J.S. Funding for open access charge: National Institutes of Health [GM124141 to I.J.F.].

Conflict of interest statement. None declared.

REFERENCES

- Greener, A., Callahan, M. and Jerpseth, B. (1996) In: Trower, M.K. (ed). *In Vitro Mutagenesis Protocols*. Humana Press, Totowa, pp. 375–385.
- Petri, R. and Schmidt-Dannert, C. (2004) Dealing with complexity: evolutionary engineering and genome shuffling. *Curr. Opin. Biotechnol.*, **15**, 298–304.
- Chen, K. and Arnold, F.H. (1991) Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Bio/Technology*, **9**, 1073.
- Cramer, A., Raillard, S.-A., Bermudez, E. and Stemmer, W.P.C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, **391**, 288.
- Sharan, S.K., Thomason, L.C., Kuznetsov, S.G. and Court, D.L. (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.*, **4**, 206–223.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281.
- Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R. and Church, G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.
- DiCarlo, J.E., Conley, A.J., Penttilä, M., Jantti, J., Wang, H.H. and Church, G.M. (2013) Yeast oligo-mediated genome engineering (YOGIE). *ACS Synth. Biol.*, **2**, 741–749.
- Liu, R., Liang, L., Garst, A.D., Choudhury, A., Nogué, V.S.I., Beckham, G.T. and Gill, R.T. (2018) Directed combinatorial mutagenesis of *Escherichia coli* for complex phenotype engineering. *Metab. Eng.*, **47**, 10–20.
- Lampson, B.C. (2007), *Industrial Enzymes*, pp. 403–420.
- Lampson, B.C., Inouye, M. and Inouye, S. (2005) Retrons, msDNA, and the bacterial genome. *Cytogenet. Genome Res.*, **110**, 491–499.
- Farzadfar, F. and Lu, T.K. (2014) Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science*, **346**, 1256272.
- Sharon, E., Chen, S.A., Khosla, N.M., Smith, J.D., Pritchard, J.K. and Fraser, H.B. (2018) Functional genetic variants revealed by massively parallel precise genome editing. *Cell*, **175**, 544–557.
- Simon, A.J., Morrow, B.R. and Ellington, A.D. (2018) Retroelement-based genome editing and evolution. *ACS Synth. Biol.*, **7**, 2600–2611.
- Yee, T., Furuichi, T., Inouye, S. and Inouye, M. (1984) Multicopy single-stranded DNA isolated from a gram-negative bacterium, *Myxococcus xanthus*. *Cell*, **38**, 203–209.
- Inouye, S., Hsu, M.-Y., Eagle, S. and Inouye, M. (1989) Reverse transcriptase associated with the biosynthesis of the branched RNA-linked msDNA in *Myxococcus xanthus*. *Cell*, **56**, 709–717.
- Hsu, M.-Y., Eagle, S.G., Inouye, M. and Inouye, S. (1992) Cell-free synthesis of the branched RNA-linked msDNA from retron-Ec67 of *Escherichia coli*. *J. Biol. Chem.*, **267**, 13823–13829.
- Shimamoto, T., Inouye, M. and Inouye, S. (1995) The formation of the 2',5'-phosphodiester linkage in the cDNA priming reaction by bacterial reverse transcriptase in a cell-free system. *J. Biol. Chem.*, **270**, 581–588.
- Simon, D.M. and Zimmerly, S. (2008) A diversity of uncharacterized reverse transcriptases in bacteria. *Nucleic Acids Res.*, **36**, 7219–7229.
- Toro, N. and Nisa-Martínez, R. (2014) Comprehensive phylogenetic analysis of bacterial reverse transcriptases. *PLoS One*, **9**, e114083.
- Zimmerly, S. and Wu, L. (2015) An unexplored diversity of reverse transcriptases in bacteria. *Microbiol. Spectrum*, **3**, doi:10.1128/microbiolspec.MDNA3-0058-2014.
- Lampson, B., Sun, J., Hsu, M., Vallejo-Ramirez, J., Inouye, S. and Inouye, M. (1989) Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science*, **243**, 1033–1038.
- Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.K., Dryden, D.T.F., Dybvig, K. et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
- Xie, X. and Yang, R. (2017) Multi-copy single-stranded DNA in *Escherichia coli*. *Microbiology*, **163**, 1735–1739.
- Shimamoto, T., Kawanishi, H., Tsuchiya, T., Inouye, S. and Inouye, M. (1998) *In vitro* synthesis of multicopy single-stranded DNA, using separate primer and template RNAs, by *Escherichia coli* reverse transcriptase. *J. Bacteriol.*, **180**, 2999–3002.
- Shimamoto, T., Shimada, M., Inouye, M. and Inouye, S. (1995) The role of ribonuclease H in multicopy single-stranded DNA synthesis in retron-Ec73 and retron-Ec107 of *Escherichia coli*. *J. Bacteriol.*, **177**, 264–267.
- Kim, K., Jeong, D. and Lim, D. (1997) A mutational study of the site-specific cleavage of EC83, a multicopy single-stranded DNA (msDNA): nucleotides at the msDNA stem are important for its cleavage. *J. Bacteriol.*, **179**, 6518–6521.
- Lease, R.A. and Yee, T. (1991) Early events in the synthesis of the multicopy single-stranded DNA-RNA branched copolymer of *Myxococcus xanthus*. *J. Biol. Chem.*, **266**, 14497–14503.
- Lim, D. (1992) Structure and biosynthesis of unbranched multicopy single-stranded DNA by reverse transcriptase in a clinical *Escherichia coli* isolate. *Mol. Microbiol.*, **6**, 3531–3542.
- Inouye, S., Herzer, P.J. and Inouye, M. (1990) Two independent retrons with highly diverse reverse transcriptases in *Myxococcus xanthus*. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 942–945.
- Inouye, S., Hsu, M.-Y., Xu, A. and 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.
- Shimamoto, T., Hsu, M.-Y., Inouye, S. and Inouye, M. (1993) Reverse transcriptases from bacterial retrons require specific secondary structures at the 5'-end of the template for the cDNA priming reaction. *J. Biol. Chem.*, **268**, 2684–2692.
- Lima, T.M.O. and Lim, D. (1997) A novel retron that produces RNA-less msDNA in *Escherichia coli* using reverse transcriptase. *Plasmid*, **38**, 25–33.
- Shimamoto, T., Ahmed, A.M. and Shimamoto, T. (2013) A novel retron of *Vibrio parahaemolyticus* is closely related to retron-Vc95 of *Vibrio cholerae*. *J. Microbiol.*, **51**, 323–328.
- Inouye, M., Ke, H.P., Yashio, A., Yamanaka, K., Nariya, H., Shimamoto, T. and 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.
- Paul, B.G., Burstein, D., Castelle, C.J., Handa, S., Arambula, D., Czornyj, E., Thomas, B.C., Ghosh, P., Miller, J.F., Banfield, J.F. et al. (2017) Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. *Nat. Microbiol.*, **2**, 17045.
- Papadopoulos, J.S. and Agarwala, R. (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, **23**, 1073–1079.
- Rest, J.S. and Mindell, D.P. (2003) Retroviruses in archaea: phylogeny and lateral origins. *Mol. Biol. Evol.*, **20**, 1134–1142.
- Xiong, Y. and Eickbush, T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, **9**, 3353–3362.
- Qu, G., Kaushal, P.S., Wang, J., Shigematsu, H., Piazza, C.L., Agrawal, R.K., Belfort, M. and Wang, H.-W. (2016) Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.*, **23**, 549.
- Zhao, C. and Pyle, A.M. (2017) The group II intron maturase: a reverse transcriptase and splicing factor go hand in hand. *Curr. Opin. Struct. Biol.*, **47**, 30–39.

43. Jeong, D.W., Kim, K. and Lim, D.B. (1997) Evidence for the complex formation between reverse transcriptase and multicopy single-stranded DNA in retron EC83. *Mol. Cells*, **7**, 347–351.
44. Rychlik, I., Sebkova, A., Gregorova, D. and Karpiskova, R. (2001) Low-molecular-weight plasmid of *Salmonella enterica* serovar Enteritidis codes for retron reverse transcriptase and influences phage resistance. *J. Bacteriol.*, **183**, 2852–2858.
45. Lampson, B.C., Viswanathan, M., Inouye, M. and Inouye, S. (1990) Reverse-transcriptase from *Escherichia coli* exists as a complex with msDNA and is able to synthesize double-stranded DNA. *J. Biol. Chem.*, **265**, 8490–8496.
46. Pilousova, L. and Rychlik, I. (2011) Retron Se72 utilizes a unique strategy of the self-priming initiation of reverse transcription. *Cell. Mol. Life Sci.*, **68**, 3607–3617.
47. Matiasovicova, J., Faldynova, M., Pravcova, M., Karpiskova, R., Kolackova, I., Damborsky, J. and Rychlik, I. (2003) Retron reverse transcriptase rrtT is ubiquitous in strains of *Salmonella enterica* serovar Typhimurium. *FEMS Microbiol. Lett.*, **223**, 281–286.
48. Sun, J., Herzer, P.J., Weinstein, M.P., Lampson, B.C., Inouye, M. and Inouye, S. (1989) Extensive diversity of branched-RNA-linked multicopy single-stranded DNAs in clinical strains of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, **86**, 7208–7212.
49. Rice, S.A. and Lampson, B.C. (1995) Bacterial reverse transcriptase and msDNA. *Virus Genes*, **11**, 95–104.
50. Rice, S.A. and Lampson, B.C. (1995) Phylogenetic comparison of retron elements among the myxobacteria: evidence for vertical inheritance. *J. Bacteriol.*, **177**, 37–45.
51. Guevara, G., Castillo Lopez, M., Alonso, S., Perera, J. and Navarro-Llorens, J.M. (2019) New insights into the genome of *Rhodococcus ruber* strain Chol-4. *BMC Genomics*, **20**, 332.
52. Labonté, J.M., Field, E.K., Lau, M., Chivian, D., Van Heerden, E., Wommack, K.E., Kieft, T.L., Onstott, T.C. and Stepanauskas, R. (2015) Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface *Firmicutes* population. *Front. Microbiol.*, **6**, 349–349.
53. Lescot, M., Hingamp, P., Kojima, K.K., Villar, E., Romac, S., Veluchamy, A., Boccara, M., Jaillon, O., Iudicone, D., Bowler, C. *et al.* (2015) Reverse transcriptase genes are highly abundant and transcriptionally active in marine plankton assemblages. *ISME J.*, **10**, 1134.
54. Olsson, B.E., Korsakova, E.S., Anan'ina, L.N., Pyankova, A.A., Mavrodi, O.V., Plotnikova, E.G. and Mavrodi, D.V. (2017) Draft genome sequences of strains *Salinicola socius* SMB35T, *Salinicola* sp. MH3R3-1 and *Chromohalobacter* sp. SMB17 from the Verkhnekamsk potash mining region of Russia. *Standards Genomic Sci.*, **12**, 39.
55. Das, R., Shimamoto, T. and Arifuzzaman, M. (2011) A novel msDNA (multicopy single-stranded DNA) strain present in *Yersinia frederiksenii* ATCC 33641 contig01029 enteropathogenic bacteria with the genomic analysis of its retron. *J. Pathog.*, **2011**, 693769.
56. Yoo, W., Lim, D. and Kim, S. (2016) Compiling multicopy single-stranded DNA sequences from bacterial genome sequences. *Genomics Informatics*, **14**, 29–33.
57. Inouye, S., Sunshine, M., Six, E. and Inouye, M. (1991) Retronophage phi R73: an *E. coli* phage that contains a retroelement and integrates into a tRNA gene. *Science*, **252**, 969–971.
58. Sun, J., Inouye, M. and Inouye, S. (1991) Association of a retroelement with a P4-like cryptic prophage (retronophage phi R73) integrated into the selenocystyl tRNA gene of *Escherichia coli*. *J. Bacteriol.*, **173**, 4171–4181.
59. Dodd, I.B. and Egan, J.B. (1996) The *Escherichia coli* retrons Ec67 and Ec86 replace DNA between the *cos* site and a transcription terminator of a 186-related prophage. *Virology*, **219**, 115–124.
60. Lampson, B.C. and Rice, S.A. (1997) Repetitive sequences found in the chromosome of the myxobacterium *Nannocystis exedens* are similar to msDNA: a possible retrotransposition event in bacteria. *Mol. Microbiol.*, **23**, 813–823.
61. Linquist, S., Cottenie, K., Elliott, T.A., Saylor, B., Kremer, S.C. and Gregory, T.R. (2015) Applying ecological models to communities of genetic elements: the case of neutral theory. *Mol. Ecol.*, **24**, 3232–3242.
62. Elfenbein, J.R., Knodler, L.A., Nakayasu, E.S., Ansong, C., Brewer, H.M., Bogomolnaya, L., Adams, L.G., McClelland, M., Adkins, J.N. and Andrews-Polymenis, H.L. (2015) Multicopy single-stranded DNA directs intestinal colonization of enteric pathogens. *PLoS Genet.*, **11**, e1005472.
63. Pilousova, L., Matiasovicova, J., Sisak, F., Havlickova, H. and Rychlik, I. (2005) Retron reverse transcriptase (rrtT) can be lost in multidrug resistant *Salmonella enterica* serovar Typhimurium DT 104 strains and influences virulence for mice. *Vet. Microbiol.*, **111**, 191–197.
64. Inouye, K., Tanimoto, S., Kamimoto, M., Shimamoto, T. and Shimamoto, T. (2011) Two novel retron elements are replaced with retron-Vc95 in *Vibrio cholerae*. *Microbiol. Immunol.*, **55**, 510–513.
65. Jeong, M.-A. and Lim, D. (2004) A proteomic approach to study msDNA function in *Escherichia coli*. *J. Microbiol.*, **42**, 200–2004.
66. Shimamoto, T., Kobayashi, M., Tsuchiya, T., Shinoda, S., Kawakami, H., Inouye, S. and Inouye, M. (1999) A retroelement in *Vibrio cholerae*. *Mol. Microbiol.*, **34**, 631–632.
67. Herzer, P.J. (1996) Starvation-induced expression of retron-Ec107 and the role of ppGpp in multicopy single-stranded DNA production. *J. Bacteriol.*, **178**, 4438–4444.
68. Thakur, P., Chopra, C., Anand, P., Dhanjal, D.S. and Chopra, R.S. (2018) In: Singh, J., Sharma, D., Kumar, G. and Sharma, N.R. (eds). *Microbial Bioprospecting for Sustainable Development*. Springer, Singapore, pp. 137–163.
69. Maas, W.K., Wang, C., Lima, T., Zubay, G. and Lim, D. (1994) Multicopy single-stranded DNAs with mismatched base pairs are mutagenic in *Escherichia coli*. *Mol. Microbiol.*, **14**, 437–441.
70. Maas, W.K., Wang, C., Lima, T., Hach, A. and Lim, D. (1996) Multicopy single-stranded DNA of *Escherichia coli* enhances mutation and recombination frequencies by titrating MutS protein. *Mol. Microbiol.*, **19**, 505–509.
71. Mao, J.-R., Inouye, S. and Inouye, M. (1996) Enhancement of frame-shift mutation by the overproduction of msDNA in *Escherichia coli*. *FEMS Microbiol. Lett.*, **144**, 109–115.
72. Brewster, R.C., Weinert, F.M., Garcia, H.G., Song, D., Rydenfelt, M. and Phillips, R. (2014) The transcription factor titration effect dictates level of gene expression. *Cell*, **156**, 1312–1323.
73. Todeschini, A.-L., Georges, A. and Veitia, R.A. (2014) Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet.*, **30**, 211–219.
74. Jens, M. and Rajewsky, N. (2014) Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat. Rev. Genet.*, **16**, 113.
75. Wu, L., Gingery, M., Abebe, M., Arambula, D., Czornyj, E., Handa, S., Khan, H., Liu, M., Pohlschroder, M., Shaw, K.L. *et al.* (2018) Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey. *Nucleic Acids Res.*, **46**, 11–24.
76. Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., Maskell, D.J., Simons, R.W., Cotter, P.A., Parkhill, J. *et al.* (2002) Reverse transcriptase-mediated tropism switching in *Bordetella* bacteriophage. *Science*, **295**, 2091–2094.
77. Doulatov, S., Hodes, A., Dai, L., Mandhana, N., Liu, M., Deora, R., Simons, R.W., Zimmerly, S. and Miller, J.F. (2004) Tropism switching in *Bordetella* bacteriophage defines a family of diversity-generating retroelements. *Nature*, **431**, 476.
78. Naorem, S.S., Han, J., Wang, S., Lee, W.R., Heng, X., Miller, J.F. and Guo, H. (2017) DGR mutagenic transposition occurs via hypermutagenic reverse transcription primed by nicked template RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E10187–E10195.
79. Handa, S., Jiang, Y., Tao, S., Foreman, R., Schinazi, R.F., Miller, J.F. and Ghosh, P. (2018) Template-assisted synthesis of adenine-mutagenized cDNA by a retroelement protein complex. *Nucleic Acids Res.*, **46**, 9711–9725.
80. Toro, N., Martínez-Abarca, F., Mestre, M.R. and González-Delgado, A. (2019) Multiple origins of reverse transcriptases linked to CRISPR-Cas systems. *RNA Biol.*, **16**, 1–8.
81. Mao, J.R., Shimada, M., Inouye, S. and Inouye, M. (1995) Gene regulation by antisense DNA produced *in vivo*. *J. Biol. Chem.*, **270**, 19684–19687.
82. Viswanathan, M. and Lovett, S.T. (1999) Exonuclease X of *Escherichia coli*. A novel 3'-5' DNase and DnaQ superfamily member involved in DNA repair. *J. Biol. Chem.*, **274**, 30094–30100.
83. Miyata, S., Ohshima, A., Inouye, S. and Inouye, M. (1992) *In vivo* production of a stable single-stranded cDNA in *Saccharomyces cerevisiae* by means of a bacterial retron. *PNAS*, **89**, 5735–5739.

84. Mirochnitchenko, O., Inouye, S. and Inouye, M. (1994) Production of single-stranded DNA in mammalian cells by means of a bacterial retron. *J. Biol. Chem.*, **269**, 2380–2383.
85. Lim, D. (1991) Structure of two retrons of *Escherichia coli* and their common chromosomal insertion site. *Mol. Microbiol.*, **5**, 1863–1872.
86. Lim, D. and Maas, W.K. (1989) Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. *Cell*, **56**, 891–904.
87. Herzer, P.J., Inouye, S. and Inouye, M. (1992) Retron-Ec107 is inserted into the *Escherichia coli* genome by replacing a palindromic 34bp intergenic sequence. *Mol. Microbiol.*, **6**, 345–354.
88. Mao, J.R., Inouye, M. and Inouye, S. (1996) An unusual bacterial reverse transcriptase having LVDD in the YXDD box from *Escherichia coli*. *Biochem. Biophys. Res. Commun.*, **227**, 489–493.
89. Lampson, B.C., Xu, C., Rice, S.A. and Inouye, S. (2002) A partial copy of msDNA from a new retron element is likely a retrotransposed DNA found in the myxobacterium *Nannocystis exedens*. *Gene*, **299**, 251–261.
90. Ahmed, A.M. and Shimamoto, T. (2003) msDNA-St85, a multicopy single-stranded DNA isolated from *Salmonella enterica* serovar Typhimurium LT2 with the genomic analysis of its retron. *FEMS Microbiol. Lett.*, **224**, 291–297.
91. Rice, S.A., Bieber, J., Chun, J.Y., Stacey, G. and Lampson, B.C. (1993) Diversity of retron elements in a population of rhizobia and other gram-negative bacteria. *J. Bacteriol.*, **175**, 4250–4254.
92. Dhundale, A.R., Furuichi, T., Inouye, S. and Inouye, M. (1985) Distribution of multicopy single-stranded-DNA among Myxobacteria and related species. *J. Bacteriol.*, **164**, 914–917.