Identification of Novel Types of Prokaryotic Retroelements Based on Gene Neighborhood and Protein Architecture

Kenji K. Kojima
kojimakk@kuicr.kyoto-u.ac.jp

Minoru Kanehisa
kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

Keywords: retroelement, reverse transcriptase, gene neighborhood, protein architecture

1 Introduction

Reverse transcriptase (RT) is RNA-dependent DNA polymerase that is thought to have played the central role in the transition from the “RNA world” to the “DNA world”. At present, however, RTs are observed only in mobile genetic elements, called “retroelements”, with only one exception, telomerases. Retroelements include various types of mobile elements, such as transposons, viruses, plasmids, and so on. The variety of life styles of retroelements results from the coupling with various genes. For example, the combination of RT and endonuclease enables non-long terminal repeat (non-LTR) retrotransposons to replicate themselves. Despite the detailed knowledge of eukaryotic retroelements, only three groups of prokaryotic retroelements have been characterized: group II self-splicing introns, retrons, and diversity-generating retroelements (DGRs). Recent genome sequencing revealed many RT genes from prokaryotes, but most of them are unclassified. Here, we classify prokaryotic retroelements based on the phylogeny, the gene neighborhood, and the protein architecture, and identify various types of prokaryotic retroelements.

2 Materials and Methods

We performed PSI-BLAST to RefSeq protein database with three iterations using the protein sequences of representatives for all characterized RT groups as queries. We performed PSI-BLAST repeatedly using all hit prokaryotic RT proteins as queries until no new prokaryotic RT proteins were hit. We obtained the neighboring gene sequences of each RT gene, and clustered all neighboring genes using BLASTCLUST program. We removed clusters in which all sequences are from the same genus, and clusters of transposases. After that, we manually checked all remained clusters whether neighboring RT genes share high sequence similarity or not. We aligned all prokaryotic reverse transcriptase sequences with MAFFT, and constructed the phylogenetic tree based on the neighbor-joining method using CLUSTALX.

3 Results and Discussion

We found 516 RT genes in prokaryotic genomes. The gene neighborhood search revealed that 15 gene families are flanked with more than two RT genes. In addition, we found 16 RT proteins including other domains except H-N-H endonucleases, which is common in group II introns. Genes flanked with RTs and domains fused with RTs contain various types of motifs, including SWIM zinc-finger, OLD family nuclease, XRE family transcriptional regulator, ribosomal protein S23, HRDC DNA/RNA-binding domain, RNase H1, DNA polymerase A, bacterial primase, archaeo-eukaryotic primase, and amidohydrolase. It is noteworthy that most of them are DNA/RNA-catalyzing enzymes or DNA/RNA-binding proteins. These genes seem to act in the life cycle of retroelements. Some RTs are phylogenetically members of known retroelements, but others are not similar to known retroelements. They

Figure 1: Structure of retron Mx65
are considered to represent novel types of retroelements.

Figure 1 shows the structure of retron Mx65. Retrons have an ability to produce an unusual satellite DNA known as msDNA. Retrons consist of three regions, RT gene, msd and msr [4]. msd encodes the DNA strand of the msDNA, and msr encodes the RNA strand of the msDNA. We found that retron Mx65 and its relatives contain additional genes encoding SWIM zinc-finger proteins. These retron SWIM proteins seem to bind msDNA and stabilize it.

Figure 2 represents one group of novel retroelements, which contain RT and DNA polymerase A. DNA polymerase A domain is observed in bacterial DNA polymerase I, mitochondrial DNA polymerase gamma, and T-odd bacteriophage DNA polymerases [5]. The possible function of DNA polymerase A is the second strand cDNA synthesis after reverse transcription.

One of the most interesting results is that different types of primases are associated with RTs. Three RTs neighbor on DnaG bacterial primases. One RT is fused with DnaG. Two RTs are fused with archaeo-eukaryotic primases. Iyer et al. [2] pointed out that archaeo-eukaryotic primases and bacterial primases are functionally equivalent and they have repeatedly displaced each other in various extrachromosomal replicons. In eukaryotic retroelements, several displacements of equivalent enzymatic domains have been observed. Non-LTR retrotransposons exchanged their endonucleases from restriction-like to apurinic/apyrimidinic-like [3]. Several groups of LTR retrotransposons acquired lambda-like recombinases and simultaneously lost D-D-E integrases [1]. Our finding of different primase genes associated with RTs suggests that displacements of equivalent enzymes have also occurred in the evolution of prokaryotic retroelements.

At least two independent acquisitions of primases by prokaryotic retroelements show the importance of primers for reverse transcription. Like other DNA polymerases, most RTs cannot start synthesizing DNA without primers. Retroviruses and LTR retrotransposons use tRNAs as primers, whereas non-LTR retrotransposons and group II introns use target DNA nicked by their endonucleases. Encoding primase is, therefore, one solution to supply primers for reverse transcription with ease.

References


