5-2013

Detection of Bacterial Retroelements Using Genomics

Sen Mu
East Tennessee State University

Detection of Bacterial Retroelements Using Genomics

A thesis

presented to

the faculty of the Department of Health Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Biology

---

by

Sen Mu

May 2013

Dr. Bert C. Lampson, Chair

Dr. Christopher Pritchett

Dr. Dhirendra Kumar

Keywords: Retroelements, Genomics

ABSTRACT

Detection of Bacterial Retroelements Using Genomics

by

Sen Mu

The reverse flow of genetic information can occur when a special DNA polymerase called Reverse Transcriptase (RT) copies the genetic information in an RNA molecule back into a complementary DNA. One type of RT encoding gene found in bacteria is called a retron element. Recent bacterial genome sequencing projects have revealed many examples of retron RT genes. This gene assignment is based on comparison with a few known retron RT proteins. However, RT proteins are highly diverse in their amino acid sequences, and thus the assigned identity of these RT proteins as retrons in genome databases is questionable. One way to prove that these postulated RTs are indeed from retron elements is to see if they can produce msDNA. Retron RTs are known to synthesize a structurally unique satellite DNA called msDNA in the bacterial cells that contain them. Based on GenBank database matches to a known protein, 7 proteins designated as retron RTs were tested for their ability to synthesize msDNA. Five of these retron RTs did show evidence of producing msDNA and are from very different bacterial hosts. The other 2 RT proteins did not show any evidence that they produce msDNA.

DEDICATION

I dedicate this thesis to my parents, Yanlai Zhang and Jiangang Mu, who gave me very valuable financial and moral support to finish my master's degree. Also I dedicate this to my wife Shanxin Liang who supported me all the time. Without their support I would never go further.

ACKNOWLEDGEMENTS

I really would like to express my appreciation to Dr. Bert Lampson, my advisor and committee chair. He is a very wonderful mentor in this project. I really appreciate his training, help, guidance, and support whenever I had trouble. Without his help, I could not complete this work. I also really want to thank to my other committee members, Dr. Christopher L. Pritchett and Dr. Dhirendra Kumar, for their valuable help and support, reviewing my thesis, and troubleshooting my work. I also would like to thank to my departmental chair Dr. Ranjan Chakraborty for his encouragement and support. Also thank to Mr. Bill Wright and Mrs. Robin Grindstaff for their help and support in the lab. I also would like to thank Mrs. Phyllus Cooper for all kinds of official paper work she helped me do. Moreover, many thanks go to the all faculty in the Department of Health Sciences for giving me this opportunity to study here. Also, I am really grateful for my labmates: Thomas Barber, Ralitha Borisova, Megan Carr, Brian Hill, Sean Stacy, Huston Heatherly, Pablo Palazon, Jane Dean, and Yuxi Sun. They helped me make my life more fabulous during the research. Finally, thanks a lot my parents and my wife, without their help, I can not make it further.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Reverse Transcriptase, Reverse Transcription, and Retroelements

The first discovery of Reverse Transcriptase (RT) was in 1970. It was discovered by Howard Temin at the University of Wisconsin–Madison in Rous Sarcoma Virus (RSV) virions (1), and by David Baltimore in 1970 at MIT isolated the Reverse Transcriptase from 2 RNA tumor viruses: Rauscher Murine Leukemia Virus (R-MLV) and Respiratory Syncytial Virus (RSV) virions (2). The work from Temin and Baltimore was recognized very quickly and both of them got the Nobel Prize for the discovery of Reverse Transcriptase. The significance is that it reveals the reverse flow of genetic information and this is the exception of the central dogma of genetics. For the molecular biology research, the ability of RT to convert mRNA to DNA can be used to create a large number of cDNA libraries. This has facilitated the cloning and study of genes involved in all areas of biology. The discovery also caused an explosion of research into retroviruses, RNA viruses that replicate via Reverse Transcription. This discovery work was critical 15 years later, when HIV was shown to be a retrovirus.

The main function of RT is Reverse Transcription. Reverse transcription is the process by which DNA is synthesized from an RNA template. It is usually accomplished by a specially dedicated RNA-dependent DNA polymerase which is RT. Retroelements are genetic elements that contain a RT gene, but some Retroelements may not encode for RT but are simply produced by Reverse Transcription like Alu elements. Alu elements are retrotransposons and look like DNA copies made from RNA polymerase III-encoded RNAs. Alu elements do not encode for RT and depend on LINE retrotransposons for their replication (3). Most Retroelements are

mobile. During mobilization, the element is transcribed to RNA, then reverse-transcribed to DNA, and integrated elsewhere in the genome. Retroelements have had a large impact on eukaryotic cells. They are abundant in eukaryotic genomes and they take up a large amount of chromosome DNA, including ~40% of the human genome and ~60% in maize (4, 5).

Class of Retroelements

In eukaryotic cells, *retroviruses* are typical retrolements because of their Reverse Transcription. Some other examples are the *Hepadnaviruses*, transposons including the LTR retrotransposons and Non-LTR retrotransposons. Also the mitochondrial plasmids and telomerase are also considered to be Retroelements. Because the telomerase is a ribonucleoprotein and it carries its own RNA molecules that can act as a template when telomerase elongates the ends of chromosomes via Reverse Transcription (6).

Generally there are 3 well defined types of RTs in prokaryotic cells. The 3 types of Retroelements are: Group II introns found in both eubacterial and archaeal genomes. Retrons, are are also found in eubacteria and some archaea, and the diversity generating Retroelements (DGRs) have been found in eubacteria (7).

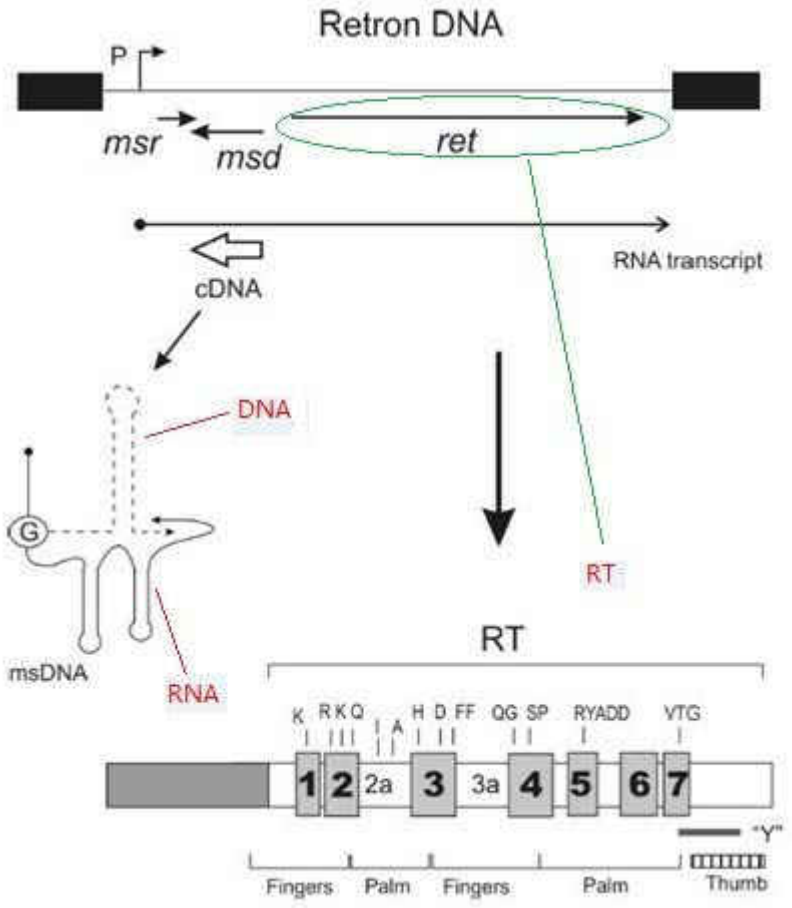
The best understood Retroelements among prokaryotic organisms are group II introns. They were initially discovered in the genomes of organelles like chloroplast and the mitochondria of yeast (8), but some have since been widely found in bacteria. Group II introns are also similar to eukaryotic intron DNA in that they may interrupt a gene between 2 exons and have to be excised out after they are transcribed into mRNA in order to yield a mature RNA that codes for a functional protein. After the group II introns are transcribed into mRNA, an autocatalytic RNA will be formed on mRNA by the group II introns, so group II introns can be

spliced out from the mRNA without any protein factors from the host cell. After that, a special lariat will be formed by the excised intron RNA. The mechanism of this process is similar to the splicing of nuclear introns by the eukaryotic spliceosome. The group II intron DNA contains an open reading frame (ORF). The ORF encodes for a protein that has many functions. First it can code for a maturase that aids in intron splicing, and second the intron protein has RT activity that converts the spliced intron RNA into a cDNA, and finally also in some cases the intron protein can perform as a DNA endonuclease that helps the intron insert into a new site in DNA (9).

There is another type of bacterial retroelement called DGRs. They are not mobile like group II intron, but they function to make the diverse DNA sequence instead (10). There is a phage called BPP-1 of *Bordetella pertussis*. It encodes a RT, RNA template (TR), an accessory protein (*atd*), and a target protein gene (*mtd*) that contains a C-terminal variable region (VR). VR and TR are the matching pair with directly repeated sequences. All of these "genes" compose the DGR retroelement. DGRs produce DNA sequence variation by Reverse Transcription. DGRs contain an ORF that encodes a bacterial RT. Upstream of RT is the region TR and VR. After the TR is transcribed to RNA then it is reverse transcribed to cDNA. Then this cDNA will combine with VR gene via site –specific homing. This produces the variation leading to changes in the amino acid sequences of VR. So, DGR produces the diversity sequences in the VR region of the phage tail protein. This is the region of phage that contacts the bacterium during infection. In this behavior, when *Bordetella* cells change between the pathogenic phases and free-living phases the DGRs can mediate tropism switching and allows the phage to infect cells with altered surface composition. RTs of DGRs appear to be more closely related to group II introns than retrons (11, 12).

Retrons and Organization

The first retroelement to be discovered in bacteria is the retron. These Retroelements are very strange and their function is still unknown. Also, they only do one thing: They synthesize a large number of multicopy single stranded DNA (msDNA). They are not thought to be independently mobile or have a clear phenotype associated with them. The RT (*ret*), msDNA (*msd*) and msRNA (*msr*), (Fig.1) make up the retron genes. Retrons are usually about 2000 base pairs long and the DNA sequence is highly varied. Some retrons can be found inserted into the bacterial chromosome or as a part of prophage DNA. The only known activity of these retrons is to produce msDNA. All retron elements have at least one large ORF designated *ret*. The amino acid sequence of this ORF product, which can range from about 298 to ~700 amino acids, is similar to eukaryotic RTs. If the ORF is inactivated, these retrons fail to produce msDNA. Therefore, it must have a functional RT to do the synthesis of msDNA and it is the first example of a reverse transcribed cDNA in bacteria. Upstream of *ret* gene are the *msd* gene and *msr* gene. The *msd* and *msr* are oppositely positioned with some overlaps at their 3' ends, but they are under the same control of a common promoter. The *msd* and *msr* are required to synthesize msDNA, *msd* codes for DNA part, and *msr* codes for RNA part of msDNA. The *ret* gene codes for RT part that catalyzes the msDNA Synthesis. Because the msDNA can not replicate on its own, the genes *msr, msd*, and *ret* essentially form an operon together to synthesize msDNA.



(BC. Lampson, 2007)

Fig 1: Organization of retron DNA, which is the operon to make msDNA. On the very top, the thin line indicates the retron DNA that contain about 2Kb of unique DNA that is usually inserted into chromosome or phage DNA (black boxes). RT coding region is the ORF (ret) and the adjacent region is the *msd* and *msr* coding for DNA and RNA part of msDNA. After the Reverse Transcription, the msDNA forms. All RT have the conserved regions from 1-7 boxes and form the fingers and palm structure. (7)

msDNA Structure

msDNAs are very small and structurally unique satellite DNAs found in a number of Gram-negative and Gram-positive bacteria. They are composed of hundreds of single stranded DNA, so they are called multicopy single stranded DNA or msDNA. msDNA is actually a complex of DNA, RNA, and maybe protein. RT catalyzes the synthesis of msDNA by Reverse Transcription. So far, studies hae determined the structure of msDNA, the replication mechanism,

the organization and function of the retron, the structure and functions of the RT protein required to make it, and even its potential usefulness as a practical genetic tool (7, 13).

First discovery of the strange structure of msDNA was from the soil bacterium *Myxococcus xanthus.* For example in the msDNA-Mx162 from *Myxococcus xanthus,* the satellite DNA is composed of a 162 nucleotide single-stranded DNA that folds into a stable stem-loop structure. There is a single stranded RNA about 78 bases that is covalently linked to the 5' end of the DNA. (Fig.2). However, the RNA strand exposes a free 5' end and a free 3' end and it is instead joined to the DNA strand at an internal guanosine nucleotide through a 2'-5' phosphodiester bond. That's why msDNA is a unique molecule that contains a DNA molecule that is joined to an RNA molecule by a 2'-5' linkage. Also at the end of the 3' end of RNA and DNA, they both base-pair with each other. On the msDNA there are not only the DNA, RNA, but also protein. The RT is the most typical protein associated with msDNA. Also, other host proteins may bind to msDNA. For example, there is an enzyme called RNaseH produced by the host that might bind to msDNA and may be involved in the msDNA Synthesis (14).

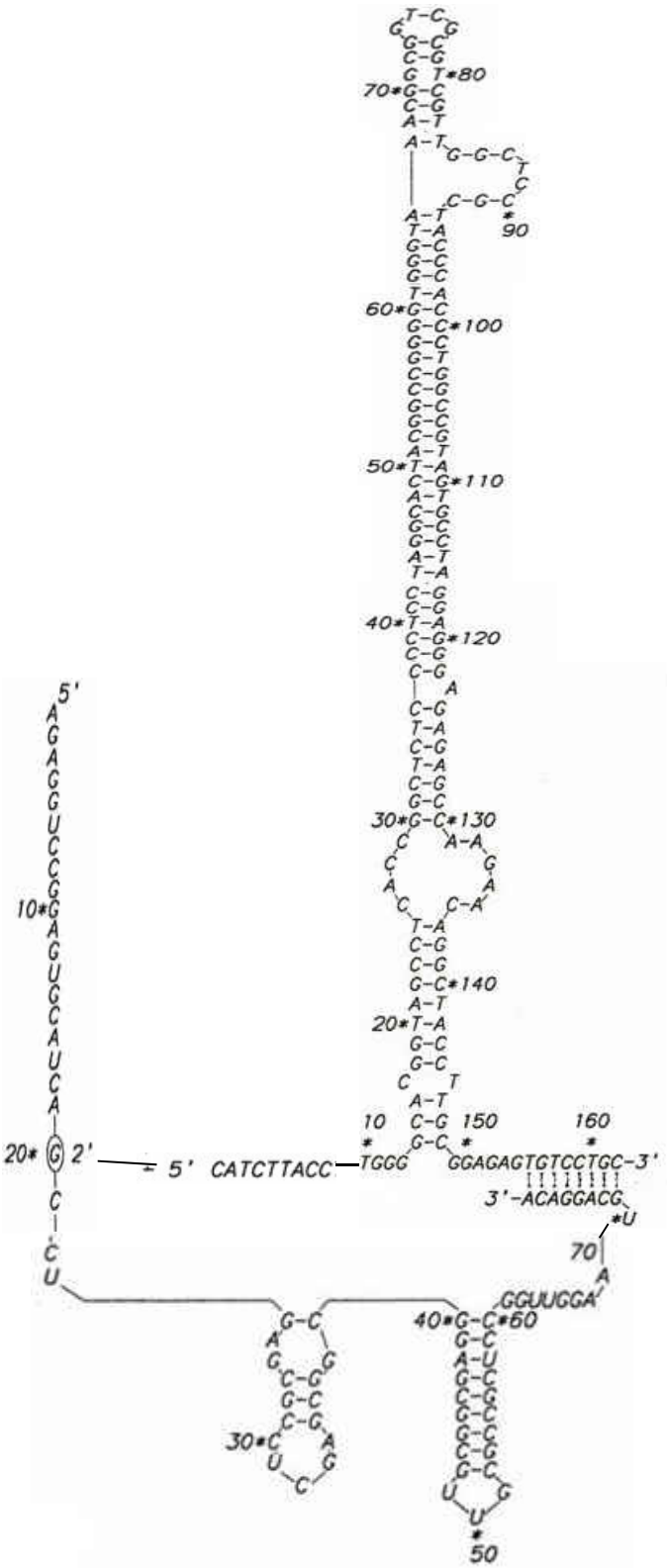


Fig 2: Structure of Mx162, stem-loop structure. RNA part has about 78 bases, and has a free 5' and 3' end. There is a unique 2'-5' linkage between RNA and DNA instead of normal 3'-5' linkage. At the RNA and DNA 3' end, they base-pair together.

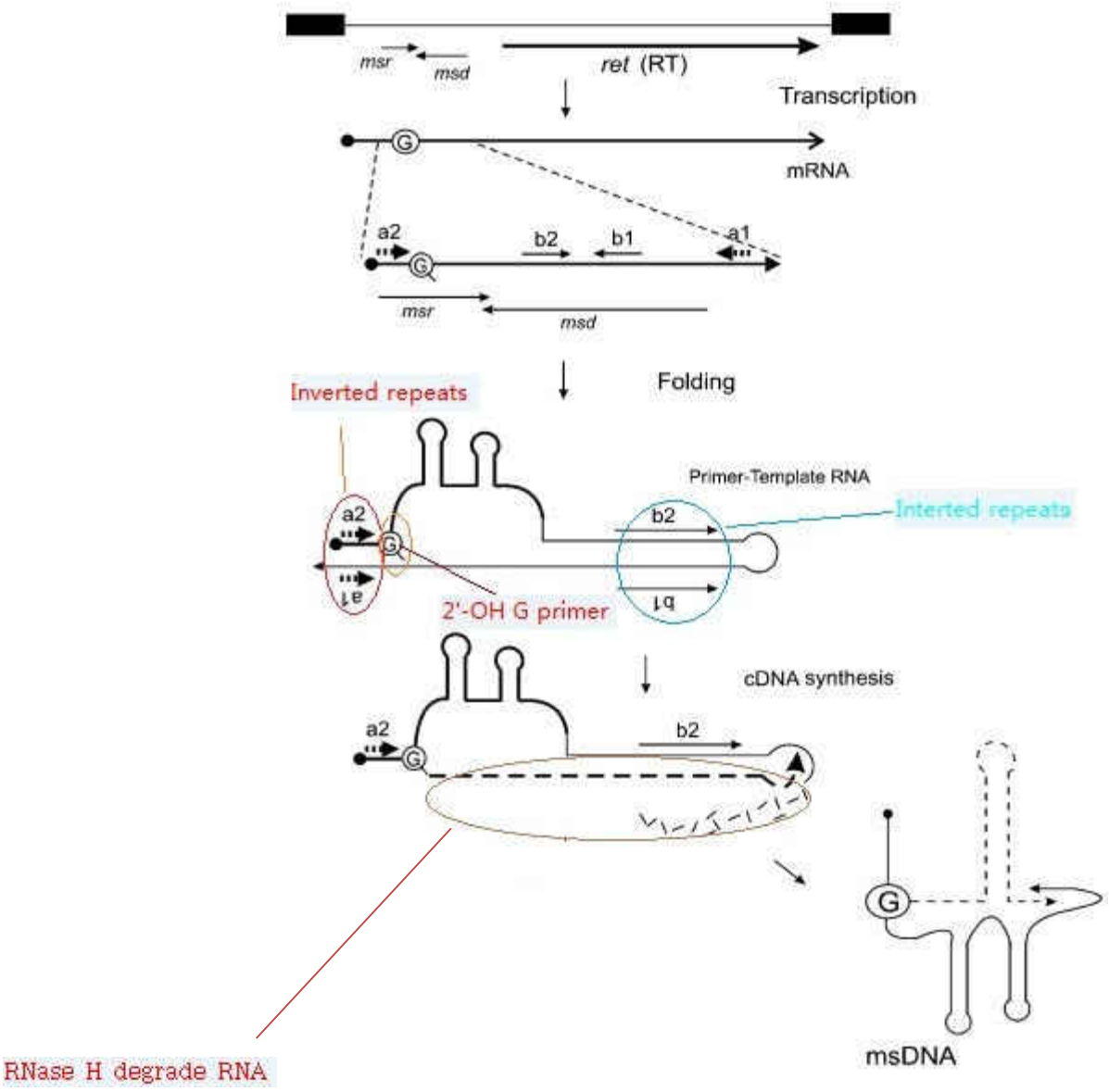
msDNA Synthesis

Synthesis of msDNA begins with the transcription of the retron encoded *msr-msd-ret* operon into a long mRNA (Fig.3). The steps are the RT (*ret*) gene is translated first to form the polymerase in order to use the *msr-msd* region of mRNA as template and primer for msDNA Synthesis. Next, 2 sets of the inverted repeat sequences a1, a2 and b1, b2 (Fig.3) are very important. These 4 inverted repeats base-pair together and fold the mRNA between a1, a2 and b1, b2, this positions the 2'-OH of an internal G that serves as a primer for the initiation. The sequence on mRNA with a stem loop will be the RNA part of msDNA, and the other part of mRNA will serve as a template for cDNA synthesis. Therefore, the proper folding of the RNA into the secondary structures allows this RNA to serve as both a primer and template for cDNA synthesis. Next, the RT binds to the 2'-OH G primer and initiates the cDNA synthesis. During the elongation, a long cDNA forms and meanwhile the cDNA synthesis continues and along with the removal of part of the RNA template (dished line) by host RNase H (15) until a specific termination point is reached within the primer-template RNA. Finally, both RNA and DNA parts of msDNA form and, after several transformations, become the mature msDNA.

Prevalence of Retroelements Among Bacteria

Of the 3 types of retro elements found in bacteria, the group II introns are the most prevalent when searching bacterial genome sequence database. But retron elements are also commonly found and widely distributed among the hundreds of bacterial genome sequences. The presence of a retron in a particular bacterial genome sequence is based on the presence of the retron encoded gene (*ret)* for the retron RT. Based on amino acid sequence similarity to known retron proteins, a particular bacterial genome is annotated as containing a retron type RNA

dependent DNA polymerase. But a comparison of these genome annotated retron proteins reveals a great degree of amino acid sequence variation. In addition, before the availability of DNA sequence data based, retron elements were discovered by detecting the presence of the small satellite DNA (msDNA) in DNA or RNA preparations from bacterial cells. So, how accurate is the identification of retron elements based on genomics when as a group retron RT proteins are so diverse in their amino acid sequences? I plan to investigate this question by looking for the production of msDNA among the selected group of bacteria that are postulated to contain a retron element based of their genome annotation.



(BC. Lampson, 2007)

Figure 3: msDNA Synthesis by Reverse Transcription. After the retron is transcribe to mRNA, it has 4 inverted repeats on mRNA, a1,a2,b1,b2 that are important to form the 2'-5' primer at the specific G to initiate cDNA synthesis by RT on the next step. Then the primer-template structure forms, the cDNA is synthesized by RT. After the cDNA synthesis is complete, part of the RNA template remains joined to the 5' end of the cDNA to yield the complete msDNA. (13)

CHAPTER 2

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

Myxococcus xanthus strain (DZF1) was grown on the CYE solid agar plates. They were inoculated on CYE agar plates and grown 3 days in a 30°C incubator. *Rhodococcus jostii* strain (RHA1) was inoculated in RM broth or RM solid agar plates. RHA1 was cultivated at 30°C for 3 days, with shaking in the case of liquid cultures. *Microscilla marina* strain (ATCC 23134) was inoculated in Marine broth (Difco) or Marine solid plates. *Microscilla marina* was cultivated at room temperature for 3 days, with shaking in the case of liquid cultures. *Shewanella* sp. strain (MR-7) was inoculated in LB broth or LB solid agar plates. *Shewanella* MR-7 was cultivated at room temperature for 24 hrs, with shaking in the case of liquid cultures. *Ralstonia eatropha* strain (H16) was inoculated in LB broth or LB solid agar plates. *Ralstonia etropha* H16 was cultivated at room temperature for 24 hrs, with shaking in the case of liquid cultures. *Pseudomonas aeruginosa* strain (PAO1) was inoculated in PIA broth or PIA solid agar plates. *Pseudomonas aeruginosa* (PAO1) was cultivated at room temperature for 24 hrs, with shaking in the case of liquid cultures. *Methylobacterium radiotolerans* strain (JCM 2831) was inoculated in nutrient broth+1% methanol or Nutrient broth+1% methanol solid agar plates. *Methylobacterium radiotolerans* was cultivated at 30°C for 3 days, with shaking in the case of liquid cultures. *Gemmata obscuriglobus* strain (UQM 2246) was inoculated on *Caulobacter* solid agar plates. *Gemmata obscuriglobus* was cultivated at 30°C for 2 weeks.

Growth Media

CYE Medium

CYE agar plates were prepared as listed below (Table 1). All ingredients were mixed together and autoclaved together. After autoclaving, the medium was cooled in a 50°C water bath, then plates were poured.

Table 1.CYE medium recipe

Ingredient	**Amount**
Casitone	5g
Yeast extract	2.5g
$MgSO_4*7H_2O$	1g
Agar	7.5g
Distilled water	500mls

RM Medium

The preparation of this medium was the same way as in above. The ingredients were as follows: (Table 2)

Table 2. RM medium recipe

Ingredients	**Amount**
Glucose	5g
Nutrient broth	4g
Yeast Extract	0.25g
Agar	7.5g
Distilled water	500mls

Marine Medium

The marine medium we used was Difco Marine Agar 2216 that was premade. The way we prepared was the same as in above. (Table 3)

Table 3.Marine medium recipe

Ingredients	**Amount**
Difico Marine Agar	37.8g
Distilled water	1L

LB Medium

The way we prepared LB medium was the same as in above. The ingredients were as follows: (Table 4)

Table 4.LB medium recipe

Ingredients	**Amount**
Tryptone	5g
NaCl	2.5g
Yeast Extract	2.5g
Agar	7.5g
Distilled water	500mls

PIA Medium

The way we prepared PIA medium was dissolving the enzymatic digest of gelatin, MgCl, K_2SO_4, Irgasan and glycerol in distilled water first by heating for 5min. After heating, we cooled it down and adjusted the final pH to 7.0 with HCl. The ingredients were as follows: (Table 5)

Table 5.PIA medium recipe

Ingredients	**Amount**
Enzymatic digest of gelatin	20g
MgCl	1.4g
K_2SO_4	10g
Irgasan	0.025g
Glycerol	20mls
Agar	13.6g
Distilled water	1L

Nutrient Medium+1% Methanol

The way we prepared this medium was dissolving the Difco nutrient agar powder into distilled water and added 5mls of methanol (Table 6). After that, medium was autoclaved.

Table 6. Nutrient Medium+1% Methanol medium recipe

Ingredients	**Amount**
Difco nutrient agar powder	11.5g
Methanol	5mls
Distilled water	1L
Agar	15g

Caulobacter Medium

To prepare the *Caulobacter* medium, we mixed all components together and autoclaved together. After autoclaving, it was cooled down in a 50°C water bath. Then plates were poured. The preparation of this solution was as follows: (Table 7)

Table7. *Caulobacter* medium recipe

Ingredients	**Amount**
Peptone	1g
Yeast extract	0.5g
Distilled water	500mls
Agar	7.5g
$MgSO_4*7H_2O$	0.05g

Isolation of Genomic DNA from for all bacterial strains

Genomic DNA was isolated by first growing DZF1 on CYE agar plates. Cells were scraped (grown ~3 days at 30°C) with a loop from CYE plates and dispersed cells into 1ml of a sterile test tube containing 500µl 50mM Tris-HCL, pH 8.0 buffer, then vortexed to disperse cells into an even suspension. Cell suspension was transferred to a 1.5ml microfuge tube. It was centrifuged at 14,000 rpm for 50 seconds, then aspirated off all liquid and saved the cell pallet. 160µl of sucrose was added (25%-50mM Tris pH 8.0+60µl lysozyme solution in 50mM Tris pH 8.0 and incubated at 37°C for 30 min. Next, 33.5µl 0.5M EDTA was added and mixed, let sit for 5 min. 50µl of 10% SDS was added and the tube was inverted many times to lyse the bacterial cells to clear the suspension and 5µl of proteinase K (20mg/ml) was added to mix, then incubated at 37°C for 38 min. Next, 190µl 50mM Tris pH 8.0 was added. After that, phenol extraction was

done twice: 500µl phenol saturated with 1xTE was added and the tube was inverted many times, then tubes were centrifuged for 3min at 14,000 rpm. The aqueous top layer was removed to a new sterile microcentrifuge tube. Chloroform: isoamyl alcohol (24:1) extraction was done twice: 450µl of chloroform was added to aqueous phase in a new tube and the tube was inverted several times, centrifuged for 1min, and top aqueous phase was removed to a new tube. The DNA was then precipitated by adding 1ml cold 100% ethanol and 1/10 volume of 3M NaoAc (pH 7.5) to the remaining aqueous phase. The tube was inverted many times until the chromosome clots form. A pipette tip was used to spool out the chromosome clots and they were transferred to a new sterile microcentrifuge with 70% ethanol. The chromosome clots were rinsed gently by inverting the tube. After that, the tube was centrifuged at 14,000 rpm for 1 min, and all liquid was removed by an aspirator, then it was air dried briefly. Finally, the chromosome DNA was redisolved in about 50-100µl of 0.1xTE buffer.

Isolation of Plasmid DNA for All Strains

Bacterial cells were scraped from agar plates and transferred into a sterile microcentrifuge tube containing 500µl 50mM Tris-HCL, pH 8.0 until the buffer was very turbid with cells. The cell suspension was centrifuged at 14,000 rpm for 50 seconds, and the supernatant was discarded. The cell pellet was resuspended in 1ml STE buffer (9.66mls 1X TE, 0.33mls 3M NaCl). Then the suspension was centrifuged at 14,000 rpm, the cell pellets were saved, and the supernatant was discarded. After that, cells were resuspended in a) 200µl solution I+ fresh lysozyme (2mg/ml) and mixed completely at room temperature for 10min, b) after that, 400µl fresh solution II was added and tube was inverted gently several times to lyse the cells, DNA could not be exposed for more than 5min, and c) then 300µl solution III was added and

kept cold, the tube was inverted several times gently but enough to ensure good mixing, formation of a white flocculant was observed and it was kept in ice for 15min.

1) Solution I— 50mM glucose, 10mM EDTA, 25mM Tris pH 8.0
2) Solution II—Distill water 1.4mls, 1M NaOH 400µl, SDS 10% 200ul
3) Solution III— 5M potassium acetate pH 6.0

Next, the tube was centrifuged for 15min at 14,000 rpm. Half of the clear supernatant volume was ransferred into a microfuge tube. 200µl phenol and 200µl chloroform were added into the clear layer and extracted in ice for 5min. After that, the top layer was collected and transferred to a new microfuge tube. Then chloroform was added to extract twice (400µl each time) and the tube was centrifuged for 1min. After spinning, the final top layer was collected into a new tube and 1ml EtOH was added, tube was iced for 5min, then the tube was centrifuged for 10min at 14,000 rpm, then the EtOH was removed. The plasmid DNA pellet was washed with 400ul 70% EtOH and centrifuged, all EtOH was removed, let it air dry. Finally, DNA was redissolved in 0.1TE+RNase (30ul/tube)

DNA Electrophoresis

For large DNA molecules (plasmid and chromosome), the separation was by electrophoresis in 0.75% agrose gels. The loading dye (~ 1/10 volume of total DNA sample) was mixed with DNA sample, then the mixture was carefully loaded into the wells. The gel was run in 1x TBE (5 mM Tris, 1mM EDTA, 45mM boric acid, pH 8.3) at 120 volts for about 1 hour. After running, the gel was carefully removed from the tank and stained in 100mls of 10µg/ml ethidium bromide for 15min. After staining, the gel was placed on the UV box, then the UV light was turned on to observe the DNA bands. After that, the gel was removed from the UV box and

was placed in distilled water to wash with shaking for about 15 min. Next, the gel was ready to be photographed.

Five percent polyacrylamide gels were used to detect the smaller size DNA molecules (msDNA) in 1x TBE buffer. To prepare a 5% mini-acrylamide gel, the following was mixed in a small vacuum flask: 8.6ml H_2O, 1.2 ml 10x TBE, 2.0ml 29:1Acrylamide: Bis solution. After they were mixed, 100µl 10% ammonium persulfate (in water) were added to the small vacuum flask. After that, the mixed solution was degased, then 8µl TEMED was added to start polymerization. The solution was poured between 2 glass plates (11cm x 11cm) and let solidify for 50 min, and then the DNA samples plus loading dye (1/10 of total DNA sample volume) were mixed together, then the samples were carefully loaded into the wells. The gel was run at 120 volts in 1x TBE buffer for about 1 hour. After running, the gel was stained with ethidium bromide for 15 min for observation. Then the gel was washed in distilled water for 15 min with shaking for photograph

PCR Amplication of Retron DNA

All the primers used were selected by using the program "primer3" (http://frodo.wi.mit.edu/), and all the restriction sites were mapped using the "NEB cutter 2.0" (http://tools.neb.com/NEBcutter2/) program. The plasmid we used for cloning was pUC9 (Fig.4). The primers I used were as follows:

1. (a) Upstream primer:

   HindIII
   Mx162(58)—5'-ATAAAAGCTTCTCTGGTTCGACTCGGATG-3'

   (b) Downstream primer:

EcoRI

Mx162(2083)—5’-AATAGAATTCGCGGCATTGAGGACTCCGT-3’

2. (a) Upstream primer:

HindIII

RHA1 retron53— 5’-TATCAAGCTTGATCATGTGGCATTCTCACG-3’

(b) Downstream primer:

EcoRI

RHA1 retron1593— 5-TATTGAATTCTTGGTCAACCCCGACAC-3’

3. (a) Upstream primer:

PstI

Microscilla217— 5’-ATTACTGCAGGGCGGCCAGGTAGAA- 3’

(b) Downstream primer:

EcoRI

Microscilla842—5’-ATATGAATTCAACTGCTGCCCCGATA-3’

4. (a) Upstream primer:

HindIII

Shewanella1—5’-TATTAAGCTTGGGGATGGTGAACTTCTGTG-3’

(b) Downstream primer: Shewanella1672

EcoRI

5’-ATATGAATTCAGTTATACCGCCAACCCAAG-3’

5. (a) Upstream primer:

HindIII

Ralstonia1— 5’-TATTAAGCTTCCCACAGGCCGTAGAGATAG-3’

(b) Downstream primer:

EcoRI

Ralstonia1400— 5’-ATATGAATTCGCCAGCTCGACTGAGAAGTG-3’

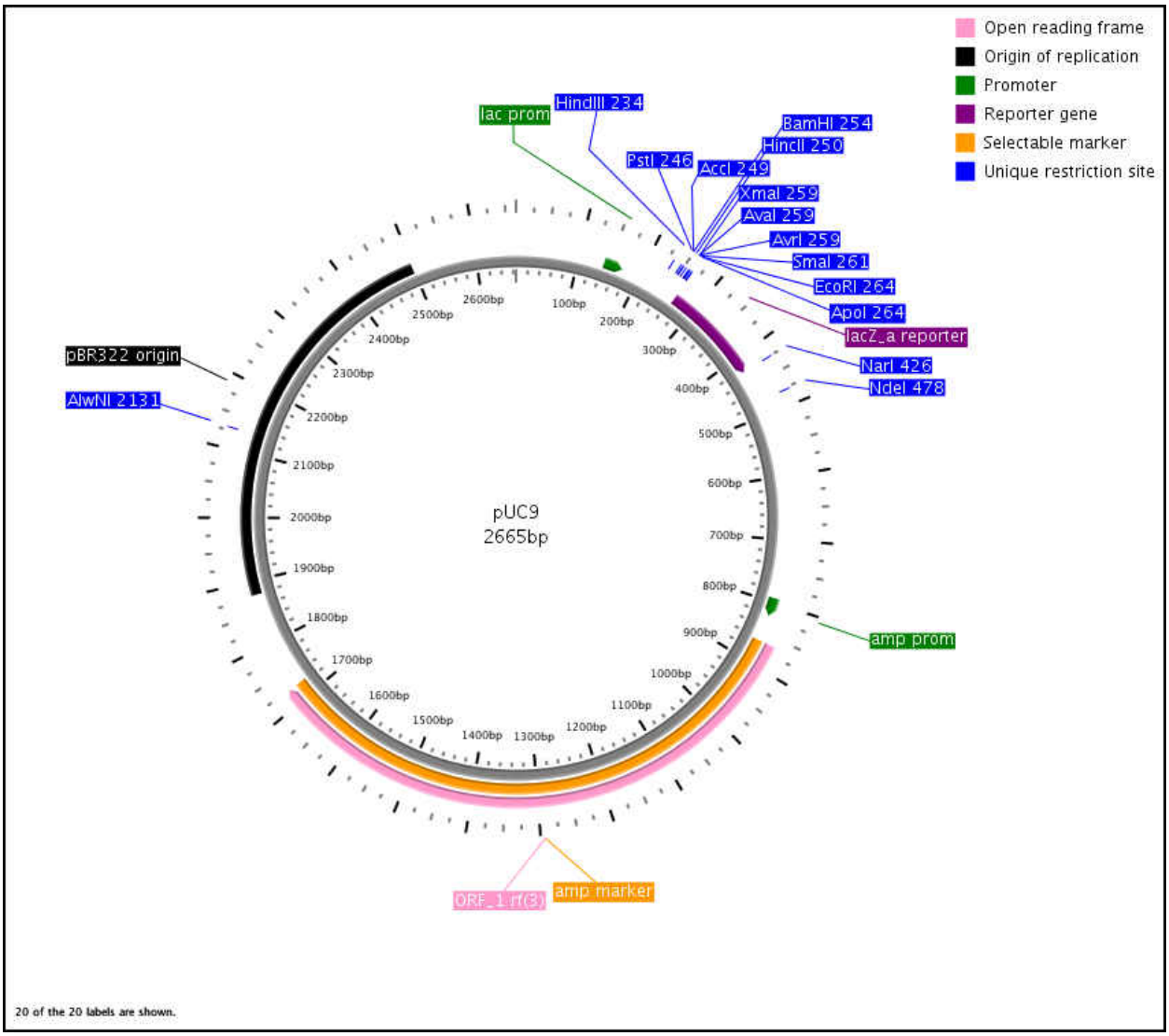


Figure 4: Map of pUC9 plasmid, it has the lac promoter and Hind III, EcoRI sites.

DNA Purification by Electro-Elution

After electrophoresis and staining of DNA sample, the gel was placed on UV box and observed fluorescent DNA bands with a “long” wave UV lamp. Then the DNA band was cut out

with a knife and the gel slice was placed in a 1.5ml microfuge tube. Then a section of dialysis tubing was cut to about 6-8 cm long and the tubing was rinsed in distilled water. After that, end of tubing was closed off one with a dialysis clip. Then the tubing was filled with 0.5X TBE electrophoresis buffer. Then the gel slice was carefully dropped into the openend of tubing and most of the buffer was drained off in the tubing. After that, the other end of tubing was closed off with a clip, then the tubing was placed in a large electrophoresis tank. The tank was filled with 0.5X TBE buffer and the electrophoresis tank was run with tubing at 125 volts for 30-45 min. After the run, the tubing was removed and excess buffer was drained off outside of tubing, then the clips were carefully removed. The end of the tubing was cut off to give a flush end. Then the inside of the edge of the tubing was washed with buffer. All liquid were removed from the tubing to a new 1.5ml microfuge tube and the volume was measured, then 1/10 of this volume of 10M ammonium acetate and 2.5 volume of cold EtOH were added to the tube. They were Mixed and placed in -20^{o}C overnight to precipitate the purified DNA.

Preparation of Electro-Competent Cells and Transformation Conditions

Electro-Competent Cells

Preparation of competent cells for *E. coli* strain DH5α was started by a seed culture. A single colony was inoculated to 2ml of YENB (Tryptone 16g, Yeast extract 10g, NaCl 5g, Distill water 1L) broth was then grown at 37°C with shaking overnight. Next day, 1ml of the seed culture was transferred to 1L of YENB broth in a side-armed flask to grow to early log phase. The Klett is about 0.5-0.9 Klett units. The culture was chilled on ice for 5 min. Then the culture was transferred to a sterile Oakridge tube and centrifuged at 5,000 rpm for 5 minutes at 4°C. The supernatant was discarded and the cell pellet was resuspended in 25mls of cold distilled water

twice. Then the cells were centrifuged at 5, 000 rpm for 5 minutes at 4°C. The supernatant was discarded and the cell pellet was resuspended in 5mls of 10% sterile glycerol in water. Next, the cells were transferred to fresh Oakridge centrifuge tubes and centrifuged at 5,000 g at -4^{o}C. The cells were resuspended in a final volume of 2mls of cold 10% glycerol and 110µl of above resuspended cells were aliquoted into small microfuge tubes and store at -80°C.

Transformation Conditions

The *E.coli* electro-competent cells were taken out from -80°C freezer and they were placed on ice to thaw. After the cells thawed, 1-2µl plasmid DNA was added into 100µl of competent cells and they were mixed together with pipette and transferred into a sterile electroporation cuvette, then they were placed on ice for 2 minutes. After that the mixture was placed into an electroporation cuvette and pulsed at 2500 volts. If the pulse time were at least 3-5 seconds, then 0.5ml of SOC broth were added to the mix. The mixture was transferred into a new tube and then was incubated at 37°C for 1 hour with shaking. Then the transformed cells were spreaded at different dilution factors (undiluted, 10^{-1}, 10^{-2}, 10^{-3} dilutions) on LB agar that contains the antibiotic ampicillin and incubated at 37°C overnight.

RNA Extraction from Bacterial Cells

Bacterial colonies were scraped from 3-5 days old plates and cells were dispersed in 1ml of RNA protect reagent from Qiagen (Kit) in a 1.5 microfuge tube. We let the microfuge tube sit 10 min at room temperature. After cenrifugation for 1min, all liquid were aspirated from cell pellet, then 200ul buffer (0.1M NaoAc, 10mM Tris pH 7.4, 1mM EDTA) + 1mg fresh lysozyme were added. The mixture plus lysozyme were incubated for 10mins. Next, the following solutions were added in a 65^{0}C water bath.

1) 200µl 4M Guanidine
2) 200µl phenol-TE (saturated)
3) 400µl chloroform

The mixture was extracted by inverting the tube many times and the tube was iced for 5min. Then the tube was centrifuged for 10min and the top layer was collected to a new tube. Then it was extracted with 200µl phenol+200µl chloroform once, then was extracted by chloroform twice (400µl each). Then the top layer was collected to a new tube. Finally, 2 vol EtOH was added and tube was stored at -80^{o}C. Next tubes were centrifuged for 15min and the supernatant was discarded and the RNA pellet was redissolved in 0.1x TE buffer.

RT Labeling Reaction for msDNA

msDNA can be specifically labeled with either alpha-^{32}P-dCTP or DIG from a total RNA preparation from bacterial cells. The labeling reaction was as follows: (Table 8)

Table 8. RT Labeling Reaction for msDNA

Ingredients	**Amount**
0.1M DTT	5µl
5x RT buffer	10µl
2.5mM dA,T,G mix	3µl
50uM cold dCTP	0.5µl
Alpha ^{32}P dCTP	(3000Ci/mMol), 1µl
mMLV RT	20units /µl, 1µl

After the reagents were added, they were mixed and incubated at 37^{o}C for 1hr. After that, 50µl DEPC water and 200µl EtOH were added, then they were mixed and placed at -80^{o}C for 15min. Next the tube was centrifuged for 10min, EtOH was removed, the pellet was dried and was resuspended in 15µl distilleded water. Then it was divided:

A: 1) 7µl of label reaction 2) 0.5µl RNase A and incubate at 37^{o}C for 10min

3) Add 7µl of 8M urea dye

B: 1) 7µl of label reaction 2) 7µl of 8M urea dye

And finally the entire sample was loaded on a 4% acrylamide-8M urea gel.

Labeling Reaction Mechanism

Based on the structure of msDNA, msDNA can be specifically labeled with alpha ^{32}P or DIG. The mechanism (Fig.5) was as follows: 3' end of DNA part of msDNA can serve as a self-primer for Reverse Transcription and the RNA part can serve as a template. After addedRT+dNTPs+DIG or (alpha ^{32}P dCTP), new synthesis of DNA incorporates DIG or alpha ^{32}P dCTP and can be formed during the RT extension. Therefore, the labeled newly synthesized DNA can be detected via gel electrophoresis or blotting.

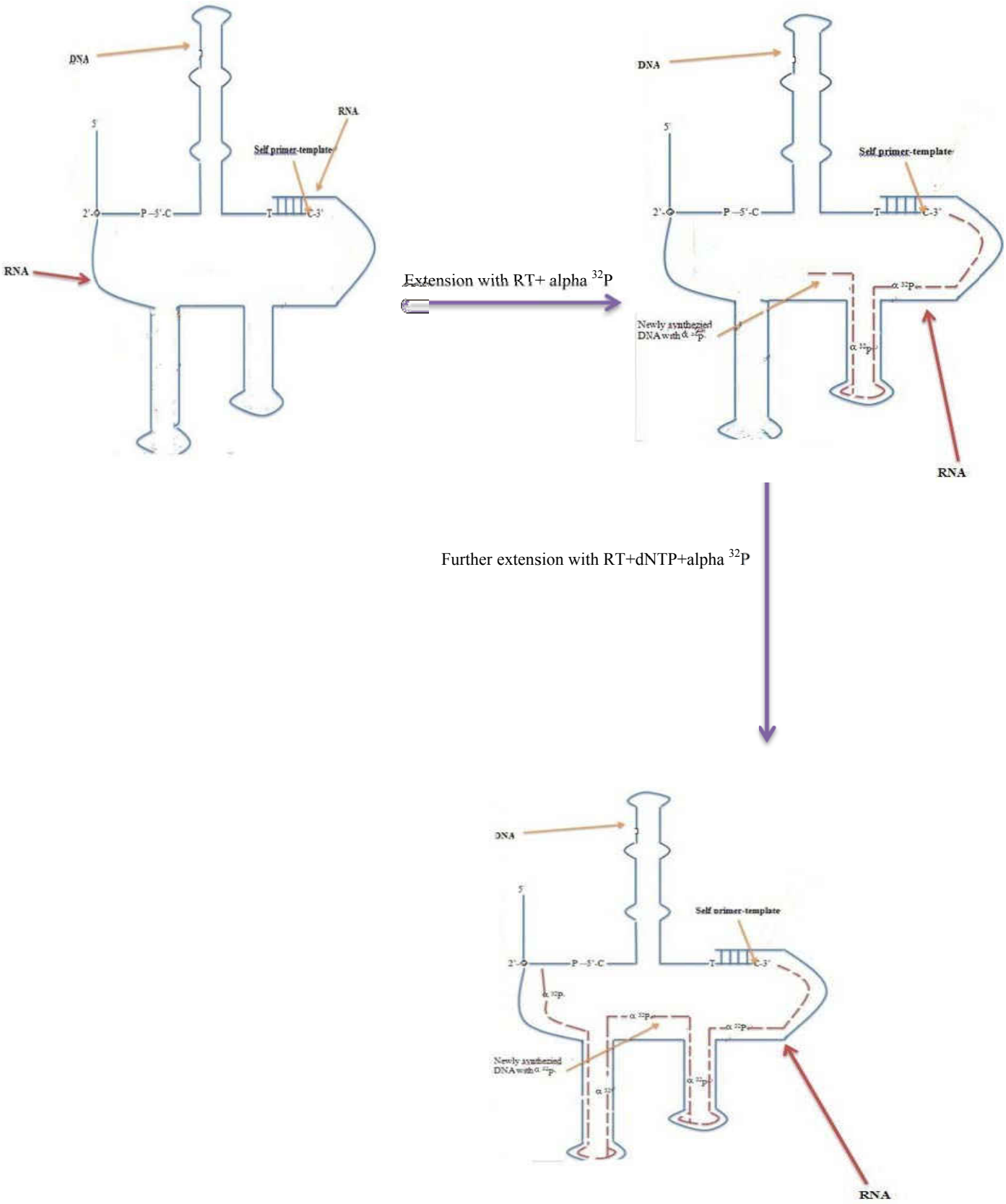


Fig 5: Labeling mechanism of msDNA with alpha ^{32}P dCTP+RT

The gel system for detection of labeled msDNA with alpha ^{32}P dCTP

Four percent acrylamide -8M urea gel was poured between large glass plates with thin spacers as follows: (Table 9)

Table 9. The gel system for detection of labeled msDNA with alpha ^{32}P dCTP

Ingredients	**Amount**
Distilled water	10mls
20% acrylamide solution (19:1)	10mls
10x TBE buffer	10mls
Urea	24.9g
Ammonium persulfate	50mg

After all ingredients were added, urea was dissolved completely with a stir bar in a glass beaker. Next the solution was filtered through a paper filter into a vacuum flask and degased for 10min. Then the solution was transferred to a small beaker and 20µl TEMED was added, mixed, and poured immediately between large glass plates and spacers. It was polymerized for 1hr. Next, after the gel was ready, samples were loaded. The gel was run at 300v in 2x TBE buffer until PBP dye was distant. After the ran, glass plates were split apart, but gel was left on 1 side of the plates and gel plate was placed in a glass dish and was washed in 250mls 10% MeOH+ 10% acetic acid for 10min with gentle shaking solution and the wash was repeated. The gel was rinsed with distill water and was transferred to a premoistured sheet of blotting paper. Then the gel was dried for 1-2 hrs. Finally, an autoradiaogram with x-ray film was set up.

Direct sequence determination of msDNA

We prepared oligo $(dC)_{18}$ as a sequencing primer and set-up "tailing" reaction on msDNA with terminal transferase. All the ingredients were as follows: (Table 10)

Table 10. Direct sequence determination of msDNA

Ingredients	Amount
msDNA (gel purified)	15µl
dGTP(100mM)	1µl
TdT 5X buffer	4µl
TdT enzyme	20units, 1µl

After all the required reagents were added, they were incubated reaction at 37^{o}C for 1hr. The reaction was stopped by heating at 70^{o}C for 10min. Next, "tailed" msDNA was gel purified via an acrylamide gel. After msDNA was purified, tailed msDNA for sequencing with oligo $(dC)_{18}$ as a primer was set up .

Sequencing Mechanism of TdT (Terminal Deoxylnucleotide Transferase)

TdT catalyses the addition of nucleotides to the 3' free terminus of a msDNA molecule. Unlike most DNA polymerases, it does not require a template. Based on the characteristics of msDNA Structure, the preferred substrate of this enzyme is a 3'-overhang, but it can also add nucleotides to blunt or recessed 3' ends. Therefore, TdT can add nucleotides to the 3' end of msDNA (Fig.6).

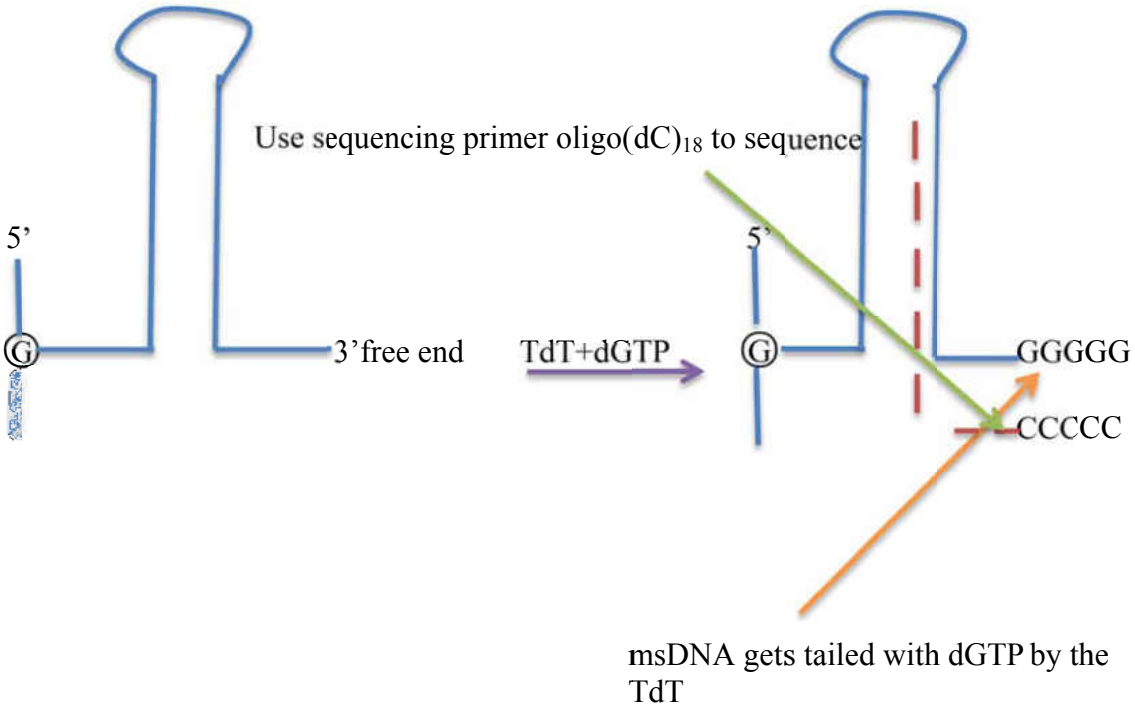


Fig.6: Tailing reaction mechanism for msDNA

CHAPTER 3

RESULTS

The amino acid sequence of the retron RT from *Myxococcus xanthus* (Mx162) was used as a query sequence to perform a BLASTp (protein to protein) search of the GenBank database (most recently performed on 8/9/12). The first 100 matches from many different bacterial genomes were used for further analysis (out of many additional matches). These 100 protein sequences were used to produce a large multiple sequence alignment (using the constraint-based multiple alignment tool; Cobalt from NCBI). Multiple sequence alignments provided a measure of the relative similarity of the query protein with the various matches in the database. The multiple alignment can also produce a hierarchy of similarity with the query protein that can be used to generate a phylogenetic tree of similarity or evolutionary relatedness.

From the large (100 protein) multiple sequence alignment (not shown) it is apparent that there is a great deal of amino acid sequence variation among the proteins that match the Mx162 retron RT. However, all the postulated retron RTs do share 7 short regions of high similarity designated domains I through VII (see Fig.7). These 7 domains correspond to conserved structures in the folded protein (the so called "fingers, palm, and thumb") and are shared by all RTs from eukaryotic viruses to bacteria (16). Also, from a phylogenetic tree produced from this large alignment, most of the protein matches to the query protein (Mx162 RT) fall into 3 broad groups: 1) Those protein matches that are very similar to the query sequence and come from other myxobacteria or related families, 2) those protein matches that are quite distant from the query protein and 3) those proteins that broadly fall in between these 2 groups. Regardless how distant or similar these protein matches appear to be with the Mx162 RT, they are all predicted to be retron type RTs. If this is true, then they all should produce msDNA.

From the list of 100 protein BLAST matches with the Mx162 RT, a small number were chosen to test the hypothesis that they should all produce msDNA. This select list of proteins is shown in the smaller protein sequence alignment of Fig.7. The first protein in this alignment is the query sequence P23072 and is the Mx162 retron encoded RT from *Myxococcus xanthus* (a member of the delta-proteobacteria). The next 6 proteins are all known retron RTs from different genera of the myxobacteria and have been shown previously to produce msDNA (P23071 retron Mx65 from *M. xanthus* (17); YP_004664596 protein from *M. fulvus* (18); AAA66173 RT from *Melittangium lichenicola*(18); ZP_01463804 retron Sa163 from *Stigmatella aurantiaca* (19); YP_001612552 RT from *Sorangium cellulosum* (18); AAL40743 retron RT Ne144 from *Nannocystis exedens* (20). With the possible exception of the RT from *Nannocytis* (AAL40743), these 6 proteins are the most similar to the query protein (see phylogenetic tree, Fig.8). Also similar to these myxobacterial RTs is a protein from *Gemmata obscuriglobus*, a member of the *Planctomycetes* (21). Based on this similarity, it is predicted that the *Gemmata* protein will produce an msDNA similar to Mx162 from *M. xanthus*.

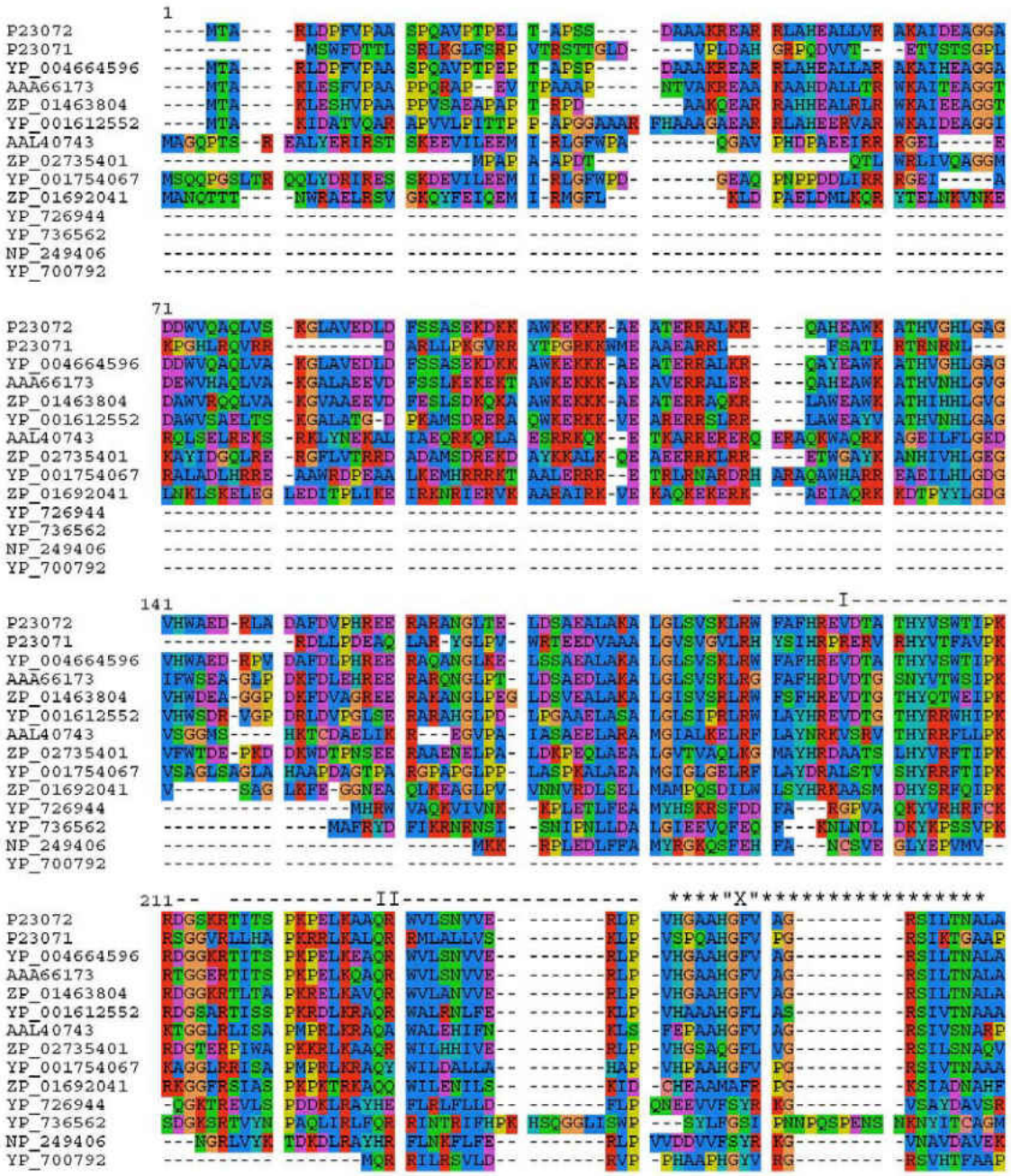


Fig.7: RT protein sequence alignment. The query sequence P23072 is the Mx162 retron RT from *Myxococcus xanthus*. P23071 is a second retron RT (Mx65), also from *M.xanthus*. YP_004664596 is the retron RT from *M.Fulvus*, AAA66173 is the retron RT from the myxobacterium, melittangium lichemicola. ZP_01463804 is the Sa163 retron RT from the myxobacterium *Stigmattella aurantiaca*. YP_001612552 is an RT protein from the myxobacterium *Sorangium cellulosum* AAL40743 is the Na144 retron RT from the myxobacterium *Nannocystis exedens*. ZP_02735401 is a protein from *Gemmata obscuriglobus*. YP_001754067 is a protein from *Methylobacterium radiotolerans*. ZP_0169201 is a protein from *Microscilla marina*. YP_726944 is a protein from *Ralstonia eutropha*. YP_73562 is a protein from *Shewanella sp*. MR7. NP_249406 is a protein from *Pseudomonas aeruginosa* PAO1. And YP_700792 is a protein from *Rhodococcus jostii* RHA1.

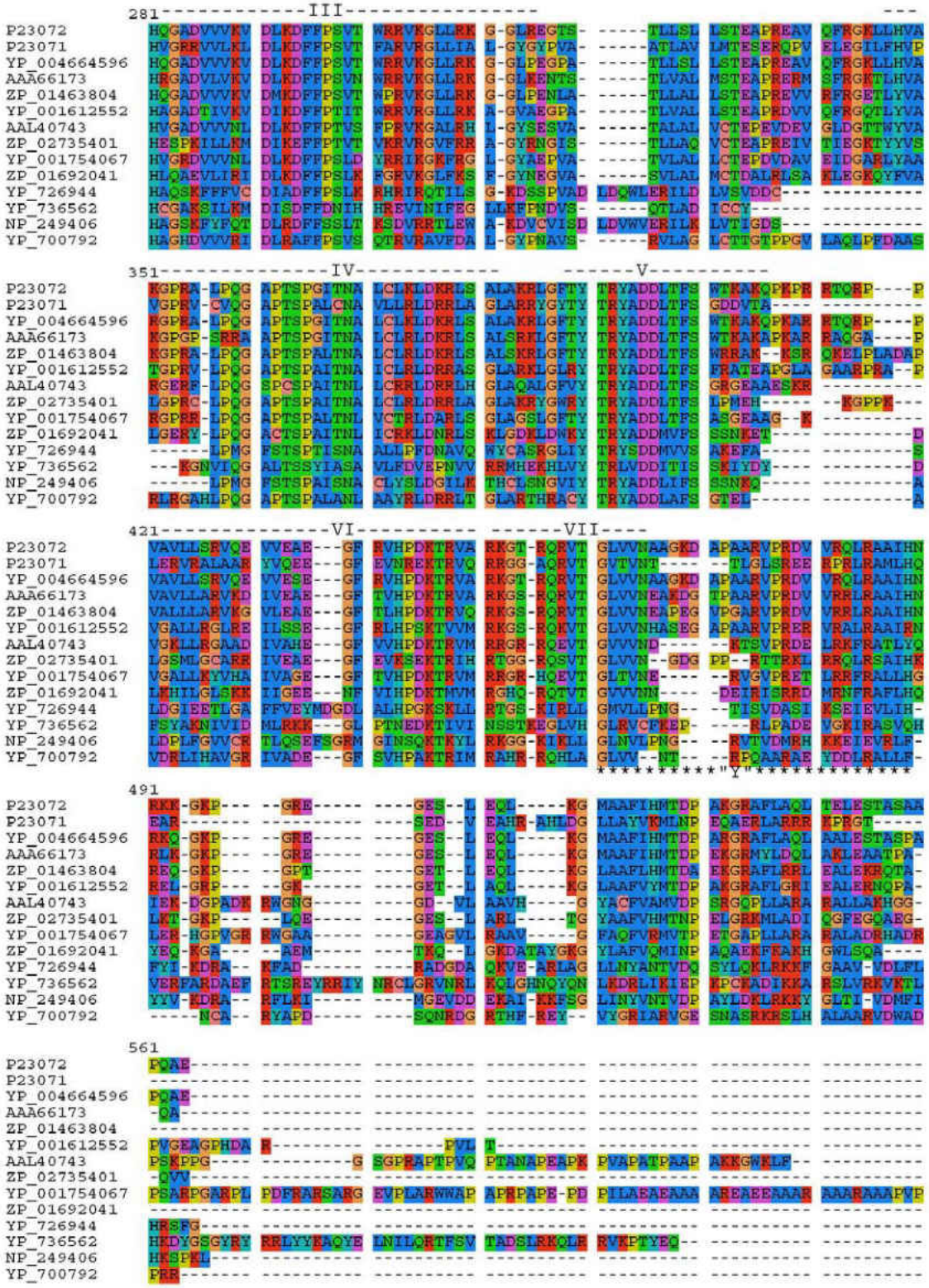


Fig.7: (cont.)

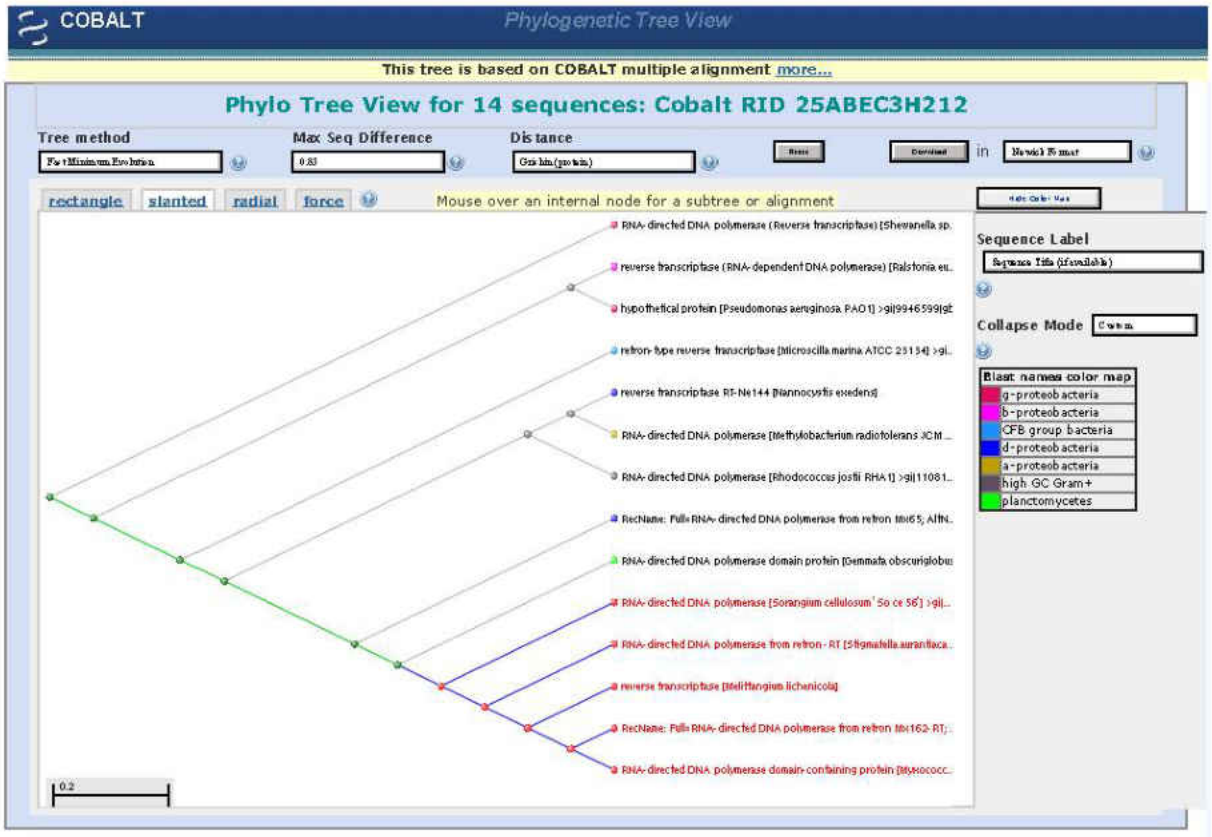


Fig.8: Phylogenetic tree showing the relationship between the query protein and unknown RT proteins. The phylogenetic tree was constructed from the protein sequence alignment of Fig.7. by the cobalt program from NCBI web page.

Falling in the middle region of relatedness to the query protein are retrons from *Methylobacterium radiotolerans* (an alph-proteobacterium) and *Microscilla marina* (a Bacteroidetes-Cytophagia group member). Interestingly, like the retrons from the myxobacteria, these 2 retron RTs have a long N-terminal extension prior to the first conserved domain I (see Fig.7). Again, it is predicted that the retrons from *Methylobacterium* and *Microscilla* will produce large msDNAs similar to Mx162.

Finally, 3 proteins matches were chosen because they are quite distantly related to the query protein (see phylogenetic tree, Fig.8). These were proteins from *Ralstonia* (YP_726944, Beta-proteobacteria), *Shewanella* sp. (YP_736562, gamma-proteobacteria), and

Pseudomonas aeruginosa (NP_249406, gamma-proteobacteria). These 3 proteins have significant variation in their amino acid sequence at some conserved regions. For example, at the highly conserved "YADD" box sequence in domain V, the protein from *Ralstonia* is "YSDD" and the protein sequence from *Shewanella* is "LVDD" (see Fig.7). In addition, a region known as "X" in the RT protein has been proposed to be a retron specific sequence, with a somewhat conserved "AHGF" sequence followed by a "RSI" box. For the proteins from *Ralstonia*, *Shewanella*, and *Pseudomonas*; however, there is a great degree of sequence variation in this region. Likewise, a region designated "Y" (Fig. 7) is also a possible retron RT specific sequence with a conserved "VTG" box followed by a "RK/A" sequence. Here again, there is great sequence variation at this position in the RT proteins from *Ralstonia* ("LLG…EV"), *Shewanella* ("VHG…RA"), and *Pseudomonas* ("LLG…EV"). Based on the significant differences in the amino acid sequence of these 3 proteins to the query protein, there is some question whether these proteins are actually retron encoded RTs or whether they have some other unknown function.

Detection of msDNA

Three different methods were used to detect the production of msDNA by the retron RTs shown in the sequence alignment of Fig 7. For the first method, total DNA was prepared from each host bacterium where the retron RT is originally found. The DNA was separated via electrophoresis on a 5% acrylamide gel to detect small satellite DNA. In the second method, retron encoding DNA from each bacterium was cloned into the plasmid pUC9 to detect heterologous expression and production of msDNA in *E.coli* cells. Finally, a sensitive radioactive method was used to specifically radio-label msDNA from total RNA prepared from

each bacterial strain. Table 11 shows a summary of the results obtained for each bacterial strain using these 3 methods.

Table 11. Summary of Results from 3 different methods used to detect the presence of msDNA

Strain names	Satellite DNA on gel	Heterologous expression in *E.coli*	Radioactive RT label
Myxococcus.xanthus(control)	+++	+++	+++
Gemmata obscuriglobus	---	N/A	++
Methylobacterium radiotolerants	++	N/A	++
Microscilla marina	++	+	N/A
Ralstonia etropha H16	---	---	+
Shewanella MR-7	---	++	+
Pseudomonas (PAO1)	---	N/A	---
Rhodococcus (RHA1)	---	---	---

+++: Very strong. ++ : Strong. +: Weak. N/A: Not done.

Gemmata obscuriglobus

Because it shows close similarity to the Mx162 retron RT of *M. xanthus*, the protein ZP02735401 from *Gemmata* is predicted to be a retron RT that likely produces a large msDNA similar to Mx162. Gel electrophoresis analysis of total DNA prepared from *Gemmata* cells did not reveal any obvious satellite DNA band (Fig.9, lane 7). Heterologous expression of the *Gemmata* RT gene in *E. coli* was not done due to the difficulty in isolating genomic DNA from this bacterium. However, specific radioactive labeling with the RT extension method shows a strong band migrating at about the same size as the Mx162 msDNA from *M. xanthus* (Fig.10, lane 12-Gemmata, and lane 3-Mx162). When the labeling reaction is treated with RNase, there is a characteristic shift to a faster migrating band, indicating the presence of RNA as well as DNA in this molecule. The appearance of multiple bands after RNase treatment may be due to

intermediate extension of the 3'end of the msDNA by Reverse Transcription during incorporation of ^{32}P-dCTP in the RT extension reaction (see methods).

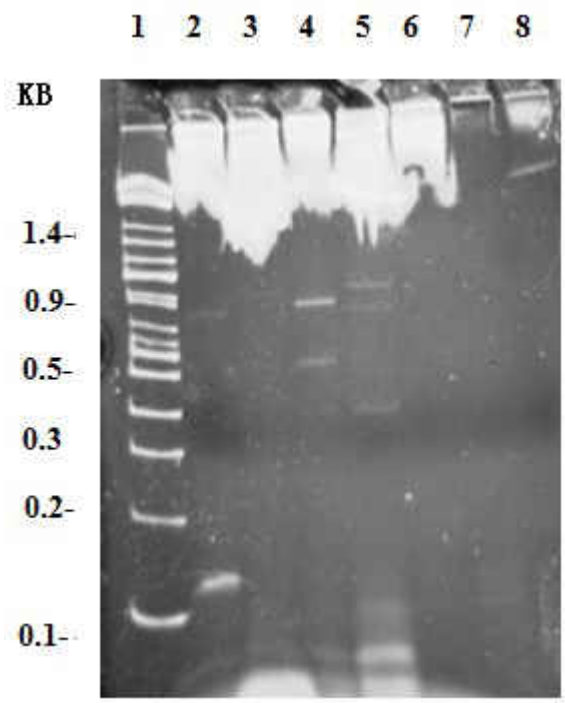


Figure 9: Detection of small molecule weight Satellite DNA via electrophoresis on a 5% acrylamide gel. Lane 1 is a 100bp molecular weight standard. Lane 2 contains plasmid DNA from *E.coli* cells over-expressing a plasmid clone of the Mx162 retron from *M.xanthus*. Lane 3 contains DNA from *E.coli* cells expressing a plasmid clone of the RT from *Rhodococcus jostii* RHA1. Lane 4 contains DNA from *E.coli* cells expressing a plasmid clone of the RT from *Microscilla marina*. Lane 5 contains expressing a plasmid clone of the RT from *Shewanella* MR-7. Lane 6 contains expressing a plasmid clone of the RT from *Ralstonia eutropha*. And Lane 7 contains plasmid DNA prepared from *Gemmata obscuriglobus* cells. Lane 8 contains a plasmid DNA prepared from *E.coli* cells containing just the pUC9 plasmid with no cloned insert DNA.

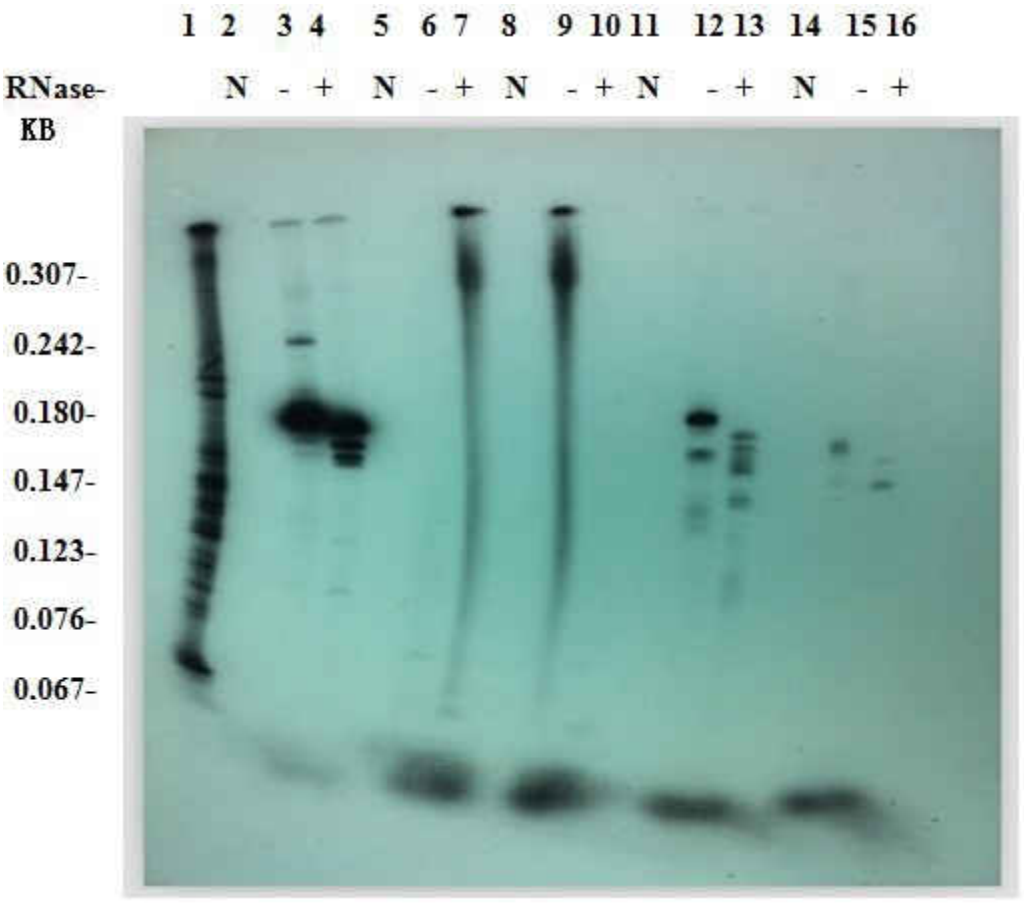


Figure 10: Auto radiograph of radiolabeled msDNA by the RT extension reaction. N: nothing. --: no Rnase added. +: RNase added. Lane 1 is radiolabeled pBR322 cut with mspI used as molecular weight standard. Lane 2 has nothing. Lane 3 is the radiolabeled DZF1 total RNA without RNase. Lane 4 is the radiolabeled DZF1 total RNA with RNase. Lane 5 has nothing. Lane 6 is the radiolabeled *Psdudomonas* total RNA without RNase. Lane 7 is the radiolabeled *Rhodococcus* total RNA with RNase. Lane 8 has nothing. Lane 9 is the radiolabeled *Rhodococcus* total RNA without RNase. Lane 10 is the radiolabeled *Psdudomonas* total RNA with RNase. Lane 11 has nothing. Lane 12 is the radiolabeled *Gemmata* total RNA without RNase. Lane 13 is the radiolabeled *Gemmata* total RNA with RNase. Lane 14 has nothing. Lane 15 is the radiolabeled *Methylobacterium* total RNA without RNase. Lane 16 is the radiolabeled *Methylobacterium* total RNA with RNase.

Methylobacterium radiotolerans

Like *Gemmata*, the RT protein from *Methylobacterium* (YP001754067) also contains a long N-terminal region prior to domain I and may produce a large msDNA. Indeed, gel electrophoresis of total DNA prepared from *Methylobacterium* cells shows a distinct satellite DNA band migrating at about 100-110 bp in size (Fig.11, lane 1). The band near the top of lane 1 (Fig.11) appears to be a small plasmid. A band migrating at about 140bp can be seen clearly after

radioactive labeling with the RT extension method (Fig.10, lane 15). Treatment of the label reaction with RNase produces a characteristic faster migrating band (Fig.10. Lane 16). Therefore, *Methylobacterium* is positive for msDNA production. After purification, several attempts were made to determine the DNA sequence of the msDNA band shown in Fig.11 but were not successful (see *Microscilla* below). Heterologous expression of the *Methylobacterium* RT has not yet been done.

Figure 11: Detection of Satellite DNA from *Methylobacterium radiotolerans*. Lane 1 is total plasmid DNA prepared from *Methylobacterium radiotolerans,* the small satellite DNA can be seen. Lane 2 is the 100 molecular weight standard.

Rhodococcus josti RHA1

The RT protein (YP700792) from *Rhodococcus* is similar to the known retron RTs from the myxobacteria. However, this protein is truncated at the N-terminus such that domain I and part of domain II are missing (Fig.7). This RT protein is, therefore, probably not functional and will likely not make msDNA. No msDNA was detected on electrophoresis gels of total DNA from *Rhodococcus* cells (Fig.12, lane 2). The gene and some flanking DNA that code for the RT protein was amplified from the chromosome of *Rhodococcus* using the primers: RHA1 retron 53 and RHA1 retron 1593. These primers amplify a DNA of about 1500bp in size (Fig.13C, lane

2,3). This amplified DNA was cloned into the *E. coli* plasmid pUC9 at the *Eco*RI and *Hin*dIII sites using the same restriction sites engineered into the primers, respectively. A map of the cloned DNA from *Rhodococcus* is shown in Fig.14. Proper orientation of the cloned DNA, with the lac promoter upstream of the RT encoding ORF (330aa ORF, Fig.14) was confirmed by restriction site mapping. The plasmid was digested with EcoRI plus BamHI (Fig.12, lane 5) and with EcoRI plus SacI (Fig.12, lane 6) that produced the expected sized restriction fragments for the plasmid map shown in Fig.14.

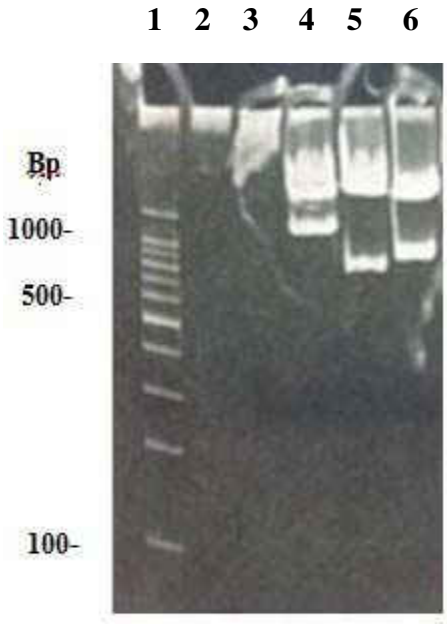


Figure 12: The plasmid clone of the RT gene from *Rhodococcus jostii* and expression in *E.coli*. Lane 1 is the 100 bp ladder. Lane 2 is a total plasmid midi prep from *Rhodococcus*, there is no satellite DNA production. Lane 3 contains plasmid DNA from *E.coli* cells expressing the plasmid clone of the *Rhodococcus* retron RT gene. Lane 4 is the retron of clone in pUC9 cut with in HindIII+EcoRI. Lane 5 is the retron of clone in pUC9 cut with EcoRI+BamHI. Lane is the retron of clone in pUC9 cut with EcoRI+SacI.

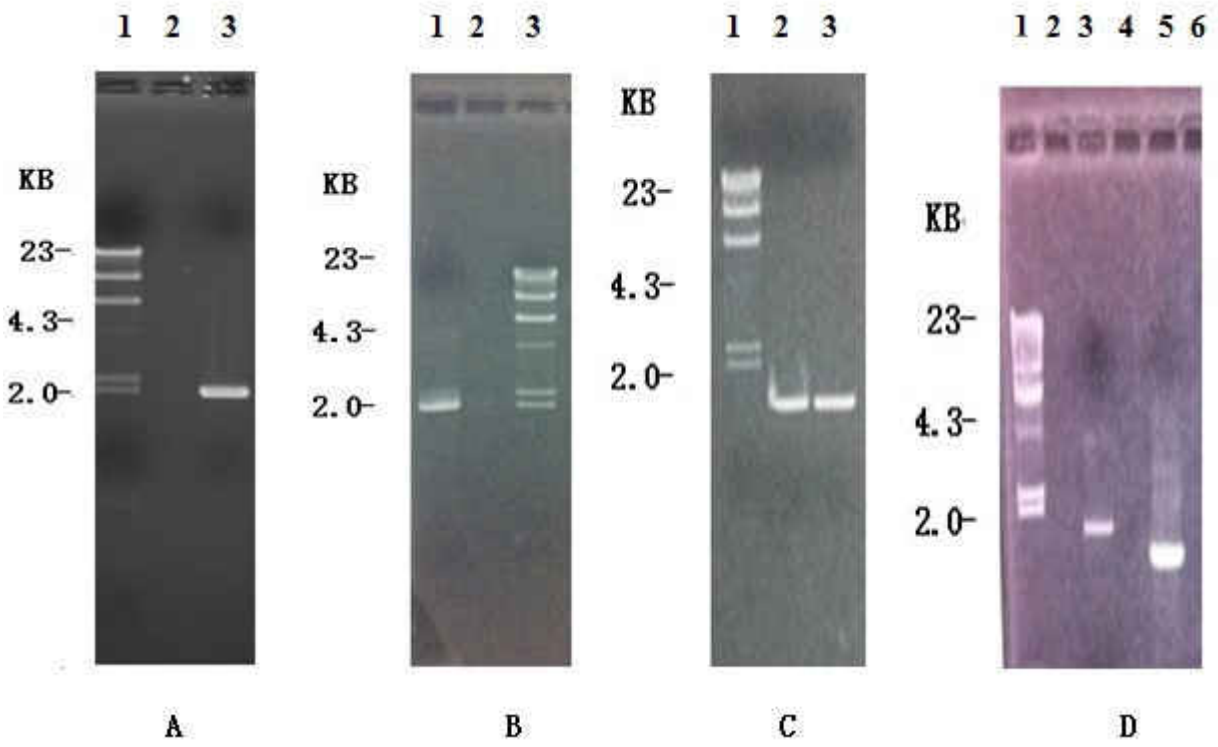


Figure 13: PCR amplification of retron RT genes from different bacteria. (A) Lane 1 contains lambda DNA cut with HindIII as a molecular weight standard. Lane 2 has nothing, lane 3 contains PCR amplified DNA of the Mx162 retron from *M.xanthus*. (B) Lane 1 contains the amplified DNA of the *Microscilla marina* RT gene, Lane 2 has nothing, lane 3 contains lambda DNA cut with HindIII. (C) Lane 1 contains lambda DNA cut with HindIII, lane 2 and 3 contain the amplified DNA of the RT gene from *Rhodococcus jostii* RHA1. (D) Lane 1 contains lambda DNA cut with HindIII, lane 2 has nothing, lane 3 contains the PCR product of the RT gene from *Shewanella* sp. MR-7, lane 4 has nothing, and lane 5 contains the amplified DNA of the RT gene from *Ralstonia eutropha* H16.

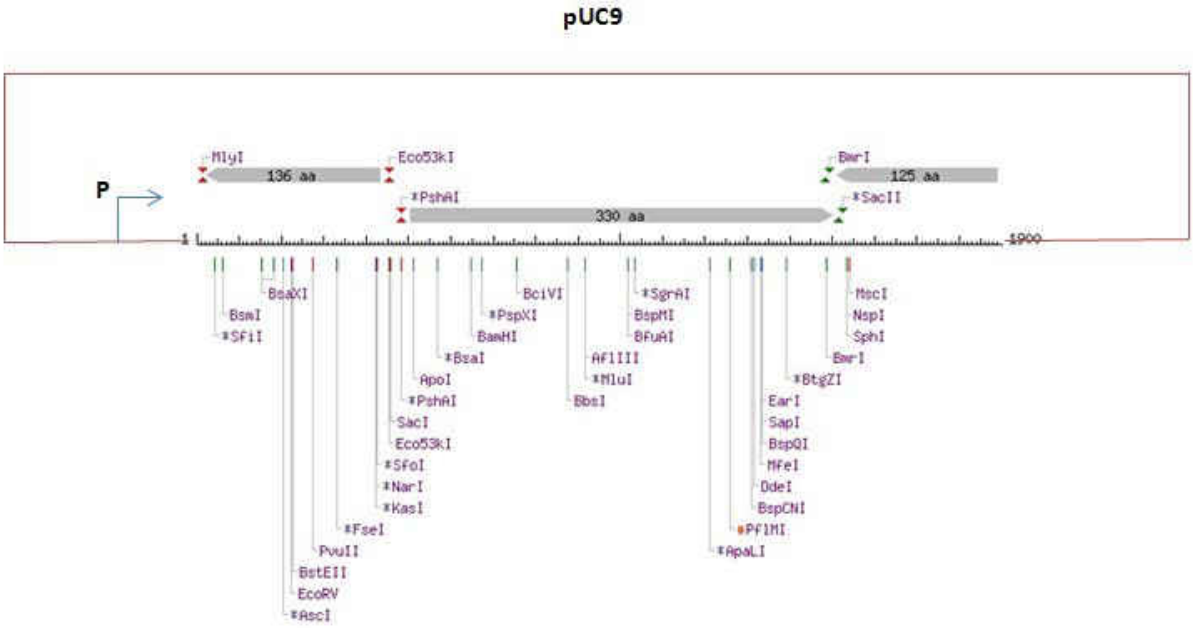


Figure 14: The retron RT gene from *Rhodococcus josti* cloned into pUC9 plasmid, the amplified DNA from Fig 13C, lane 2 was cloned into the EcoRI-HindIII sites of the plasmid pUC9. Single cutting restriction sites are shown for the cloned *Rhodococcu*s DNA.

The cloned *Rhodococcus* gene was over-expressed on the high copy number plasmid pUC9 in *E. coli* cells under 1% IPTG induction. As a positive control, the known retron Mx162 from *M. xanthus* was also cloned downstream of the lac promoter in the pUC9 plasmid and expressed in *E. coli* cells. Total DNA prepared from *E. coli* cells containing the Mx162 clone showed a strong satellite DNA band at about the expected size of 162 nucleotides (Fig.9, lane 2). Total DNA prepared from *E. coli* cells containing the *Rhodococcus* DNA clone in pUC9 did not show any satellite DNA even after over loading the gel with the DNA preparation (Fig.9, lane 3 and Fig.12, lane 3).

The most sensitive way to detect the presence of msDNA is by the radioactive labeling method (the RT extension method). However, no clear signal was produced on the autoradiogram from total RNA extracted from *Rhodococcus* cells (Fig.10, lane 7 with RNase treatment and lane 9 with no RNase treatment). As expected, therefore, msDNA was not detected with any of the 3 methods.

Pseudomonas aeruginosa PAO1

For the RT protein from *Pseudomonas* (NP249406), there was a great degree of sequence variation in postulated retron specific regions "X" and "Y" of the aminoacid sequence (Fig.7). Total DNA and plasmid DNA were extracted from PAO1 and run a 5% acrylamide electrophoresis gel. There was no msDNA production observed (Fig.15, lane 6). Also when using the radioactive method to label msDNA from total RNA from PAO1, there was no msDNA production observed. (Fig.10, lane 6 without RNase, lane 10 with RNase).

Shewanella sp. MR-7

The *Shewanella* protein YP736562 contains an unusual "LVDD" amino acid sequence rather than the super conserved "YADD" box sequence in domain V of the RT. It is possible, therefore, that this RT does not produce msDNA. Two types of DNA extractions were prepared from *Shewanella* cells. First, a total DNA preparation that includes the chromosome was done and, second, a plasmid DNA preparation that excludes the chromosome was also prepared. Both DNA preparations were analyzed via gel electrophoresis, but no small satellite DNA was detected in either DNA sample (Fig.15, lane 4 with total DNA and lane 8 with plasmid DNA).

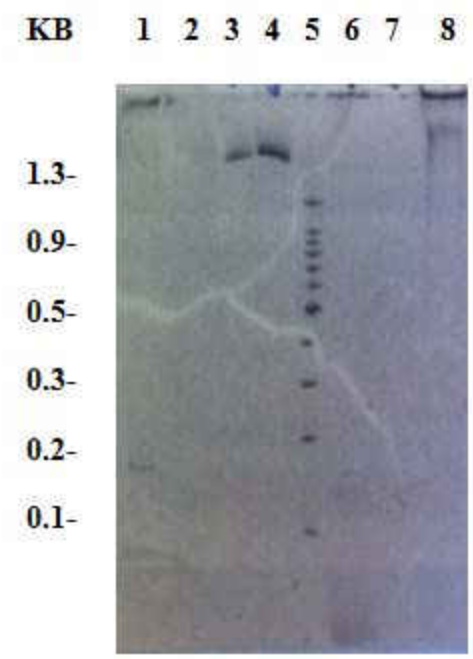


Figure 15: Detection of low molecular weight Satellite DNA via gel electrophoresis. Lane 1 is *Methylobacterium sp* chromosome. Lane 2 is the total chromosome prep of *Pseudomonas*. Lane 3 is the total chromosome prep of *Ralstonia etropha* H16. Lane 4 is the total chromosome prep of *Shewanella* MR-7. Lane 5 is the 100 bp molecular weight standard and lane 6 is the total plasmid-midi-prep of *Pseudomonas*. lane 7 is the total plasmid-midi-prep of *Ralstonia etropha* H16. Lane 8 is the total plasmid-midi-prep of *Shewanella* MR-7. Only the DNA prep from methylobacterium appears to show an msDNA band (Negative image).

Primers were designed to amplify a 1,670bp DNA fragment (Fig. 13D, lane 3) encoding the *Shewanella* RT. This amplified DNA was cloned into the *Hin*dIII plus *Eco*RI restriction sites in pUC9 using the same restriction sites engineered in the primers (*Shewanella*1 and *Shewanella*1670) (Fig.16). The cloned DNA is downstream of the lac promoter for over expression in *E.coli* cells. *E.coli* cells containing the pUC9 clone were induced with IPTG and then plasmid DNA was extracted. Gel electrophoresis of a heavily loaded plasmid DNA sample from these induced *E.coli* cells shows several light bands ranging in size from 400 to 800 bp. (Fig.9, lane 5). These DNA bands are probably too large to be msDNA but could be concatemeric forms of msDNA that have been known to form when a retron is expressed from a high copy number plasmid like pUC9 (Lampson, observation).

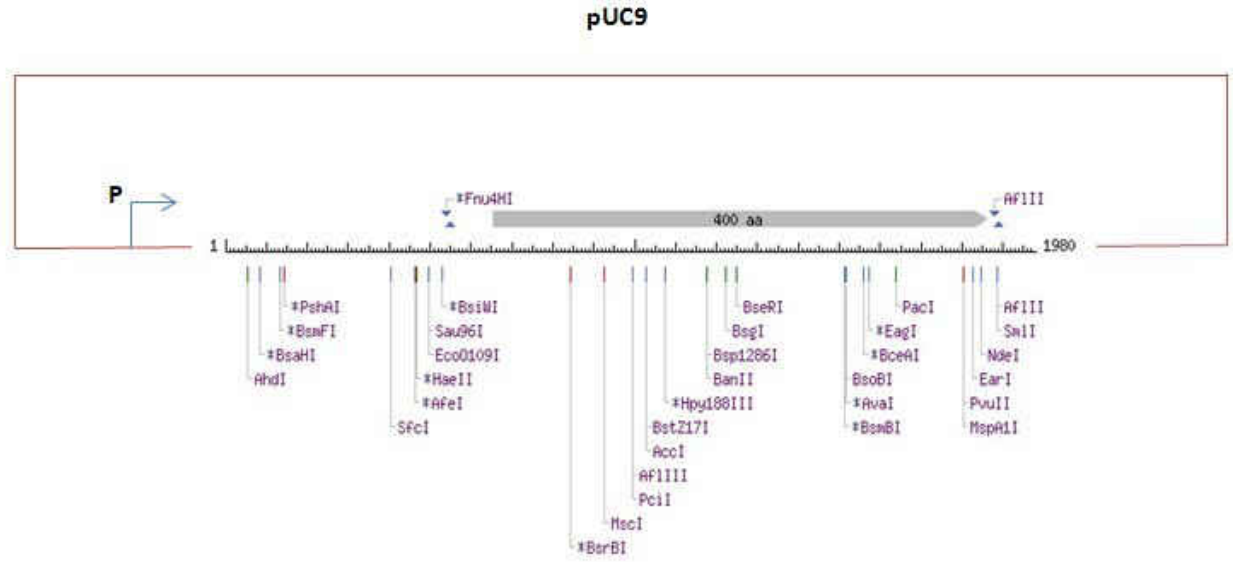


Figure 16: The retron RT gene from *Shewanella* cloned into pUC9 plasmid, the amplified DNA from Fig 13D, lane 3 was cloned into the EcoRI-HindIII sites of the plasmid pUC9. Single cutting restriction sites are shown for the cloned *Shewanella* DNA.

Total RNA prepared from *Shewanella* cells was used to specifically label msDNA with ^{32}P (by the RT extension method). RNA from *M. xanthus*, which contains the Mx162 retron was used as a positive control. As shown in Fig.17A, Mx162 msDNA produced a very strongly labeled band that shifts to a faster migrating DNA molecule after RNase treatment (lane 1 and 2 respectively). No labeled band was detected from the RNA sample prepared from *Shewanella* cells. However, when the autoradiogram was allowed to develop for a longer period (19 hours), a very small satellite DNA band was clearly visible (Fig.17B, lane 4). After RNase treatment, the labeled band shifted to a much smaller sized DNA molecule indicating the presence of RNA. This is a characteristic of msDNA (Fig.17B, lane 5). Thus, *Shewanella* appeared to produce a very small msDNA.

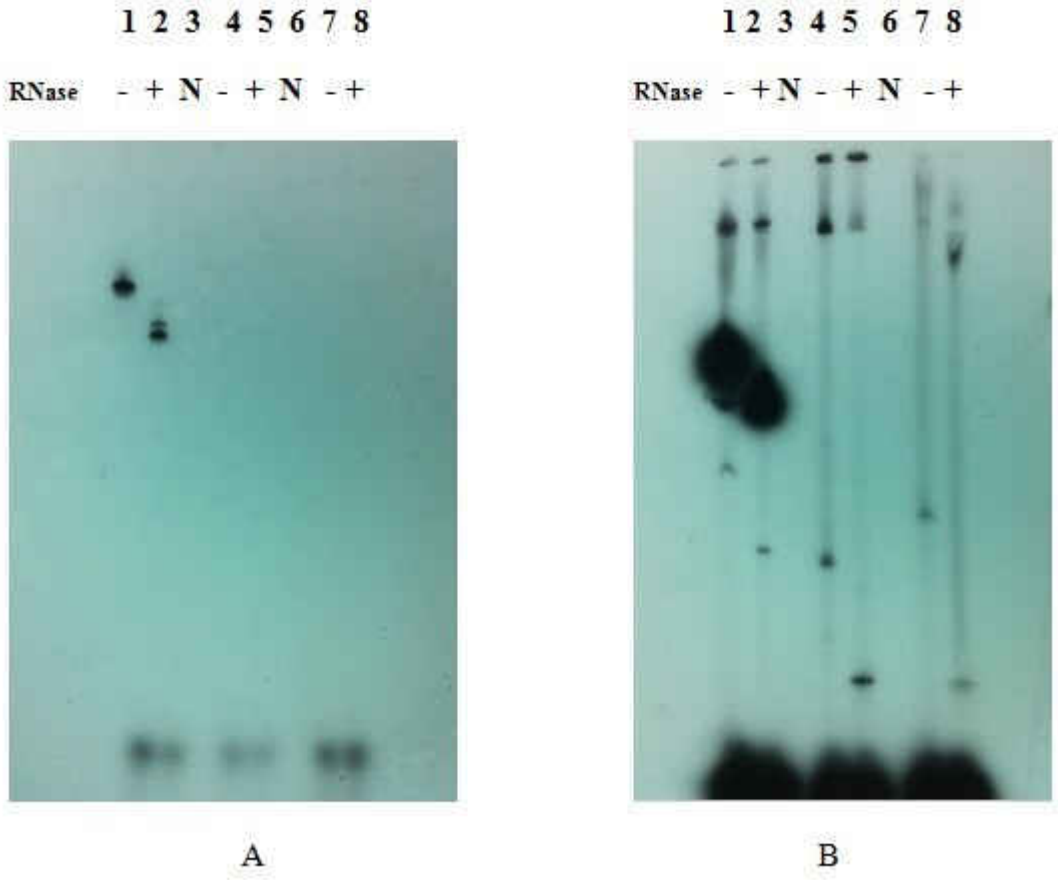


Figure 17: Specific ^{32}P labeling of msDNA by the RT extension method. Lane 1 contains Mx162 msDNA ^{32}P labeled from an RNA preparation from *M.xanthus cells*. Lane 2 is the same labeling reaction but treated with Rnase. Lane 3 has nothing. Lane 4 contains the RT extension (labeling) reaction using total RNA prepared from *Shewanella* sp. MR-7 cells. Lane 5 is the same labeling reaction treated with Rnase. Lane 6 has nothing. Lane 7 contains the RT extension reaction using total RNA prepared from *Ralstonia eutropha* H16 cells. Lane 8 is the same labeling reaction treated with RNase. (A) the auto radiograph exposed for 2hrs. (B) the same autoradiograph exposed for 19hrs. +: RNase added. -: no RNase added. N:nothing.

Ralstonia eutropha H16

The *Ralstonia* protein YP726944 also contains an atypical "YSDD" sequence instead of the conserved YADD box sequence in domain V of this RT. No satellite DNA band was observed when total DNA was separated on electrophoresis gels (Fig.15, lane3-total DNA and lane 7-plasmid DNA). Primers Ralstonia1 and Ralstonia1400 were used to amplify a 1530bp DNA containing the *Ralstonia* RT protein (Fig.13D, lane 5). This amplified DNA was cloned

into pUC9 with *Hin*dIII and *Eco*RI restriction sites engineered into the PCR primers. This plasmid clone (Fig.19) was used to over-express the *Ralstonia* retron in *E. coli* cells. However, despite a heavy amount of DNA sample loaded on the gel, plasmid DNA from these *E.coli* cells did not show any small satellite DNA band (Fig.9, lane 6). Finally, radioactive labeling of total RNA from *Ralstonia* cells was done by the RT-extension method. Similar to *Shewanella*, there appears to be a small msDNA produced in these *Ralstonia* cells (Fig.17B, lane 7-no RNase, and lane 8-with RNase treatment).

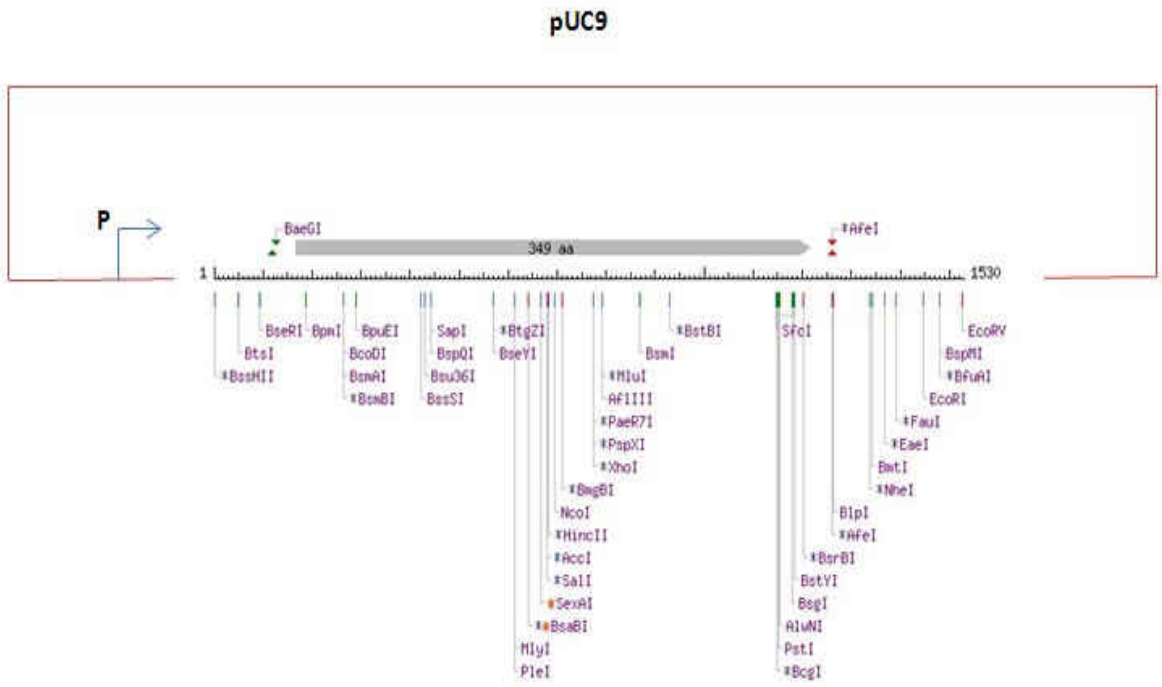


Figure 18: Map of the plasmid clone containing the RT gene from *Ralstonia eutropha*. The amplified DNA from Fig.13D, lane 5 was cloned into the EcoRI-HindIII sited of the plasmid pUC9. Single cutting restriction sites are shown for the *Ralstonia* DNA insert.

Microscilla marina

Total plasmid DNA was extracted from *Microscilla* cells and then analyzed by gel electrophoresis. Separation of this DNA sample on a 5% acrylamide gel showed a very strong band running about 240bp in size (Fig.19, lane 2). Because this satellite DNA is the only thing

visible in this small size range, it is very likely to be an msDNA. Also interesting, the size of this msDNA is the largest ever reported at about 240bp.

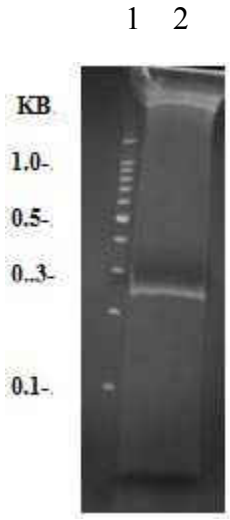


Figure 19: Detection of 240 bp Satellite DNA from *Microscilla marina*. Lane 1 contains the 100 bp molecular weight standard. Lane 2 contains plasmid (midi-prep) DNA from *Microscilla* cells.

The primers Microscilla217 and Microscilla842 successfully amplified a DNA molecule of the expected size of about 2,200bp (Fig.13B, lane 1). This amplified DNA was successfully cloned into the plasmid vector pUC9 using the restriction sites *Pst*I and *Eco*RI that had been engineered into the PCR primers. A map of this plasmid clone is shown in Fig.20. This plasmid clone was over-expressed in *E.coli* cells. A large plasmid DNA preparation from *E.coli* cells containing this plasmid was separated by gel electrophoresis. This gel shows 2 (or possibly 3) strong satellite DNA bands running at about 800bp and 600bp in size (Fig.9, lane 4). These large size bands could be concatemeric forms of the 240bp msDNA of *Microscilla*.

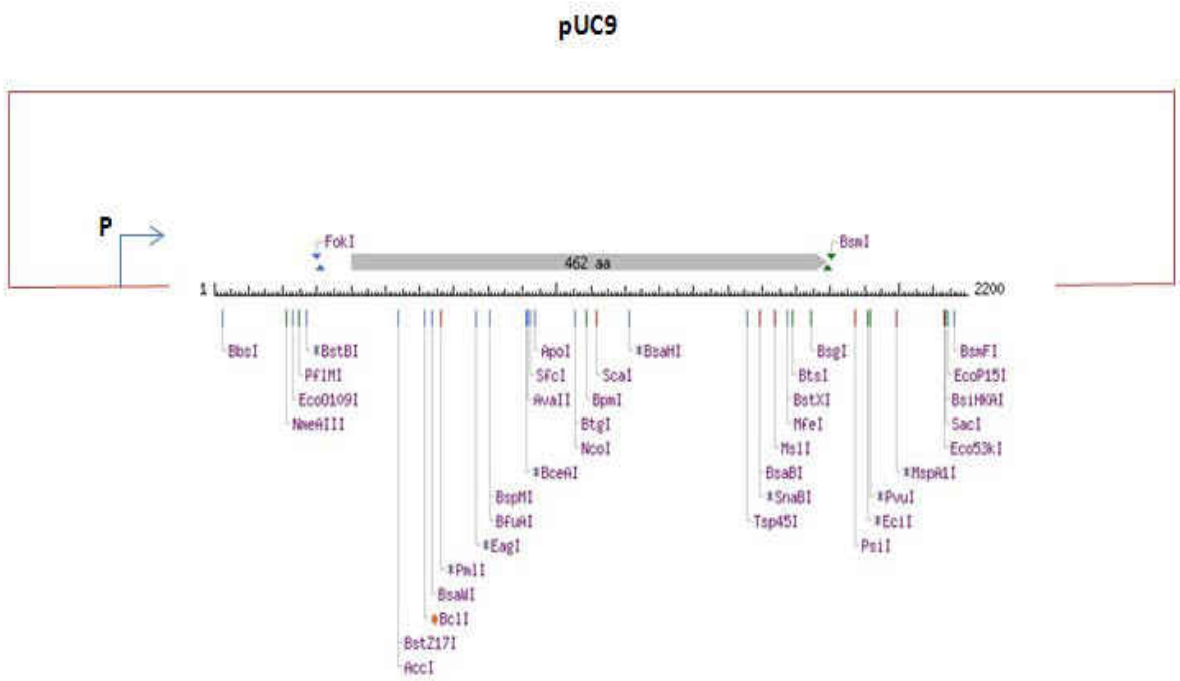


Figure 20: Map of the plasmid clone containging the RT gene from *Microscilla marina.* Amplified DNA from Fig.13B, lane 1 was cloned into the PstI-EcoRI restriction sites of plasmid pUC9.

Microscilla cells appear to produce a large msDNA in very abundant amounts. This makes it easy to purify a large amount of this msDNA from electrophoresis gels. I used this purified msDNA to directly determine its DNA sequence. This was done by adding dGs to the 3' end of the msDNA molecule in a "tailing" reaction with the enzyme terminal transferase (see methods) and using oligo $(dC)_{18}$ as a sequencing primer. About 100bp of fairly good sequence was determined with this method and is shown in Fig.21 (see"Query 1" sequence). This msDNA sequence was then used as a query sequence to search for a matching DNA sequence from the *Microscilla marina* whole genome (shotgun) sequence in the GenBank database. A very closely matching DNA sequence was found at positions 57814 to 57907. However, surprisingly, this matching DNA sequence was not located anywhere near the *Microscilla* RT protein ZP01692041 (positions 11062-12450) on this bacterium's chromosome. To confirm this finding, purified

satellite DNA (msDNA?) from *Microscilla* was labeled with DIG using the random primed method (Fig.22, lane 2). The DIG labeled satellite DNA was then used as a hybridization

Figure 21: BLAST of the sequence of msDNA from *Microscilla marina*. The "query" sequence is a 94 nucleotide sequence determined directly from the satellite DNA from Fig.18, lane 2. This DNA matches a sequences ("Subject") from the chromosome of *Microscilla* at positions 57814-57907.

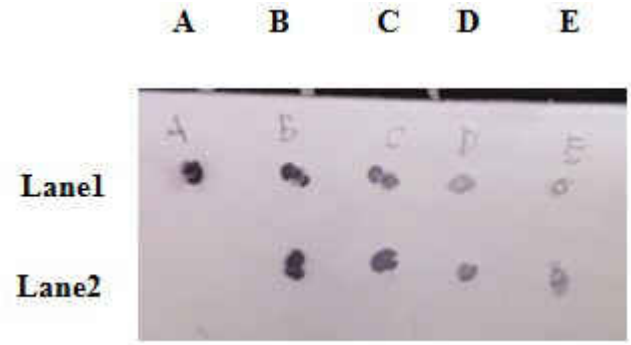


Fig.22 Random primed labeling of msDNA from *Microscilla marina* with Digoxygenin-II dUTP (DIG). Lane 1 A-E: Control DNA with different dilution factiors 10,20,25,50. Lane 2 B-E: Dig labeledmsDNA from *Microscilla marina* with different dilution factors 5,10,20,25,50.

probe against a Southern blot containing both chromosome DNA from *Microscilla* and the pUC9 plasmid clone of the RT protein (Fig.23 A and B). As expected, the probe hybridized to itself (Fig.22B, lane 2). But surprisingly, the msDNA probe did not hybridize to the plasmid DNA

clone of the RT protein from the *Microscilla* genome (Fig. 23, lane 3). The gene *msd*, which codes for msDNA, is normally found a short distance 5' to the start of the ORF encoding the retron RT protein and well within the 400bp of cloned DNA upstream of the *Microscilla* protein (gene) ZP01692041 (Fig.20).

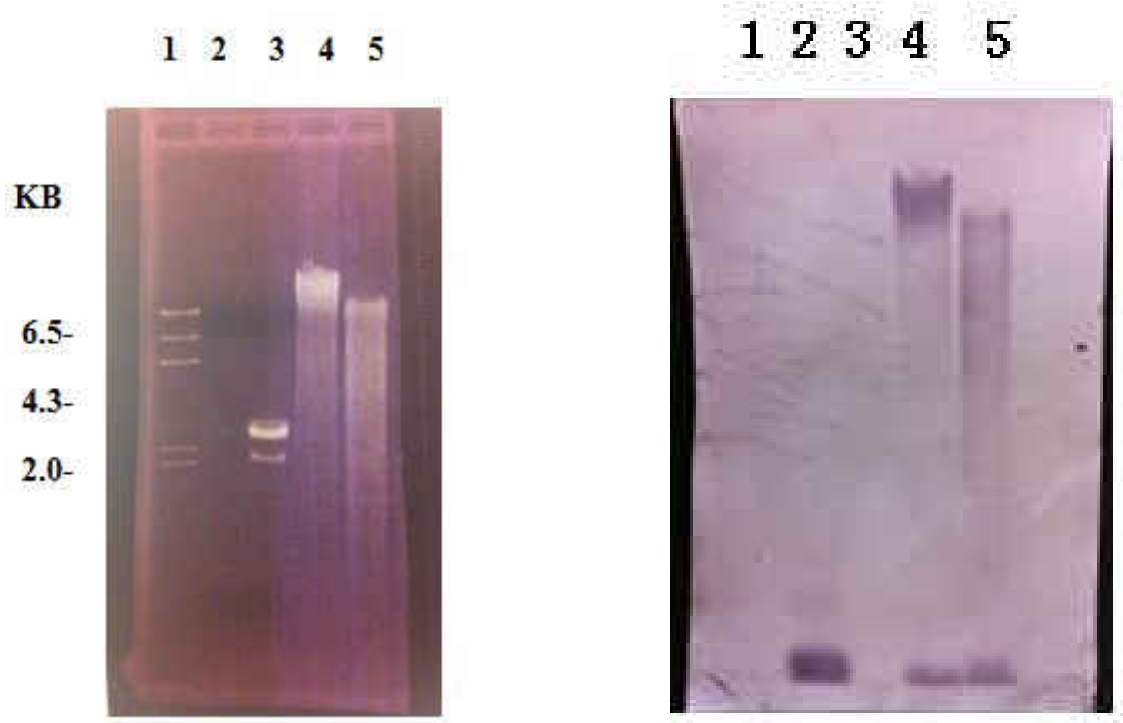


Fig.23: Southern hybridization with DIG labeled msDNA as probe. (A) Gel electrophoresis of DNA samples. (B) Southern blot of the gel and hybridization with the DIG labeled msDNA from *Microscilla*. Lane 1 contains lambda DNA digested with HindIII. Lane 2 contains purified satellite DNA from *Microscilla* cells. Lane 3 contains the plasmid clone (Fig.20) of the RT gene from *Microscilla*, digested with PstI and EcoRI. Lane 4 contains undigested *Microscilla* chromosome DNA. Lane 5 contains *Microscilla* chromosome digested with HindIII.

CHAPTER 4

DISCUSSION

Seven different bacteria, each with a genome sequence annotated to contain a retron type RT, were analyzed to detect the presence of msDNA. These bacteria come from very diverse phylogenetic groups. Three different methods were used to detect msDNA. Of the 7 bacteria, 5 of them (*Gemmata*, *Methylobacterium*, *Microscilla*, *Ralstonia*, and *Shewanella*) clearly showed the presence of msDNA. Although for 2 of these bacteria, *Ralstonia* and *Shewanella*, only the most sensitive radioactive (RT extension) method could detect the presence of a very small msDNA (Fig.17B). The other 3 bacteria produced large msDNAs of a 100 deoxyribonucleotides or longer.

msDNA was not detected in 2 of the 7 bacteria analyzed: *Rhodococcus* and *Pseudomonas*. For *Rhodococcus*, there is an easy explanation for this. The putative *Rhodococcus* retron RT protein is clearly truncated at the N-terminus. The amino acid sequence of protein YP_700792 from *R. jostii* (strain RHA1) shows that the protein is missing all of conserved domain I and most of domain II (Fig.7). It is, therefore, probably no longer a functional RT protein.

Several additional *Rhodococcus* species also appear to contain a retron type RT gene in their genome sequence (from the original BLAST search using the mx162 RT as a query sequence). The protein YP_002769309 from *R. erythropolis* PR4, the protein ZP_04388531 from *R. etrythropolis* SK121, the protein EHI 45352 from *R. opacus* PD630, and the protein YP_002777738 from *R. opacus* B4 were compared with the retron RT from *R. jostii* RHA1 and the known Mx162 retron RT from *M. xanthus* in a multiple amino acid sequence alignment (Fig. 24). This alignment clearly shows the truncated RT protein from *R. jostii* RHA1. However,

the 4 proteins from other *Rhodococcus* species, which are similar, do not appear to be truncated and are probably functional RTs and likely produce msDNA.

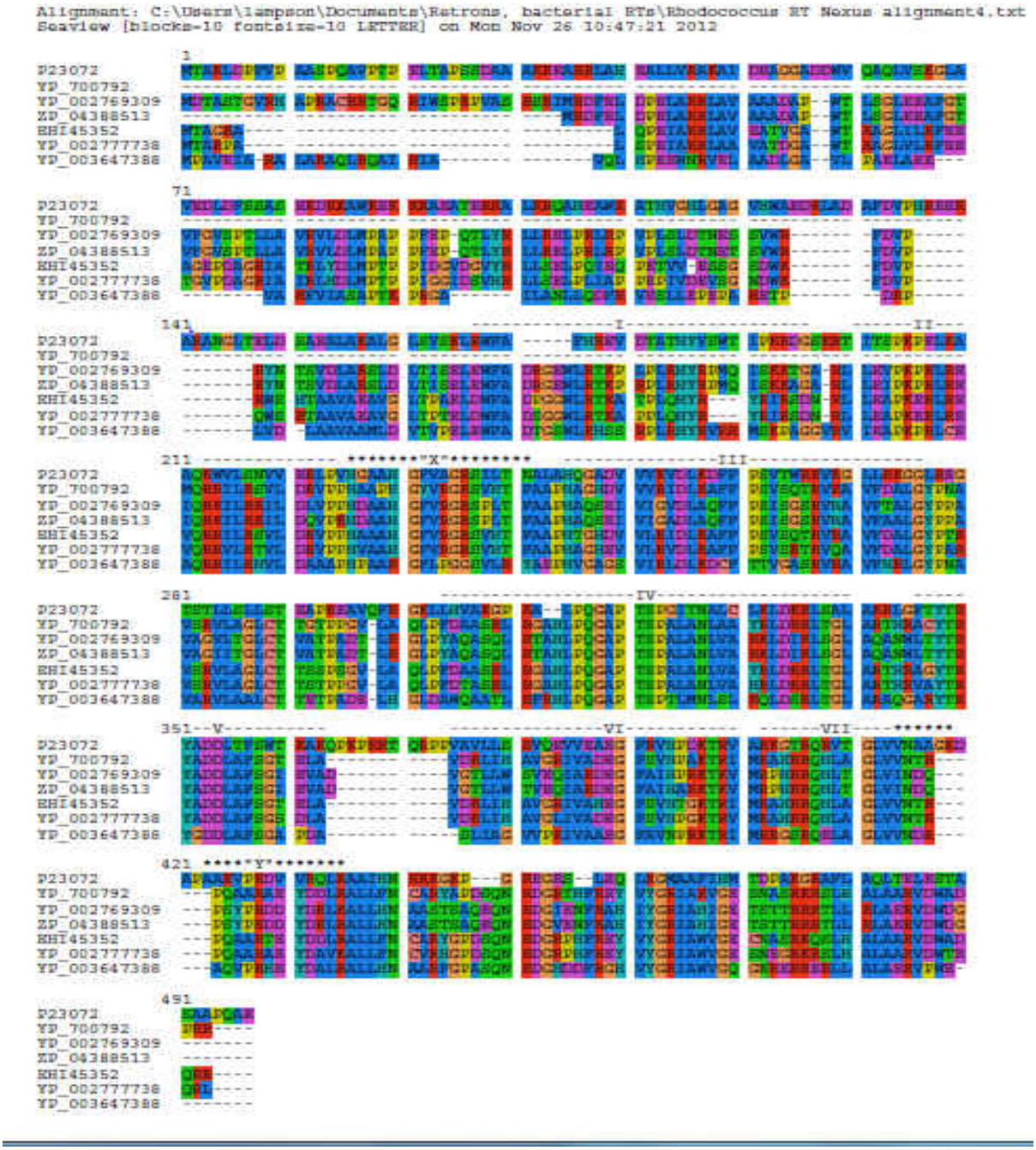


Fig.24: Amino acid alignment of RT proteins from different *Rhodococcus* species. P23072 is the protein from Mx162. YP_700792 is the protein from *Rhodococcus jostii* RHA1. YP_002769309 is the protein from *R. erythropolis* PR4. ZP_04388513 is the protein from *R. erythropolis* SK121. EHI45352 is the protein from *R. opacus* PD630 YP_002777738 is the protein from *R. Opacus* B4. YP_003647388 is the protein from *Tsukamurella paurometabola* DSM 20162.

The other bacterial species that did not produce msDNA was *Pseudomonas aeruginosa* PAO1, whose genome was annotated to contain the retron type RT protein YP_249406. The amino acid sequence of this protein is quite variable compared to the query protein, Mx162 RT (Fig.7). It is possible that this protein is not a retron RT, but perhaps a RT that has some other unknown function.

Finally, for *Microscilla marina*, the result I got is quite interesting. One reason is that when the total DNA was extracted from *Microscilla* and electrophoresed on a 4% acrylamide gel, a very clear and small satellite DNA appeared on the acrylamide gel (Lane 2, Fig.19). The other reason is that when the retron RT gene from *Microscilla* cloned into pUC9 and over-expressed in *E.coli*, the retron RT gene was expressed and a clear small satellite DNA was shown on the gel (lane 4, Fig.9). But the most unusual thing I found occurred when I directly sequenced the satellite DNA purified from on an acrylamide gel. I found that the sequence of this satellite band does appear in the microscilla genome database, but the sequence location is not near the retron RT gene (Chromosome 57814-57907, Fig.21). Based on the retron RT characteristics, normally the msDNA sequence should be a little bit upstream of the RT gene, but the msDNA sequence from *Microscilla marina* is not and it is still a mystery. Based on the alignment results, the *Microscilla marina* retron RT (ZP_01692041), it also has a long N-terminal extension prior to the first conserved domain like myxobacteria. Therefore, it is predicted that it should produce msDNA.

REFERENCES

1. Temin, H.M., Mizutani, S. (1970) Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature* 226: 1211–1213.

2. Baltimore, D. (1970) Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature* 226: 1209–1211.

3. Kramerov, D.A., Vassetzky, N.S. (2005) Short retroposons in eukaryotic genomes. *International Review of Cytology* 247: 165–221.

4. Harberer, G., Young, S., Bharti, A.K., Gundlach, H., Raymond, C., Fuks, G., Butler, E., Wing, R.A., Rounsley, S., Birren, B. (2005) Structure and architecture of the maize genome. *Plant Physiol*ogy 139: 1612-1624.

5. Hua-Van, A., Le Rouzic, A., Maisonhaute, C., Capy, P., (2005) Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. *Cytogeneics and Genome Research* 110: 9590-9594.

6. Greider, C.W., and Blackburn, E.H., (1989) A telemetric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Natur*e 337: 331-337.

7. Lampson, B.C (2007) Prokaryotic Reverse Transcriptase. *Industrial Enzymes and structure, function and application* 23: 401-418.

8. Michel, F., and Lang, B.F., (1985) Mitochondrial class II introns encode proteins related to the Reverse Transcriptase of retroviruses. *Nature* 316: 641-643.

9. Costa, M., Francois, M., Toro, N. (2006) Potential for alternative intron–exon pairings in group II intron RmInt1 from *Sinorhizobium meliloti* and its relatives. *RNA* 12(3): 338-341.

10. Medhekar, B., Miller, J.F. (2007) Diversity-generating Retroelements. *Current Option in Microbiology* 10: 388-395.

11. Doulatov, S., Hodes, A., Dai, L., Mandhana, N., Liu, M., Deora, R., Simons, R.W., Zimmerly, S., Miller, J.F. (2004) Tropism switching in *Bordetella* bacteriophage defines a family of diversity generating Retroelements. *Nature* 431: 476-481.

12. Cousineau, B., Lawrence, S., Smith, D., Belfort, M. (2000) Retrosposition of a bacterial group II intron. *Nature* 404: 1018-1021.

13. Lampson, B.C., Inouye, M., Inouye, S. (2001) The msDNAs of bacteria. *Progress in Nucleic Acid Research and Molecular Biology* 67: 65-89.

14. Shimamoto, T., Shimada, M., Inouye, M., Inouye, S. (1995) The role of ribonuclease H in multicopy single-stranded DNA synthesis in retron-Ec73 and retron-Ec107 of *Escherichia coli*. *Journal of Bacteriology* 177: 264-267.

15. Lampson, B.C., Inouye, M., Inouye,S. (1989) Survey of multicopy single-stranded DNAs and Reverse Transcriptase genes. *Journal of Bacteriology, Cell* 56: 701-707.

16. Xiong, Y., Eickbash, T.H. (1990) Origin and evolution of Retroelements based upon their Reverse Transcriptase sequences. *EMBO* Journal. 9: 3353-3362.

17. Dhundale, A., Inouye, M., Inouye, S. (1988) A new species of multicopy single-stranded DNA from *Myxococcus xanthus* with conserved structured features. *Journal of Biological Chemistry* 163: 9055-9058.

18. Rice, S.A., Lampson B.C., (1995) Phylogenic comparison of retron elements among the myxobacteria: Evidence for vertical inheritance. *Journal of Bacteriology* 177: 37-45.

19. Teiichi, F., Inouye, S., Inouye, M. (1987). Biosynthesis and structure of stable branched RNA covalently linked to the 5' end of multicopy single-stranded DNA of *Stigmatella Aurautiaca*. *Cell* 48: 55-62.

20. Lampson, B.C., Xu, C., Rice, S.A., Inouye, S. (2002) A partical copy of msDNA from a new retron element is likely a retrotransposed DNA found in the myxobacterium *Nannocystis exedens. Gene* 299: 255-261.

21. Fuerst, J.A., Webb R.I (1991) Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. *PNAS* 88: 8184-8188.

VITA

SEN MU

Personal Data:	Date of Birth: March 5th , 1986 Place of Birth: Jinan, Shandong, China Marital Status: Married
Education:	B.S, Microbiology, East Tennessee State University, Johnson City, TN 2010 B.S, Computer Science, Shandong Normal University, Jinan, Shandong, China 2012 M.S, Biology, East Tennessee State University, Johnson City, TN 2013
Professional Experience:	Undergraduate Research, Department of Chemistry, East Tennessee State University, 2009-2010 Graduate Research, Department of Health Sciences, East Tennessee State University, 2010-2012 Research Technician, Department of Health Sciences, East Tennessee State University, 2012-2013
Preparation:	Mu, S. 2013. Detection of Bacterial Retroelements Using Genomics. M. S. Thesis. East Tennessee State University, Johson city, TN
Honors and Awards:	3rd position Scholarship recipient for academic and research achievement in the Department of Computer Science, Shandong Normal University, 2005

Dean's List in College of Public Health in Fall 2009, East Tennessee State University, 2009

Excellent Graduate Honors—Cum Laude, East Tennessee State University, 2010