To be mobile or not: The variety of reverse transcriptases and their recruitment by host genomes

Irina R. Arkhipova and Irina A. Yushenova

Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543, USA. E-mail: iarkhipova@mbl.edu; iyushenova@mbl.edu

Keywords: Reverse transcription, RNA-dependent DNA polymerase, telomerase reverse transcriptase.

Abstract

Reverse transcriptases (RT), or RNA-dependent DNA polymerases, are unorthodox enzymes that originally added a new angle to the conventional view of the unidirectional flow of genetic information in the cell from DNA to RNA to protein. First discovered in vertebrate retroviruses, RTs were since re-discovered in most eukaryotes, bacteria, and archaea, spanning essentially all domains of life. For retroviruses, RTs provide the ability to copy the RNA genome into DNA for subsequent incorporation into the host genome, which is essential for their replication and survival. In cellular organisms, most RT sequences originate from retrotransposons, the type of self-replicating genetic elements that rely on reverse transcription to copy and paste their sequences into new genomic locations. Some retroelements, however, can undergo domestication, eventually becoming a valuable addition to the overall repertoire of cellular enzymes. They can be beneficial yet accessory, like the diversity-generating elements, or even essential, like the telomerase reverse transcriptases. Nowadays, ever-increasing numbers of domesticated RT-carrying genetic elements are being discovered. It may be argued that domesticated RTs and reverse transcription in general is more widespread in cellular organisms than previously thought, and that many important cellular functions, such as chromosome end maintenance, may evolve from an originally selfish process of converting RNA into DNA.

Introduction

At the dawn of molecular biology, when little was known about the underlying molecular nature of biological phenomena, numerous theoretical papers were attempting to foresee future discoveries and to make viable predictions regarding molecular explanations of fundamental genetic concepts. Notably, only a relatively small fraction of such papers withstood the test of time and the eventual experimental scrutiny that followed in the years to come. Among such visionary papers, the theoretical prediction by Alexey Olovnikov of terminal DNA under-replication in linear chromosomes and of the specialized enzyme that could overcome this problem [1, 2] occupies a well-deserved place. While simultaneous recognition of the end-replication problem is also credited to the paper by James Watson [3], its focus on phage DNA avoided the requirement for a specialized polymerase, shifting the emphasis on end-processing nucleases instead.

The Nobel prize-winning discovery of telomerase, the specialized polymerase which can add simple repetitive sequences to the ends of linear chromosomes to compensate for terminal DNA loss after each round of replication, has in turn followed a long and winding path. In the initial report by Greider and Blackburn, the discovered *Tetrahymena* enzyme was designated as a terminal transferase [4], because the detected activity was adding tandem repeats onto telomeric primers without an apparent template. An associated RNA template, however, was subsequently identified as an integral component of the ribonucleoprotein holoenzyme, providing experimental evidence in support of RNA-dependent DNA

synthesis [5], although it was still considered premature to classify the telomerase enzyme as an authentic reverse transcriptase.

The process of DNA synthesis that uses RNA as a template is universally recognized under the term "reverse transcription", and the corresponding enzyme that can perform this reaction bears the name "reverse transcriptase" (RT). Its experimental discovery by Temin and Baltimore more than 50 years ago [6, 7], which was also recognized by a Nobel prize, was similarly preceded by Howard Temin's conceptualization of DNA synthesis on viral RNA template, known as "the provirus hypothesis" [8]. Little did they know that in addition to discovering the reverse flow of genetic information from viral RNA to DNA, they also provided the foundation for the discovery of self-replicating movable genetic elements and for eventual realization that some of the accessory or even essential host functions can be taken over by the descendants of such mobile elements. Remarkably, RTs were discovered approximately at the time when the chromosome end under-replication problem first came to light (Fig. 1).

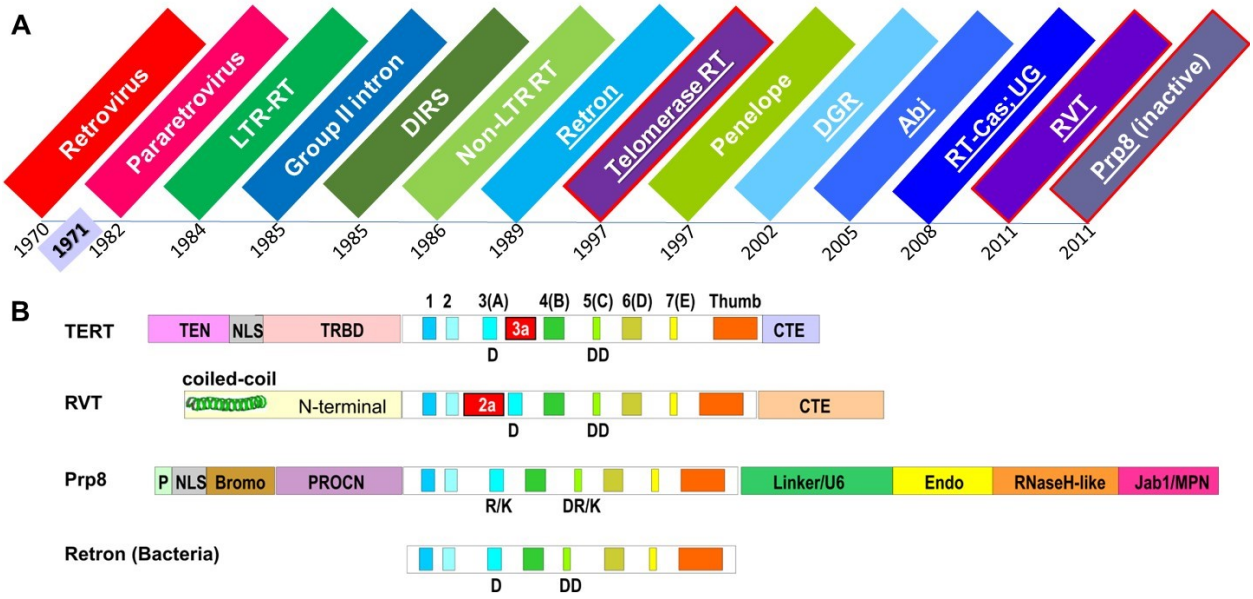

Figure 1*. The main types of reverse transcriptases (RT) from the three domains of life. (**A**) Chronology of RT discovery. The main RT types described in the text are colored as follows: viral RTs, shades of red; RTs of eukaryotic mobile elements, shades of green; prokaryotic RTs, shades of blue; domesticated eukaryotic RTs, shades of purple. Domesticated RTs are underlined. The years correspond to the first reports of identification of homology to the RT catalytic core. The year 1971 marks the first report of the chromosome end under-replication problem. (**B**) Examples of structural organization of domesticated eukaryotic RTs. Bacterial retrons are included for comparison. The centrally positioned RT catalytic core is represented by the seven conserved motifs separated by spacers of variable length, with distinctively long insertion loops 2a and 3a (also called IFD) marked in red. The D..DD active site residues and their non-catalytic replacements are indicated. Additional domains on either side of the RT core and thumb are as follows: TEN, telomerase essential N-terminal domain; TRBD, telomerase RNA binding domain; CTE, C-terminal extension; P, polyproline stretch; NLS, nuclear localization signal; Bromo, bromodomain; PROCN, PRO8 central domain; Endo, endonuclease-like; Jab1/MPN, putative deubiquitinase-like domain. The scale is approximate. Domain composition is compiled from refs. [55, 57, 59].*

Evolution of approaches to retroelement discovery

Since their discovery in retroviruses, RT diversity underwent an amazing expansion from purely viral constituents to a staggering variety of structural and functional roles in eukaryotic and prokaryotic hosts (Fig. 1A). After early advances in the field of virology, which led to further discovery of reverse transcription in the replicative cycles of hepadnaviruses and caulimoviruses (collectively named pararetroviruses [9]) and were facilitated by the availability of methods for virus isolation and biochemical RT assays, the discovery potential soon shifted towards detection of sequence homologies, spurred by the advent of sequencing technologies and the landmark identification of common amino acid sequence motifs in the catalytic core of DNA polymerases from reverse-transcribing viruses [10]. Since then, the search for the aspartates forming the D..DD catalytic triad at the RT active site has quickly become an integral part of identification of novel RTs. In the RT discovery timeline (Fig. 1A), the underlying publications in which the characteristic RT residues were first identified were given priority in comparison to those reporting initial biochemical detection of RNA-dependent DNA polymerization. This is because proper experimental validation of RT activity should inevitably include site-directed mutagenesis of the active site residues, present in two of the seven conserved motifs defining the RT catalytic core (Fig. 1B).

The first half of the timeline, prior to 1990's, is represented mainly by RTs from various types of viruses and mobile genetic elements. Indeed, multicopy transposable elements were one of the first components of eukaryotic genomes to be cloned molecularly [11, 12], along with other actively transcribed multicopy genes such as ribosomal DNA repeat units or histone gene clusters [13, 14]. The overall structural similarity between LTR-retrotransposons and retroviruses immediately became apparent upon their cloning from Drosophila and yeast [15]. However, the definitive proof of their close relationship to retroviruses came from analysis of their complete nucleotide sequences identifying the coding capacity for the RT enzyme [16, 17]. Furthermore, characteristic blocks of homology to the RT conserved motifs were soon identified not only in retrovirus-like transposable elements, but also in fungal mitochondrial group II mobile introns and other types of multicopy eukaryotic transposons, such as DIRS and LINE-like retrotransposons [18-21]. To conclude the first two decades of RT research, the existence of RTs in bacteria was reported in the form of retrons, multicopy extrachromosomal DNA-RNA chimeric molecules connected through a 2'-5' branchpoint [22, 23].

The next temporal phase in RT discovery, while also relying on detection of sequence homologies, was dominated by RTs present in lower copy numbers, most of which do not belong to transposable elements, but instead represent single-copy host genes (Fig. 1A, underlined). In fact, the currently known eukaryotic retrotransposon diversity has not expanded since the discovery of *Penelope*-like retroelements (PLEs) [24]. The first and most prominent case of RT domestication in eukaryotes emerged with the proof that telomerase represents a *bona fide* RT. Connecting the RT activity with the corresponding enzyme took a lot of time and effort, with mis-identifications along the way [25], but the ultimate success in identifying the telomerase catalytic subunit as an RT came with identification of the conserved motifs in the fingers and palm RT domains, validated by loss of activity upon site-directed mutagenesis of the three invariant catalytic aspartates [26]. Thus, a single-copy RT gene present in nearly all eukaryotic species was found to be responsible for an essential host function of elongating the ends of linear chromosomes to counteract terminal DNA loss from under-replication, or marginotomy, as it was originally named by Olovnikov [27]. In the following sections, our aim is to briefly characterize the

RTs which belong to mobile genetic elements, and to compare to those which are domesticated and accordingly non-mobile.

Eukaryotic mobile elements: Retroviruses, pararetroviruses, retrotransposons

To understand and compare the properties of viral and mobile RTs, we need to consider the architectural composition of conserved domains that occur in combination with RT, as well as the adjacent gene content within the mobilizable unit (Fig. 2). Interestingly, ***retroviruses***, the discovery of which opened the era of RT research, turned out to be strikingly similar to ***LTR-retrotransposons***, discovered over a decade later, in their gene content, organization, and replication cycle, pointing at their common evolutionary ancestry [16, 17, 28]. RTs of ***hepadnaviruses*** can be broadly assigned to the base of the viral/LTR branch of eukaryotic RTs, which harbors the C-terminal RNase H domain to ensure replication in the cytoplasm, avoiding the need to employ host nuclear RNase H enzymes for destruction of RNA in the DNA-RNA hybrid (Fig. 2). Even more unusual is the case of ***caulimoviruses***, the RT of which is closely related to that of Metaviridae (aka Ty3/mdg4(gypsy)-like LTR retrotransposons), such that their ancestry is most likely of hybrid nature, resulting from RT capture by a DNA virus [29]. The Ty1/copia-like LTR retrotransposons (Pseudoviridae) conform to the general LTR structure, but show a different domain order. All retrovirus-like elements comprising the taxonomic order Ortervirales (Retroviridae, Metaviridae, Pseudoviridae and Belpaoviridae) [29] are mobilized with the aid of the integrase (IN), which is responsible for insertion of a cDNA copy into new chromosomal locations. A distinct group called ***DIRS elements*** mobilizes by using tyrosine recombinase (YR) instead of IN.

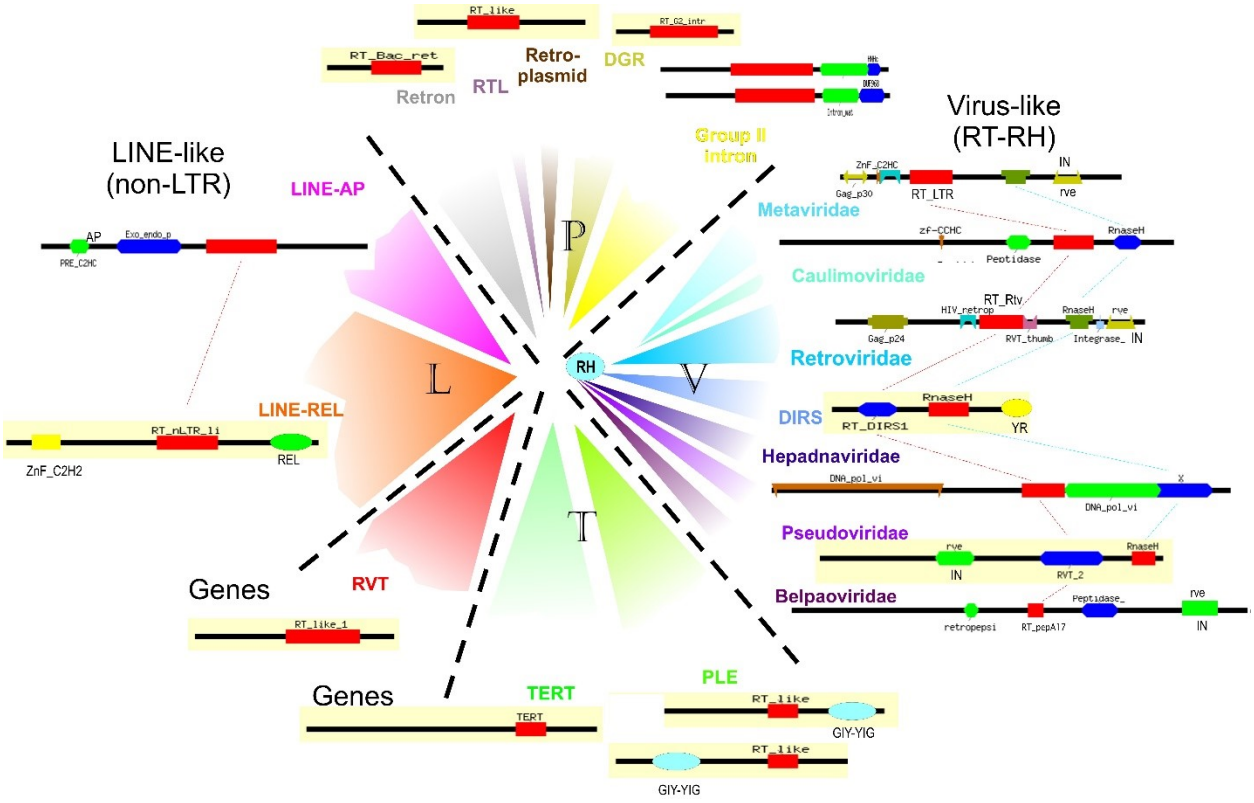

Figure 2*. Domain architecture of the major RT types described in the text. For each type, a typical architecture is presented as revealed by the CDART tool at NCBI [63]. Domain designation is according to*

the NCBI conserved domain database (CDD) [64]. The colors are assigned by the CDART tool dynamically rather than following each domain specifically; to facilitate homology tracing, the RT and RNaseH (RH) domains are connected with a dashed line. The circular arrangement follows the phylogenetic groupings in the center from ref. [55], with letters P, V, T and L corresponding to prokaryotic, virus-like, telomerase-like and LINE-like retroelements; RVT genes form a separate group which has no designation yet. Mobile elements contain six different types of associated nucleases/phosphotransferases mentioned in the text: IN, AP, REL, YR, GIY-YIG, HNH. Virus-like elements are named according to ICTV classification [29]. Domesticated eukaryotic RTs (TERT, RVT) are designated as Genes.

Non-LTR (or LINE-like) retrotransposons mobilize without producing a cytoplasmic cDNA intermediate: their RT uses the target-primed reverse transcription (TPRT) mechanism to synthesize cDNA directly at the chromosomal integration site nicked by one of the two different types of associated endonuclease (EN), either AP-like or REL-like. Finally, RTs of ***Penelope-like elements*** employ yet another EN type (GIY-YIG) for mobilization, bringing the number of retrotransposon-associated endonuclease types to five. A more detailed recent description of retromobility mechanisms can be found in ref. [30].

Prokaryotic mobile elements: Group II introns, retroplasmids

Group II introns (G2I) are self-splicing retroelements found in bacteria, some archaea, and eukaryotic organelles [31]. First discovered in fungal mitochondria, they were shown to possess the same structural organization in bacteria and archaea, and are widely regarded as evolutionary precursors to eukaryotic spliceosomal introns. Their retromobility is ensured by the combined action of the catalytically active RNA, which functions as a ribozyme in the self-splicing and reverse-splicing reactions, and the intron-encoded RT, which synthesizes a cDNA copy of the intron RNA at the target site, using the TPRT mechanism.

Retroplasmids were found in fungal mitochondria [32] and for a long time served as a model system to study the unconventional priming modes by reverse transcriptases (protein priming, when RT uses the hydroxyl group of tyrosine or serine residues for priming, or de novo RT initiation, which does not use any primer at all). Their distribution is still quite limited, as there are only a few dozen fungal species harboring them, out of hundreds of sequenced fungal genomes. As extrachromosomal entities, they are not expected to undergo integration, but technically form part of the mobilome due to their ability to replicate autonomously.

Non-mobile retroelements in Bacteria and Archaea: Retrons, DGRs, Abi/UG, Cas-associated, G2I-like

Retrons are peculiar domesticated bacterial elements composed of covalently linked RNA and multicopy single-stranded DNA (msDNA) in a single branched molecule connected by a 2'-5' phosphodiester linkage [22, 23]. Each retron module encodes an RT protein sequence, a non-coding RNA which is reverse-transcribed by the RT to form the chimeric single-stranded DNA/RNA molecules, and an effector gene needed for anti-phage activity. Despite being the first prokaryotic non-mobile retroelements discovered over 30 years ago, the cellular function of retrons was elucidated only in 2020 [33-35]. Retrons confer host defense against a broad range of phages via abortive infection and subsequent cell death. They are widespread in bacteria, being one of the main components of bacterial immune systems. However, the exact mechanisms by which they confer phage resistance via reverse transcription are still unknown. The co-occurrence of RT in tripartite modules with template RNA and a variety of putative effector genes suggests their direct interaction in eliciting anti-phage response [36]. Indeed, such interaction was

observed in a complex between RT, its cognate msDNA, and the linked effector nucleoside deoxyribosyltransferase [37].

Diversity-generating retroelements (DGRs) are non-mobile RTs that diversify adjacent target DNA sequences in bacteria, archaea, and viruses [38, 39]. Despite being non-essential retroelements, DGRs are nevertheless beneficial for their hosts. In the best-described model system, DGRs generate diversity in the C-terminal variable region of target protein gene (*mtd*) of the *Bordetella pertussis* bacteriophage BPP-1. The resulting hypervariability in the phage tail protein, the region that contacts the bacterial cell during infection, allows the phage to infect bacterial cells with altered surface receptors. By utilizing error-prone reverse transcription, DGRs help to increase diversity in gene products, especially those involved in ligand-binding and host attachment. It is still a mystery how the adenine specificity of targeted hypermutagenesis is accomplished. Moreover, inspection of adjacent genes in DGR modules suggests that hypervariability targets may not be limited to tropism switching and surface display [40, 41].

Abortive infection systems (Abi), represented by AbiA, AbiK, and Abi-P2, are bacterial retroelements that serve to protect certain bacteria from phage infections. These genes are only found in some Bacilli (mostly in *Lactococcus lactis*) genomes as plasmid-encoded genes (AbiA and AbiK), and on P2-like prophages in *Escherichia coli* (Abi-P2). While their detailed mechanism of action is still unknown, Abi proteins are required for blocking phage replication followed by programmed cell death or phage exclusion [42, 43]. Interestingly, the AbiK protein was shown to perform non-templated DNA polymerization *in vitro* and is covalently attached to DNA, which is indicative of protein priming [44]. Thus, Abi represent another, besides retrons, type of active RT which confers advantage to a subset of bacteria when attacked by phages. Of note, AbiP2 and AbiK RTs are exceptional in forming compact trimers or hexamers in solution, as well as in lacking the RT thumb domain, which is replaced by the all-helical domain composed of HEAT repeats [45, 46]. A substantial proportion of the so-called unknown groups (***UG***) [47], some of which were independently called DRT (defense RT) [33], were reported in earlier surveys as unassignable to a specific RT type, but were later found to be related to Abi RTs and to play a role in antiphage defense, with enrichment in the so-called defense islands, which contain a variety of other genes providing protection against invading foreign DNA [33, 45].

RT-Cas: RT domains were found near CRISPR-associated genes or even fused to Cas proteins [48-50]. Potentially, these RTs can confer bacterial immunity by performing cDNA synthesis on RNA from bacteriophages, and were indeed shown to mediate heritable acquisition of short sequence segments (spacers) from foreign RNA elements [51]. Fusion to Cas proteins is not necessary, although it allows more efficient cooperation of the interacting domains [52]. These RTs are not monophyletic, having been co-opted into CRISPR-Cas systems from several bacterial RT lineages [50].

Group II intron-like RTs (G2L), a heterogeneous group of non-mobile RTs that share sequence similarity with G2I but lack the ribozyme moiety, was first described in [48]. Recently, it was found that G2L RT from *Pseudomonas aeruginosa* (G2L4 RT) is involved in translesion DNA synthesis and double-strand break repair via microhomology-mediated end-joining (MMEJ) [53]. Interestingly, the substitution of YADD to YIDD in the G2L4 RT active site is responsible for a shift towards performing MMEJ instead of primer extension, which is characteristic for canonical G2I RTs with YADD at the catalytic site. Nevertheless, a canonical G2I RT was also capable of performing DNA repair.

Non-mobile eukaryotic RTs and their derivatives: Telomerase, RVT, PRP8

Telomerase reverse transcriptase (TERT), as described above, is undoubtedly the most well-known RT with a crucial cellular function. On top of the main function of maintaining the length of linear chromosomes, it has well-described roles in aging, cancer, and other human diseases (aplastic anemia, Cri du chat syndrome, Dyskeratosis congenita, etc.). Multiple approaches are being developed to target active telomerase and the associated TERT RNA template pharmaceutically in the context of anti-cancer therapy and age-related diseases (recently compiled in [54]).

Reverse transcriptase-related genes (rvt) are the most recently discovered type of domesticated eukaryotic RTs widespread in fungi and sporadically occurring in selected plants, protists, and invertebrates [55]. Strikingly, these genes are present in both prokaryotes and eukaryotes, in contrast to all other RT types. Notably, RVTs from all bacterial phyla form a monophyletic group, suggesting that they were not horizontally transferred from eukaryotes, but may have been present in Bacteria prior to eukaryogenesis [56]. *Rvt* genes encode active RT-like proteins that in fungi can polymerize both dNTPs and NTPs. RVT proteins are also capable of protein priming. While biological function of *rvt* genes is not yet fully understood, they are clearly preserved by natural selection, indicating their importance for host cells. These genes are strongly activated by starvation and certain antibiotics in fungi, suggesting their involvement in response to these agents [55].

Pre-mRNA-processing factor 8 (Prp8) is an unusual domesticated RT derivative that lost two out of three catalytic aspartates, thereby losing the ability to polymerize nucleotides [57]. Yet, Prp8 is an essential part of eukaryotic spliceosome regulating its assembly and conformation during pre-mRNA splicing [58]. The RT moiety of Prp8 was proposed to originate from mobile group II introns [59], giving us one more example of how during evolution selfish retrotransposons can give rise to essential components of eukaryotic cells, in this case as a structural element which comprises the central U5-snRNA-binding part of a large multi-domain protein (Fig. 1B). The lack of catalytic residues and very high sequence conservation due to evolutionary constraints imposed by spliceosome function impedes unambiguous phylogenetic placement of this RT-derived domain, but its origin undoubtedly dates back to the last common ancestor of all eukaryotes.

Concluding remarks

From the RT descriptions summarized above, it is easy to note that the RT types discovered in earlier years generally originated from abundant, high-copy-number sources – initially from viruses, and subsequently from cellular multicopy mobile genetic elements: from LTR, DIRS and non-LTR retrotransposons in eukaryotes, to prokaryotic mobile group II introns and retroplasmids, and to retrons producing abundant branched DNA-RNA molecules in bacterial cells. Retromobility is typically conferred by a specific type of endonuclease associated with each mobile element, providing the means for intrachromosomal insertion of a cDNA copy. At the initial stages, many eukaryotic TEs were identified by their ability to cause insertional mutations with visible phenotypes in strains experiencing transposition of multicopy elements [60]. It is now clear that RTs can perform a large variety of functions besides their role in proliferation of selfish genetic elements. We argue that the diversity of domesticated RTs has been grossly underestimated and their role has been substantially undervalued, with plenty of opportunities existing for RT recruitment by the host cells despite their overall non-essential nature and patchy distribution. It is not surprising that sometimes it may take a long time, even decades, from initial identification of an element to the proper assignment of a host function, if the selective advantage to the host is conditional. The telomerase RT, a single-copy gene, represents a notable exception in being

ubiquitously present throughout eukaryotes, and the revelation that it encodes a specialized RT, i.e. an enzyme previously thought to be characteristic only of viruses and mobile elements, has truly revolutionized the field [26]. Still, even the critical function of telomere maintenance can be supported by independent backup pathways [61].

It is worth emphasizing that RT domestication in eukaryotes is invariably associated with the appearance of additional functional domains that would prevent it from spurious cDNA synthesis using random primer/template combinations. Generally, synthesis of cDNA copies on random host RNA templates is not expected to benefit the host cell and should be prevented. The most straightforward way is to eliminate catalytic activity by replacing active site residues, as in Prp8. Another option is to change the configuration of the active site by inserting additional structural loops, as in RVT genes. Finally, TERTs have achieved strict substrate specificity via a high degree of specialization towards an unlinked highly structured RNA (called TER or TR), which contains a short reverse-complement of the telomeric repeat unit serving as a template, and interacts specifically with the TRBD domain to perform highly processive DNA synthesis by target-primed reverse transcription (TPRT) off the 3'-ends of exposed short G-rich tandem repeats at the ends of linear chromosomes [62]. It is fascinating to realize that the specialized enzyme predicted to overcome terminal DNA loss and to preserve chromosome integrity takes its origins from mobile elements initially poised to disrupt chromosomal stability.

Ethics declarations. The authors declare that they have no conflicts of interest. This article contains no description of studies involving human subjects or animals performed by any of the authors.

Funding. The work in the laboratory is funded by grants from the U.S. National Institutes of Health to I.A. (R01GM111917) and the U.S. National Science Foundation to I.A. and I.Y. (MCB-2139001, MCB-2326038).

Acknowledgments. This contribution honors the memory of Alexey Olovnikov, in lieu of a planned in-person discussion, which was originally expected to take place in Moscow but never did.

References

1. Olovnikov, A. M. (1971) [Principle of marginotomy in template synthesis of polynucleotides], *Dokl Akad Nauk SSSR*, **201**, 1496-1499,
2. Olovnikov, A. M. (1973) A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon, *Journal of theoretical biology*, **41**, 181-190, doi: 10.1016/0022-5193(73)90198-7.
3. Watson, J. D. (1972) Origin of concatemeric T7 DNA, *Nature: New biology*, **239**, 197-201, doi: 10.1038/newbio239197a0.
4. Greider, C. W., and Blackburn, E. H. (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts, *Cell*, **43**, 405-413, doi: 10.1016/0092-8674(85)90170-9.
5. Greider, C. W., and Blackburn, E. H. (1989) A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis, *Nature*, **337**, 331-337, doi: 10.1038/337331a0.
6. Temin, H. M., and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus, *Nature*, **226**, 1211-1213, doi: 10.1038/2261211a0.
7. Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses, *Nature*, **226**, 1209-1211, doi: 10.1038/2261209a0.

8. Temin, H. M. (1964) Nature of the provirus of Rous sarcoma, *Nat Cancer Inst Monogr*, **17**, 557-570,
9. Temin, H. M. (1985) Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts, *Mol Biol Evol*, **2**, 455-468, doi: 10.1093/oxfordjournals.molbev.a040365.
10. Toh, H., Hayashida, H., and Miyata, T. (1983) Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus, *Nature*, **305**, 827-829, doi: 10.1038/305827a0.
11. Georgiev, G. P., Ilyin, Y. V., Ryskov, A. P., Tchurikov, N. A., Yenikolopov, G. N., Gvozdev, V. A., and Ananiev, E. V. (1977) Isolation of eukaryotic DNA fragments containing structural genes and the adjacent sequences, *Science*, **195**, 394-397, doi: 10.1126/science.401545.
12. Finnegan, D. J., Rubin, G. M., Young, M. W., and Hogness, D. S. (1978) Repeated gene families in Drosophila melanogaster, *Cold Spring Harb Symp Quant Biol*, **42 Pt 2**, 1053-1063, doi: 10.1101/sqb.1978.042.01.106.
13. Glover, D. M., White, R. L., Finnegan, D. J., and Hogness, D. S. (1975) Characterization of six cloned DNAs from Drosophila melanogaster, including one that contains the genes for rRNA, *Cell*, **5**, 149-157, doi: 10.1016/0092-8674(75)90023-9.
14. Schaffner, W., Gross, K., Telford, J., and Birnstiel, M. (1976) Molecular analysis of the histone gene cluster of Psammechinus miliaris: II. The arrangement of the five histone-coding and spacer sequences, *Cell*, **8**, 471-478, doi: 10.1016/0092-8674(76)90214-2.
15. Georgiev, G. P. (1984) Mobile genetic elements in animal cells and their biological significance, *Eur J Biochem*, **145**, 203-220, doi: 10.1111/j.1432-1033.1984.tb08541.x.
16. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K., and Yuki, S. (1984) Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in Drosophila melanogaster, *Nature*, **312**, 659-661, doi: 10.1038/312659a0.
17. Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S., and Saigo, K. (1985) The nucleotide sequences of copia and copia-related RNA in Drosophila virus-like particles, *Nature*, **315**, 773-776, doi: 10.1038/315773a0.
18. Michel, F., and Lang, B. F. (1985) Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses, *Nature*, **316**, 641-643, doi: 10.1038/316641a0.
19. Cappello, J., Handelsman, K., and Lodish, H. F. (1985) Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence, *Cell*, **43**, 105-115, doi: 10.1016/0092-8674(85)90016-9.
20. Hattori, M., Kuhara, S., Takenaka, O., and Sakaki, Y. (1986) L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein, *Nature*, **321**, 625-628, doi: 10.1038/321625a0.
21. Fawcett, D. H., Lister, C. K., Kellett, E., and Finnegan, D. J. (1986) Transposable elements controlling I-R hybrid dysgenesis in D. melanogaster are similar to mammalian LINEs, *Cell*, **47**, 1007-1015, doi: 10.1016/0092-8674(86)90815-9.
22. Lampson, B. C., Sun, J., Hsu, M. Y., 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, doi: 10.1126/science.2466332.
23. 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, doi: 10.1016/0092-8674(89)90693-4.
24. Evgen'ev, M. B., Zelentsova, H., Shostak, N., Kozitsina, M., Barskyi, V., Lankenau, D. H., and Corces, V. G. (1997) Penelope, a new family of transposable elements and its possible role in

hybrid dysgenesis in Drosophila virilis, *Proc Natl Acad Sci U S A*, **94**, 196-201, doi: 10.1073/pnas.94.1.196.
25. Lundblad, V., and Blackburn, E. H. (1990) RNA-dependent polymerase motifs in ESTl: Tentative identification of a protein component of an essential yeast telomerase, *Cell*, **60**, 529-530, doi: 10.1016/0092-8674(90)90653-v.
26. Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T. R. (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase, *Science*, **276**, 561-567, doi: 10.1126/science.276.5312.561.
27. Olovnikov, A. M. (1996) Telomeres, telomerase, and aging: origin of the theory, *Experimental gerontology*, **31**, 443-448, doi: 10.1016/0531-5565(96)00005-8.
28. Arkhipova, I. R., Mazo, A. M., Cherkasova, V. A., Gorelova, T. V., Schuppe, N. G., and Ilyin, Y. V. (1986) The steps of reverse transcription of Drosophila mobile genetic elements and U3-R-U5 structure of their LTRs, *Cell*, **44**, 555-563, doi: 10.1016/0092-8674(86)90265-5.
29. Krupovic, M., Blomberg, J., Coffin, J. M., Dasgupta, I., Fan, H., Geering, A. D., Gifford, R., Harrach, B., Hull, R., Johnson, W., Kreuze, J. F., Lindemann, D., Llorens, C., Lockhart, B., Mayer, J., Muller, E., Olszewski, N., Pappu, H. R., Pooggin, M., Richert-Poggeler, K. R., et al. (2018) Ortervirales: A new viral order unifying five families of reverse-transcribing viruses, *J Virol*, doi: 10.1128/jvi.00515-18.
30. Paul, B. G., Yushenova, I. A., and Arkhipova, I. R. (2022) *The Diversity of Reverse Transcriptases*. in *Retrotransposons and Human Disease* (Gabriel, A. ed.), World Scientific, Singapore. pp 1-28
31. Lambowitz, A. M., and Belfort, M. (2015) Mobile bacterial group II introns at the crux of eukaryotic evolution, *Microbiol Spectr*, **3**, doi: 10.1128/microbiolspec.MDNA3-0050-2014.
32. Arkhipova, I. R., and Yushenova, I. A. (2019) Giant transposons in eukaryotes: Is bigger better?, *Genome Biol Evol*, **11**, 906-918, doi: 10.1093/gbe/evz041.
33. Gao, L., Altae-Tran, H., Böhning, F., Makarova, K. S., Segel, M., Schmid-Burgk, J. L., Koob, J., Wolf, Y. I., Koonin, E. V., and Zhang, F. (2020) Diverse enzymatic activities mediate antiviral immunity in prokaryotes, *Science*, **369**, 1077-1084, doi: 10.1126/science.aba0372.
34. Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., Oppenheimer-Shaanan, Y., and Sorek, R. (2020) Bacterial retrons function in anti-phage defense, *Cell*, **183**, 1551-1561, doi: 10.1016/j.cell.2020.09.065.
35. Bobonis, J., Mitosch, K., Mateus, A., Karcher, N., Kritikos, G., Selkrig, J., Zietek, M., Monzon, V., Pfalz, B., Garcia-Santamarina, S., Galardini, M., Sueki, A., Kobayashi, C., Stein, F., Bateman, A., Zeller, G., Savitski, M. M., Elfenbein, J. R., Andrews-Polymenis, H. L., and Typas, A. (2022) Bacterial retrons encode phage-defending tripartite toxin-antitoxin systems, *Nature*, **609**, 144-150, doi: 10.1038/s41586-022-05091-4.
36. Mestre, M. R., González-Delgado, A., Gutiérrez-Rus, L. I., Martínez-Abarca, F., and Toro, N. (2020) Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems, *Nucleic Acids Res*, **48**, 12632-12647, doi: 10.1093/nar/gkaa1149.
37. Wang, Y., Guan, Z., Wang, C., Nie, Y., Chen, Y., Qian, Z., Cui, Y., Xu, H., Wang, Q., Zhao, F., Zhang, D., Tao, P., Sun, M., Yin, P., Jin, S., Wu, S., and Zou, T. (2022) Cryo-EM structures of Escherichia coli Ec86 retron complexes reveal architecture and defence mechanism, *Nature Microbiology*, **7**, 1480-1489, doi: 10.1038/s41564-022-01197-7.
38. Guo, H., Arambula, D., Ghosh, P., and Miller, J. F. (2014) Diversity-generating retroelements in phage and bacterial genomes, *Microbiol Spectr*, **2**, MDNA3-0029-2014, doi: 10.1128/microbiolspec.MDNA3-0029-2014.
39. 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., and Valentine, D. L. (2017) Retroelement-guided protein

diversification abounds in vast lineages of Bacteria and Archaea, *Nat Microbiol*, **2**, 17045, doi: 10.1038/nmicrobiol.2017.45.
40. Roux, S., Paul, B. G., Bagby, S. C., Nayfach, S., Allen, M. A., Attwood, G., Cavicchioli, R., Chistoserdova, L., Gruninger, R. J., Hallam, S. J., Hernandez, M. E., Hess, M., Liu, W. T., McAllister, T. A., O'Malley, M. A., Peng, X., Rich, V. I., Saleska, S. R., and Eloe-Fadrosh, E. A. (2021) Ecology and molecular targets of hypermutation in the global microbiome, *Nat Commun*, **12**, 3076, doi: 10.1038/s41467-021-23402-7.
41. Paul, B. G., and Eren, A. M. (2022) Eco-evolutionary significance of domesticated retroelements in microbial genomes, *Mobile DNA*, **13**, 6, doi: 10.1186/s13100-022-00262-6.
42. Fortier, L. C., Bouchard, J. D., and Moineau, S. (2005) Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK, *J Bacteriol*, **187**, 3721-3730, doi: 10.1128/jb.187.11.3721-3730.2005.
43. Lopatina, A., Tal, N., and Sorek, R. (2020) Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy, *Annual review of virology*, **7**, 371-384, doi: 10.1146/annurev-virology-011620-040628.
44. Wang, C., Villion, M., Semper, C., Coros, C., Moineau, S., and Zimmerly, S. (2011) A reverse transcriptase-related protein mediates phage resistance and polymerizes untemplated DNA in vitro, *Nucleic Acids Res*, **39**, 7620-7629, doi: 10.1093/nar/gkr397.
45. Mestre, M. R., Gao, L. A., Shah, S. A., López-Beltrán, A., González-Delgado, A., Martínez-Abarca, F., Iranzo, J., Redrejo-Rodríguez, M., Zhang, F., and Toro, N. (2022) UG/Abi: a highly diverse family of prokaryotic reverse transcriptases associated with defense functions, *Nucleic Acids Res*, **50**, 6084-6101, doi: 10.1093/nar/gkac467.
46. Figiel, M., Gapińska, M., Czarnocki-Cieciura, M., Zajko, W., Sroka, M., Skowronek, K., and Nowotny, M. (2022) Mechanism of protein-primed template-independent DNA synthesis by Abi polymerases, *Nucleic Acids Res*, **50**, 10026-10040, doi: 10.1093/nar/gkac772.
47. Zimmerly, S., and Wu, L. (2015) An unexplored diversity of reverse transcriptases in bacteria, *Microbiol Spectrum*, **3**, MDNA3-0058-2014, doi: doi:10.1128/microbiolspec.MDNA3-0058-2014.
48. Simon, D. M., and Zimmerly, S. (2008) A diversity of uncharacterized reverse transcriptases in bacteria, *Nucleic Acids Res*, **36**, 7219-7229, doi: 10.1093/nar/gkn867.
49. Kojima, K. K., and Kanehisa, M. (2008) Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture, *Mol Biol Evol*, **25**, 1395-1404, doi: 10.1093/molbev/msn081.
50. Toro, N., Martinez-Abarca, F., Mestre, M. R., and Gonzalez-Delgado, A. (2019) Multiple origins of reverse transcriptases linked to CRISPR-Cas systems, *RNA Biol*, **16**, 1486-1493, doi: 10.1080/15476286.2019.1639310.
51. Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A. M., and Fire, A. Z. (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein, *Science*, **351**, aad4234, doi: 10.1126/science.aad4234.
52. Mohr, G., Silas, S., Stamos, J. L., Makarova, K. S., Markham, L. M., Yao, J., Lucas-Elio, P., Sanchez-Amat, A., Fire, A. Z., Koonin, E. V., and Lambowitz, A. M. (2018) A reverse transcriptase-Cas1 fusion protein contains a Cas6 domain required for both CRISPR RNA biogenesis and RNA spacer acquisition, *Mol Cell*, **72**, 700-714, doi: 10.1016/j.molcel.2018.09.013.
53. Park, S. K., Mohr, G., Yao, J., Russell, R., and Lambowitz, A. M. (2022) Group II intron-like reverse transcriptases function in double-strand break repair, *Cell*, **185**, 3671-3688.e3623, doi: 10.1016/j.cell.2022.08.014.
54. Fragkiadaki, P., Renieri, E., Kalliantasi, K., Kouvidi, E., Apalaki, E., Vakonaki, E., Mamoulakis, C., Spandidos, D. A., and Tsatsakis, A. (2022) Telomerase inhibitors and activators in aging and cancer: A systematic review, *Mol Med Rep*, **25**, 158, doi: 10.3892/mmr.2022.12674.

55. Gladyshev, E. A., and Arkhipova, I. R. (2011) A widespread class of reverse transcriptase-related cellular genes, *Proc Natl Acad Sci U S A*, **108**, 20311-20316, doi: 10.1073/pnas.1100266108.
56. Yushenova, I. A., and Arkhipova, I. R. (2018) Biochemical properties of bacterial reverse transcriptase-related (rvt) gene products: multimerization, protein priming, and nucleotide preference, *Curr Genet*, **64**, 1287-1301, doi: 10.1007/s00294-018-0844-6.
57. Dlakic, M., and Mushegian, A. (2011) Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase, *RNA*, **17**, 799-808, doi: 10.1261/rna.2396011.
58. Grainger, R. J., and Beggs, J. D. (2005) Prp8 protein: at the heart of the spliceosome, *RNA*, **11**, 533-557, doi: 10.1261/rna.2220705.
59. Galej, W. P., Oubridge, C., Newman, A. J., and Nagai, K. (2013) Crystal structure of Prp8 reveals active site cavity of the spliceosome, *Nature*, **493**, 638-643, doi: 10.1038/nature11843.
60. Lambert, M. E., McDonald, J. F., and Weinstein, I. B. (1988) *Eukaryotic transposable elements as mutagenic agents*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
61. Arkhipova, I. R. (2012) *Telomerase, retrotransposons, and evolution*. in *Telomerases: Chemistry, Biology, and Clinical Applications* (Lue, N. F., and Autexier, C. eds.), John Wiley & Sons, Inc., Hoboken, NJ. pp 265-299
62. Lue, N. F., and Autexier, C. (2006) The structure and function of telomerase reverse transcriptase, *Annu Rev Biochem*, **75**, 493-517, doi: 10.1146/.
63. Geer, L. Y., Domrachev, M., Lipman, D. J., and Bryant, S. H. (2002) CDART: protein homology by domain architecture, *Genome Res*, **12**, 1619-1623, doi: 10.1101/gr.278202.
64. Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., and Geer, L. Y. (2015) CDD: NCBI's conserved domain database, *Nucleic Acids Res.*, **43**, D222-226, doi: 10.1093/nar/gku1221.