

Computational Evidence that Array-Associated Reverse Transcriptase Loci Are Retron Systems with Multiple ncRNA Genes

Yuzhen Ye

yye@iu.edu

Indiana University Bloomington

Research Article

Keywords: reverse transcriptase (RT), array-associated reverse transcriptase (ART), retron, msr, msd, priming guanosine, noncoding RNA (ncRNA), RNA secondary structure

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11367685/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Computational Evidence that Array-Associated Reverse Transcriptase Loci Are Retron Systems with Multiple ncRNA Genes

Yuzhen Ye

Computer Science Department, Indiana University, Bloomington
yye@iu.edu

Abstract

Array-associated reverse transcriptases (ARTs) were recently discovered and described as a novel family of reverse transcriptases (RTs) associated with CRISPR-like arrays composed of repeated sequences interspersed with divergent regions. Here, we independently investigated ART-containing loci using established methods for RT classification and RNA secondary-structure analysis, together with two tools developed for this study: a scanner for candidate retron noncoding RNAs (ncRNAs; *msr/msd*) and an integrated viewer for comparative analysis of ncRNA sequence and structure. The ART-associated RTs examined here show their strongest similarity to retron-associated RTs. Moreover, multiple candidate ncRNA loci with features characteristic of retron *msr/msd* RNAs occur in tandem near the RT genes. Although individual ncRNA copies exhibit limited overall primary-sequence similarity, they retain related predicted secondary structures, including conservation of the priming guanosine and associated structural features, as well as sequence conservation concentrated in important stem-loop (SL) *msr* regions including SL1 and SL2. This organization—functionally relevant conserved regions interspersed with divergent sequences—may account for the superficial resemblance of these ncRNA arrays to CRISPR-like arrays. Together, these observations suggest that ARTs may represent an unusual retron architecture characterized by tandem arrays of structurally conserved *msr/msd*-like ncRNAs. Experimental characterization will be needed to determine whether these loci produce canonical or modified retron reverse-transcription products and to establish the functional significance of multiple ncRNA copies within a single retron locus.

Keywords: reverse transcriptase (RT); array-associated reverse transcriptase (ART); retron; *msr*; *msd*; priming guanosine; noncoding RNA (ncRNA); RNA secondary structure

1 Introduction

Reverse transcriptases (RTs) are widespread in bacterial and phage genomes and participate in a diverse collection of mobile genetic elements and defense systems. Among the best-characterized bacterial RT-containing systems are retrons, genetic elements that encode a specialized RT together with a structured noncoding RNA (ncRNA), conventionally described in terms of its *msr* and *msd* regions [1]. The ncRNA serves as both substrate and template for reverse transcription, producing a characteristic multicopy single-stranded DNA (msDNA) molecule covalently linked to RNA through an unusual 2′–5′ phosphodiester bond. Many retron systems function in antiphage defense [2, 3], and they have been repurposed for genome editing [4] and genome recombineering [5].

Although retron ncRNAs can be highly divergent at the primary-sequence level, they retain sequence and structural features required for recognition and reverse transcription by their cognate RTs [1]. Both the *msr* and *msd* regions exhibit substantial sequence diversity across retrons, although the *msr* retains key functional features, including the branching guanosine that serves as

the internal primer for reverse transcription and elements involved in RT recognition, whereas the *msd* is generally more sequence-variable [6]. Consequently, the structural organization of retron ncRNAs can provide informative signatures for identifying divergent retron systems. A characteristic feature is base pairing between complementary regions of the *msr/msd* transcript, which is critical for positioning the priming guanosine, and experimentally characterized retrons can exhibit substantial sequence divergence while retaining related RNA architectures [?]. Biochemical, mutational, and structural studies further demonstrate that specific ncRNA structural elements mediate interactions with retron RTs and are important for reverse transcription and retron function [7, 8]. For example, cryo-EM analysis revealed extensive recognition of stem-loop 2 (SL2) of the *msr*RNA by the RT thumb domain, deletion of which abolishes Retron-Eco7 antiphage defense [8]. In Retron-Eco8, SL1 and SL2 likewise have distinct but coordinated roles in phage sensing and effector activation, with disruption of either element impairing retron activation and antiphage defense [9]. Together, these observations indicate that functionally constrained ncRNA structures can remain conserved and detectable even when primary-sequence similarity has substantially eroded.

Recently, Anthropic reported the discovery of an unusual family of RT-containing systems in Jumbo phage genomes, identified through a large-scale AI-agent analysis of genomic sequence data [10]. Approximately 950 AI agents analyzed sequence data over a period of 21 hours, and one analysis identified a repeating pattern of DNA sequences adjacent to an unusual RT gene. Subsequent computational analysis and laboratory experiments led the investigators to describe these systems as *array-associated reverse transcriptases* (ARTs). ARTs were reported in bacteriophages and are characterized by an RT gene associated with an array of repeated sequences. The system was described as previously uncharacterized, while its primary biological function remains under investigation [10].

Here, we independently investigated ART-containing loci using myRT for RT identification and classification, together with RNA secondary-structure prediction, sequence alignment, and visualization approaches. We additionally developed a retron ncRNA scanner to identify candidate *msr/msd*-like sequences based on characteristic sequence and structural features, as well as an integrated interactive viewer for comparative analysis of sequence alignments and predicted RNA structures. Our analyses suggest that the ART-associated RTs examined here are most closely related to retron RTs and that their neighboring repeat arrays comprise multiple candidate retron-like ncRNAs. Although [10] also noted similarities between ART and retron systems, the model emerging from our analyses differs fundamentally from that proposed in their study.

2 Methods

2.1 Genome sequences

This study focused on the phage genomes and a Logan contig predicted to contain ARTs in [10]. Their genomes were obtained from NCBI: *Staphylococcus* phage vB_StaM_SA1 (GenBank ID: MW218148.1), MarsHill (MW248466.1), LY01 (OR836606.1), S6 (LC680885.1), and *Listeria* phage LPJP1 (MZ422438.1). The L0050 contig (ID: SRR12632059.735; assembled from shotgun metagenome sequencing of treated wastewater irrigated soil from orchard1) was obtained using Logan Search [11].

2.2 Identification and classification of reverse transcriptases

The online myRT [12] was used to identify and classify reverse transcriptases (RTs) encoded by these genomes. Given a genome sequence as input, myRT predicts protein-coding regions and searches

the predicted proteins against class-specific RT profile hidden Markov models (HMMs) to identify and classify RTs.

MEGA11 [13] was used to infer a neighbor-joining (NJ) tree of RT sequences based on a multiple sequence alignment (MSA) generated using MAFFT v7 with default parameters [14]. To assess the robustness of the phylogenetic relationships to different alignment and tree-inference methods, additional MSAs were generated using ClustalW and MUSCLE as implemented in MEGA11. Maximum-likelihood (ML) phylogenetic analysis was also performed in addition to the NJ analysis.

2.3 Identification of candidate retron ncRNAs

To identify candidate retron ncRNAs neighboring RT genes, we developed a sequence scanner that searches genomic DNA for sequence arrangements characteristic of retron *msr/msd* regions. The *msr* corresponds to the portion of the retron transcript that remains as RNA and contains the guanosine that primes reverse transcription, whereas the *msd* is reverse-transcribed to form the DNA component of msDNA. To reduce the recovery of unrelated structured sequences elsewhere in the genome, the search was restricted to genomic regions surrounding candidate RT genes.

Candidate ncRNAs were identified by searching for complementary sequence elements capable of forming the characteristic a1/a2 stem of retron RNA. The scanner required an a1/a2 complementary region of at least 8 bp, allowing at most one unpaired or mismatched nucleotide within the paired region. Candidate lengths were restricted to 100–300 nucleotides, based on the typical lengths of retron *msr/msd* regions reported by Khan et al [?]. Candidates were further evaluated based on the positions of two conserved guanosines relative to the ends of the a1/a2 stem: the guanosine that primes reverse transcription and the opposing guanosine on the complementary strand. Although both positions are reported to be conserved [9, 15], the two guanosines have distinct functional roles.

The nomenclature for the two complementary repeats, a1 and a2, is inconsistent across the retron literature. Throughout this study, we follow the nomenclature of early mechanistic studies of branched RNA-linked msDNA synthesis [16, 17], in which the repeat within the *msr* is designated a2 and the priming G (referred to as G* in the figures) is located at the end of a2. The complementary repeat on the *msd* side is therefore designated a1. Candidate sequences satisfying these criteria were extracted for subsequent secondary-structure prediction and comparative analysis.

2.4 RNA secondary-structure prediction

Candidate ncRNA sequences identified by the scanner were analyzed using [online RNAstructure](#) [18] (default parameters) to predict RNA secondary structures [18]. We used RNAstructure to predict structures from both single sequences and multiple sequences (the latter to detect common secondary structures). [Online RNAfold](#) [19] (default parameters) was also used as an alternative approach for structure prediction. Predicted structures were examined for features expected of retron-associated ncRNAs, including pairing between complementary sequence elements surrounding G* (i.e., a1 and a2) and conserved internal stem-loop structures (such as SL2).

Candidate ncRNAs identified from the same genome were compared using sequence-based and structure-based alignment approaches. Because retron ncRNAs may retain structural similarity despite extensive primary-sequence divergence, structural conservation was considered together with sequence conservation when evaluating candidate ncRNAs.

2.5 Multiple alignment of ncRNAs and comparative visualization

Candidate ncRNAs associated with each ART locus were aligned using MAFFT (v7, default parameters) [14] and RNAstructure [18]. Multiple-sequence alignments (MSAs) were inspected

together with predicted base-pairing interactions using a custom viewer developed for this study. The viewer displays the sequence alignment, an arc representation of predicted base pairs, and conventional two-dimensional RNA secondary structures, facilitating comparison of conserved sequence and structural elements among ncRNA copies. We note that the viewer was used primarily for exploratory comparison and figure preparation rather than as an independent prediction method.

Along with the MSA, the viewer provides visualization of per-column conservation. Three metrics are considered: frequency-based, entropy, and sum-of-pairs. Frequency measures the proportion of sequences sharing the most common nucleotide, treating gaps as mismatches; it is intuitive but ignores the distribution of minority characters. Shannon entropy considers the full character distribution including gaps as a fifth state. Sum-of-pairs computes the probability that two randomly drawn sequences agree at a position, with gaps penalized as mismatches; its quadratic weighting amplifies signals from strongly conserved positions while suppressing background noise, making it particularly effective for sequences with biased composition. The SA1 and MarsHill genomes are GC poor (27% and 25%, respectively), which can inflate frequency-based scores. We note that such compositional bias may also contribute to the difficulty of aligning the divergent regions of the ncRNAs predicted in these two genomes, which may otherwise resemble dissimilar spacers in CRISPR arrays [10].

2.6 AI-assisted development

Claude.ai was used to assist with the initial implementation of the ncRNA scanner and the viewer. Subsequent code modifications, selection of the sequence and structural criteria, evaluation of the resulting candidates and biological interpretation were performed by the author. Established bioinformatic and RNA-structure programs were used for the primary classification and structural analyses described above.

3 Results

3.1 ART-associated RTs are most similar to retron RTs

RTs of the retron type were predicted by myRT in *Staphylococcus* phages SA1, MarsHill, LY01, and S6, *Listeria* phage LPJP1, and contig L0050. As an example, Fig. 1A shows the genomic location of the RT gene in phage SA1. With the exception of the RTs from three *Staphylococcus* phages (MarsHill, LY01, and S6), which exhibit high sequence similarity to one another, these RTs share only moderate sequence similarity. For example, the SA1 and MarsHill RTs share approximately 60% amino-acid sequence identity over an alignment of 505 residues, whereas the SA1 and L0050 RTs share only 36% identity over 298 aligned residues. For comparison, the SA1 RT shares approximately 29% sequence identity with an RT from a retron of the Ec67 family in *Dickeya* sp. NCPPB 3274 (WP_152495267.1).

The assignment of these RTs as retron RTs was supported by HMMER searches. For example, the SA1, MarsHill, and L0050 RTs showed their strongest HMM matches to the RT_Bac_retron_II model, which represents RTs associated with type II retron systems [1]. The corresponding full-sequence HMMER [20] E-values were 4.2×10^{-35} , 2.8×10^{-37} , and 2.0×10^{-35} , respectively, providing strong support for their classification as retron-type RTs.

Phylogenetic analysis also placed these RTs together with RTs from known retron systems (Fig. 1B). To assess whether RTs cluster according to their functional roles rather than their source genomes, we included RTs associated with different systems, including diversity-generating retroelements (DGRs), CRISPR systems, and group II introns, from genomes encoding multiple

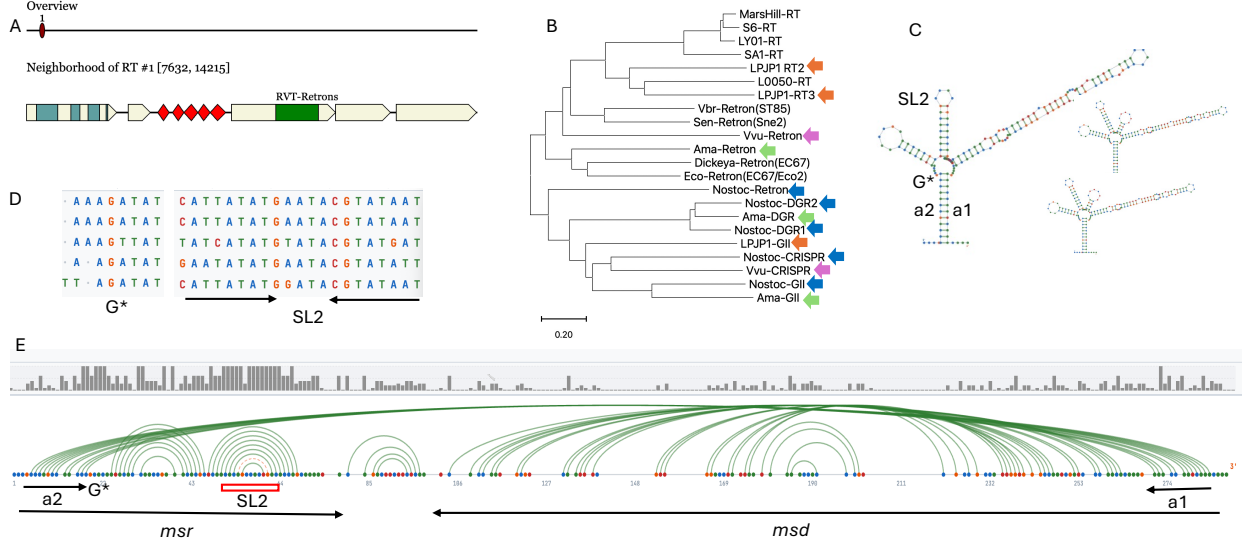


Figure 1: Putative retron system in *Staphylococcus* phage SA1 and comparison of RTs. (A) myRT prediction for SA1 showing the location of the RT locus (overview) and the neighborhood of the RT gene. In the plot, protein-coding genes are shown as arrows, and the five putative ncRNA genes are shown as diamonds. (B) NJ tree of RTs, including SA1 RT and RTs of other types (CRISPR, DGR and Group II introns). In this plot, arrows of the same color point to the multiple RTs found in the same genome. (C) Predicted structures of a SA1 ncRNA (between 10137-10320 in the genome) by different approaches: RNAstructure multilign, RNAstructure MaxExpect structure and RNAfold's minimum free energy prediction. Minor differences were observed among the predicted structures, but the main characteristics (priming guanosine and SL2) remained the same. (D) MSA of the five ncRNA sequences, showing only the most conserved regions, one around the priming G, and the other one involves SL2. (E) Column-wise conservation of the SA1 ncRNA sequences (sum-of-pairs, shown in bars) over the entire length, along with the arc representation of the secondary structure of the representative ncRNA. Important structural features are highlighted or marked, including *msr*, *msd*, a1 and a2 (containing priming G).

RTs. The resulting phylogeny showed that RTs generally grouped according to their functional associations rather than their genomes of origin; see the three RTs found in LPJP1 as an example (Fig 1A; orange arrows).

Additional RTs included Vbr-Retron(ST85) (WP_102381196.1), an RT from a retron of the St85 family in *Vibrio breoganii*; Dickeya-Retron(EC67) (WP_152495267.1), an RT from a retron of the Ec67 family in *Dickeya* sp. NCPPB 3274; Eco-Retron(EC67/Eco2) from *Escherichia coli* strain S10 (CP010229.1); Sen-Retron(Sen2) from *Salmonella enterica* subsp. *enterica* serovar Typhimurium str. LT2 (NC_003197.2); five RTs from *Nostoc* sp. PCC 7120 (NC_003272.1); three RTs from *Acaryochloris marina* MBIC11017 (CP000828.1); and two RTs from *Vibrio vulnificus* (BA000037.2). The clustering of ARTs with retron RTs was robust across different multiple-sequence alignment and phylogenetic inference methods.

These results indicate that the ART-associated RTs are closely related to known retron RTs and motivated examination of the surrounding genomic regions for retron-like ncRNAs.

3.2 ART-associated ncRNA arrays show conserved predicted structures

Analysis of the regions surrounding predicted RT genes in *Staphylococcus* phages SA1, MarsHill, LY01, and S6, *Listeria* phage LPJP1, and contig L0050 identified multiple candidate ncRNA sequences with features characteristic of retron *msr/msd* regions, including complementary sequence elements (a1 and a2) and an appropriately positioned candidate priming guanosine. However, rather than occurring as a single ncRNA locus, multiple candidate ncRNA genes were found adjacent to each RT, giving rise to the array-like organization originally described for ART systems.

The SA1 locus provides a representative example (Fig. 1). Five candidate ncRNA genes were identified adjacent to the RT gene (red diamonds in Fig. 1A). These putative ncRNA genes exhibit substantial primary-sequence divergence, with the *msd* regions being more divergent than the *msr* regions (Fig. 1E). Nevertheless, secondary-structure analysis (Fig. 1C) and sequence alignment (Fig. 1D) revealed a shared structural organization among the ncRNAs despite their limited overall sequence similarity. Sequence and structural conservation was concentrated in several features, including the priming G, SL1, and SL2 (Fig. 1E). Notably, SL2 (Fig. 1D) contains the repeat sequences previously reported by [10].

Similarly, multiple copies of putative *msr/msd* genes were identified in MarsHill (5 copies), LY01 (5 copies), S6 (6 copies), and LPJP1 (7 copies). These ncRNAs exhibit little overall primary-sequence similarity, with detectable local similarity largely confined to a stem-loop region. Despite this sequence divergence, all candidate ncRNAs retain characteristic retron structural features, including the priming guanosine and its associated structural context (see examples of predicted structures of these ncRNAs in Fig. 2).

Analysis of the L0050 contig identified an even larger array comprising 14 candidate ncRNA genes. All copies contain a conserved priming G and the stem structure that supports the priming G. In contrast to the ncRNAs from the *Staphylococcus* phages, in which SL2 is highly conserved among ncRNA copies, the most highly conserved region among the L0050 ncRNAs is located within SL1 (see a predicted structure and the MSA focusing on SL1 in Fig. 2).

Thus, the repetitive organization of the ART locus in the *Staphylococcus phage* and *Listeria phage* genomes and L0050 contig don't simply represent an array of identical sequence repeats or a CRISPR-like array of nearly identical repeats interspersed with spacers. Instead, the array can be interpreted as multiple related ncRNA genes in tandem, and the ncRNAs they encode retain conserved structural features characteristic of retron *msr/msd* despite sequence divergence.

4 Discussion

Our analyses provide several independent lines of computational evidence connecting ART systems to retrons. First, the ART-associated RTs examined here show the strongest similarity to retron-associated RT profiles. Second, the genomic regions surrounding these RTs contain multiple sequences with characteristics expected of retron *msr/msd* ncRNAs. Third, despite relatively limited primary-sequence conservation, these candidate ncRNAs retain related predicted secondary structures, including conservation concentrated in the SL2 region. Taken together, these observations suggest that ARTs are evolutionarily related to retrons and may represent an unusual retron architecture in which multiple ncRNA genes occur in tandem adjacent to a retron-like RT.

4.1 Why are multiple ncRNAs associated with a single RT?

The presence of multiple candidate ncRNA genes associated with a single RT raises an important question about the functional organization of these systems. Canonical descriptions of retrons

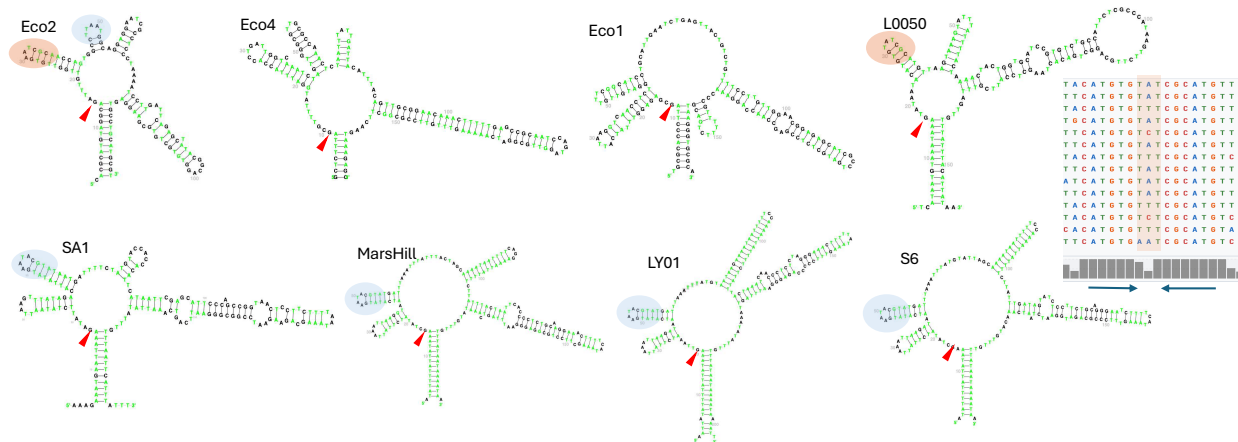


Figure 2: Comparison of known retron ncRNAs and those predicted in this study. The 132-nt ncRNA of the characterized Eco2 retron from *E. coli* NCTC 86 was extracted from PDB 9I2F [3]. Its predicted secondary structure is shown with the two stem-loops important for Eco2 function [3] highlighted in orange (SL1) and blue (SL2). Also shown are the experimentally characterized ncRNAs of retrons Eco4 and Eco1 [21], as well as the predicted ncRNAs of SA1 (positions 10137–10320 in the corresponding genome), MarsHill (214166–214361), LY01 (75543–75745), and S6 (166634–166828). The 14 CRISPR-like repeats reported in [10] correspond to the SL1 regions of the 14 ncRNAs predicted from the L0050 contig (Logan, SRR12632059.735); these repeats are shown alongside the predicted structure of one representative ncRNA. All RNA secondary structures shown in this figure were predicted using RNAstructure (MaxExpect) and visualized using R2DT [22]. Red arrows indicate the priming G.

typically center on a cognate RT-*msr/msd* pair, and biochemical studies have demonstrated specificity between retron RTs and their associated ncRNAs. The occurrence of several related but non-identical ncRNAs adjacent to an ART RT therefore suggests several possible models.

The ncRNA copies could be functionally redundant, could produce distinct reverse-transcription products from a common RT, or could constitute a diversified repertoire involved in phage biology or interactions with host defense systems. Alternatively, some array members could represent nonfunctional evolutionary remnants. The computational analyses presented here cannot distinguish among these possibilities. Determining whether multiple array members are transcribed and recognized by the same RT will be important for understanding what distinguishes these systems from more conventional retrons.

The evolutionary origin of the arrays is also unclear. Tandem duplication followed by sequence divergence could generate multiple related ncRNA genes while retaining structural constraints imposed by recognition by a shared RT. If so, the regions showing the strongest conservation may identify RNA features under functional constraint. The pronounced conservation observed around the predicted SL2 region is consistent with this hypothesis, although experimental analysis will be required to determine whether this region has a common functional role across the ART-associated ncRNAs.

4.2 Experimental tests of the retron model

Our conclusions are based primarily on RT classification, sequence analysis, and predicted RNA secondary structures and therefore require experimental validation. Structural similarity alone

does not demonstrate that the candidate ncRNAs are processed as retron substrates or that the ART-associated RT produces msDNA.

Several experiments could directly test the retron model. Expression of an ART RT together with individual candidate ncRNAs could determine whether they produce reverse-transcribed DNA products. Mapping the resulting products could establish whether reverse transcription initiates from the predicted priming guanosine and whether the characteristic retron RNA–DNA linkage is formed. Mutational analysis of the predicted complementary arms, priming guanosine, and conserved SL2 region could further test whether the structural features identified computationally are required for activity. Finally, testing individual versus combined members of the ncRNA array could determine whether a single RT recognizes multiple array members and whether different copies generate distinct products.

Demonstration of an msDNA-like product with the characteristic retron linkage would provide particularly strong evidence for the retron interpretation. Conversely, failure to observe canonical msDNA synthesis would not necessarily exclude evolutionary relatedness to retrons but could indicate that ART systems have diverged toward a distinct biochemical function.

5 Conclusions

Our analyses suggest that ART-associated RTs are evolutionarily related to retron RTs and that the neighboring arrays contain multiple structurally related, retron-like ncRNAs. These results support a model in which at least the ART systems examined here represent an unusual retron architecture characterized by tandem arrays of *msr/msd*-like ncRNA genes. The functional significance of these arrays remains unknown. Experimental characterization of the predicted ncRNAs, their reverse-transcription products, and their interactions with the associated RTs will be necessary to determine how these systems relate mechanistically to canonical retrons and what biological advantage is conferred by maintaining multiple ncRNA copies.

Data and Code Availability

The ncRNA scanner and interactive viewer are available in the GitHub repository [retron2](#). Analysis results (including sequences of predicted RTs and ncRNA genes) and their visualizations are available at [myRT website](#).

References

- [1] Mario Rodríguez Mestre, Alejandro González-Delgado, Luis I Gutiérrez-Rus, Francisco Martínez-Abarca, and Nicolás Toro. Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems. *Nucleic acids research*, 48(22):12632–12647, 2020.
- [2] Adi Millman, Aude Bernheim, Avigail Stokar-Avihail, Taya Fedorenko, Maya Voichek, Azita Leavitt, Yaara Oppenheimer-Shaanan, and Rotem Sorek. Bacterial retrons function in anti-phage defense. *Cell*, 183(6):1551–1561, 2020.
- [3] Monika Jasnauskaitė, Jonas Juozapaitis, Toma Liegutė, Rokas Grigaitis, Aistė Skorupskaitė, W Steinchen, Algirdas Mikšys, Lidija Truncaitė, K Kazlauskaitė, MF Torres Jiménez, et al. Structure and mechanism of antiphage retron eco2. *Nature Structural & Molecular Biology*, 33(2):330–340, 2026.

- [4] Kazuo L Nakamura, Karen Zhang, Mario R Mestre, Matías Rojas-Montero, and Seth L Shipman. New retron systems from environmental bacteria identify triggers of anti-phage defense and expand tools for genome editing. *PLoS Biology*, 23(10):e3003042, 2025.
- [5] Alejandro González-Delgado, Laura Bonillo-Lopez, Milo S Johnson, Nastassia Knödlseeder, Ching-Chung Ko, Yassir Lekbach, Jee-Hwan Oh, Hema Selvakumar, Michael C Wold, Zihan Yu, et al. Genome editing of phylogenetically distinct bacteria using cross-species retron-mediated recombineering. *Nature Biotechnology*, pages 1–13, 2026.
- [6] Anna J Simon, Andrew D Ellington, and Ilya J Finkelstein. Retrons and their applications in genome engineering. *Nucleic acids research*, 47(21):11007–11019, 2019.
- [7] Arturo Carabias, Sarah Camara-Wilpert, Mario Rodríguez Mestre, Blanca López-Méndez, Ivo A Hendriks, Ruiliang Zhao, Tillmann Pape, Anders Fuglsang, Sean Hoi-Ching Luk, Michael L Nielsen, et al. Retron-eco1 assembles nad⁺-hydrolyzing filaments that provide immunity against bacteriophages. *Molecular cell*, 84(11):2185–2202, 2024.
- [8] Yanjing Wang, Zeyuan Guan, Chen Wang, Yangfan Nie, Yibei Chen, Zhaoyang Qian, Yongqing Cui, Han Xu, Qiang Wang, Fen Zhao, et al. Cryo-em structures of escherichia coli ec86 retron complexes reveal architecture and defence mechanism. *Nature microbiology*, 7(9):1480–1489, 2022.
- [9] Chao-Guang Ji, Zhuolin Li, Xin-Yang Wei, Yuelong Li, Jun-Tao Zhang, Xueyan Liu, and Ning Jia. Mechanistic insights into activation of bacterial retron-eco8 immunity by phage protein ssb. *Nature Communications*, 17(1):7374, 2026.
- [10] Peter H. Yoon, Januka S. Athukoralage, Emmanuel Ameisen, Eric Kauderer-Abrams, Nicholas T. Perry, and Matthew G. Durrant. Autonomous ai agents discover reverse transcriptases with tandem repeat arrays (preprint: <https://www-cdn.anthropic.com/22573675ada52a8ca8a97a1a4b4326b2f208a071.pdf>), 2026.
- [11] Rayan Chikhi, Téo Lemane, Raphaël Loll-Krippelbein, Mercè Montoliu-Nerin, Brice Raffestin, Antonio Pedro Camargo, Carson J Miller, Mateus Bernabe Fiamenghi, Daniel Paiva Agostinho, Sina Majidian, et al. Logan: planetary-scale genome assembly surveys life’s diversity. *Biorxiv*, pages 2024–07, 2025.
- [12] Fatemeh Sharifi and Yuzhen Ye. Identification and classification of reverse transcriptases in bacterial genomes and metagenomes. *Nucleic acids research*, 50(5):e29–e29, 2022.
- [13] Koichiro Tamura, Glen Stecher, and Sudhir Kumar. Mega11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, 38(7):3022–3027, 2021.
- [14] Kazutaka Katoh, John Rozewicki, and Kazunori D Yamada. Mafft online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*, 20(4):1160–1166, 2019.
- [15] Junichiro Ishikawa, Kanta Yoneyama, Aa Haeruman Azam, Asuteka Nagao, Yoshihisa Mitsuda, Ren Nakazaki, Kotaro Chihara, Masahiro Hiraizumi, Keitaro Yamashita, Tsutomu Suzuki, et al. Structural mechanism of the retron-eco7 anti-phage defense system. *Nature Communications*, 16(1):10821, 2025.

- [16] Tadashi Shimamoto, Hideki Kawanishi, Tomofusa Tsuchiya, Sumiko Inouye, and Masayori Inouye. In vitro synthesis of multicopy single-stranded dna, using separate primer and template rnas, by escherichia coli reverse transcriptase. *Journal of bacteriology*, 180(11):2999–3002, 1998.
- [17] Mei-Yin Hsu, Susan G Eagle, Masayori Inouye, and Sumiko Inouye. Cell-free synthesis of the branched rna-linked msdna from retron-ec67 of escherichia coli. *Journal of Biological Chemistry*, 267(20):13823–13829, 1992.
- [18] Stanislav Bellaousov, Jessica S Reuter, Matthew G Seetin, and David H Mathews. Rnastructure: web servers for rna secondary structure prediction and analysis. *Nucleic acids research*, 41(W1):W471–W474, 2013.
- [19] Ivo L Hofacker, Walter Fontana, Peter F Stadler, L Sebastian Bonhoeffer, Manfred Tacker, and Peter Schuster. Fast folding and comparison of rna secondary structures. *Monatshefte für Chemie/Chemical Monthly*, 125(2):167–188, 1994.
- [20] Sean R Eddy. Accelerated profile hmm searches. *PLoS computational biology*, 7(10):e1002195, 2011.
- [21] Asim G Khan, Matías Rojas-Montero, Alejandro González-Delgado, Santiago C Lopez, Rebecca F Fang, Kate D Crawford, and Seth L Shipman. An experimental census of retrons for dna production and genome editing. *Nature biotechnology*, 43(6):914–922, 2025.
- [22] Holly McCann, Caeden D Meade, Loren Dean Williams, Anton S Petrov, Philip Z Johnson, Anne E Simon, David Hoksza, Eric P Nawrocki, Patricia P Chan, Todd M Lowe, et al. R2dt: a comprehensive platform for visualizing rna secondary structure. *Nucleic acids research*, 53(4):gkaf032, 2025.