Improving Storage Efficiency of Cipher Text Using Palindrome Compression Technique

[1]Dr.R.Kiran Kumar, [2]P.Bharathi Devi

[1]Associate Professor, [2]Research Scholar
[1]Department of Computer Science
[1]Krishna University, Machilipatnam, Andhra Pradesh, India
[1]kirankreddi@gmail.com, [2]bharathipatnala@gmail.com

Abstract: In present days data exchange rate is high through electronic media. Huge numbers of them utilized Whatsapp, Twitter, Instagrams, messages and other data media to share their data. The extent of data is enormous. Offering security to the data is likewise significant angle. With a specific end goal to give security numerous Cryptography approaches accessible. Out of them DNA Cryptography is a developing field. In DNA Cryptography the figure content is as DNA arrangements. There are tremendous genomic successions in the real world. Putting away of these genomes in Gene Bank is likewise troublesome.This adds up to tremendous information stockpiling costs, in this manner making the utilization of this information, for example, examination and recovery very difficult. DNA arrangement examination is valuable in differing territories, for example, crime scene investigation, medicinal research, drug store, agribusiness and so on, It is extremely important to address the capacity issue of these exponentially developing information. Present day natural science produces tremendous measures of genomic arrangement information. This is fuelling the requirement for effective calculations for grouping pressure and investigation. Information pressure and the related strategies originating from data hypothesis are regularly seen as being of enthusiasm for information correspondence and capacity. In the present paper the authors proposed the DNA Compression by splitting based on the given length and find whether the subsequence contains palindrome or not. It it contains palindrome half of the string only will be stored in the compressed string.

IndexTerms - DNA, DNA Compression, Genebank, DNA Cryptography

I. Introduction

As all we know, the cryptography plays a vital role to provide security in the field of network or any storage media. There are various cryptographic techniques available now a days. Out of which, DNA cryptography is new born field in the field of Cryptography. The data which will be hide or converted the transmitted the data in the form DNA. While encrypting the cipher text is in the form of DNA which is of very long length sequences. To provide efficient storage for the ciphertext we need to provide compression technique after performing any cryptography algorithm using DNA. There are various DNA compression techniques available to compress the DNA sequences of spices in Genebank. In the present paper, the compression technique which will be used for providing efficient sThese days, we are living in the period of web of things due to the utilization of billions of gadgets associated together and the critical increment in the dataflow put away and transmitted between them. Along these lines, each day, a gigantic amount of computerized data is utilized, shared and broke down. Great PCs are utilized to appropriately investigate and store this information. Thusly, two issues have emerged. In the first place, the encoding of the information and second, the time required to process them. Therefore, to diminish information sizes, numerous information pressure strategies have been actualized.. Mechanical advancement has prompted the introduction of bioinformatics investigate region which forms and breaks down various living creatures' information. The fundamental component in accomplishing these medicines is the Deoxyribonucleic Acid or DNA, which is a biomolecule present in all cells. This biomolecule contains the hereditary data required for the working and advancement of every single living being. Every monomer establishing it is a nucleotide, which is made out of a nitrogenous base; Adenine (A), Cytosine (C), Guanine (G) or Thymine (T). GenBank, overseen by the International Nucleotide Sequence Database Collaboration, is a free access database that contains a lot of DNA groupings which are put away in crude design and that may thusly prompt excess information.

II.Related Work

The following table illustrated various works done in the field of DNA compression.

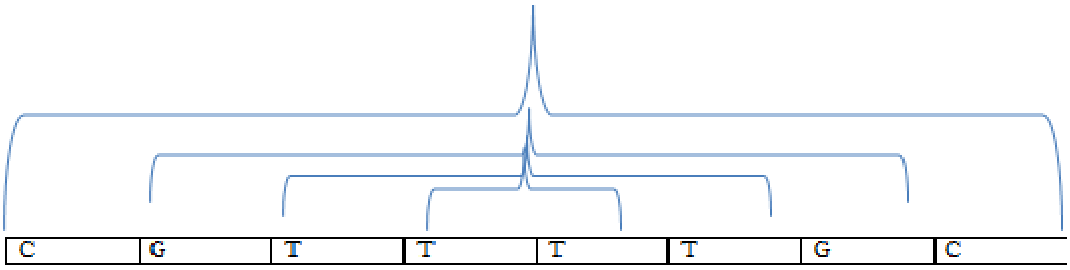
Table1: Literature survey on various compression techniques.

Year & Publication	Title of the Paper & Author	Name of the Compressor	Methodology Used	Compression Ratio
1993 & Proceedings of Data Compression Conference[1]	Compression of DNA Sequences &Grumbach	Biocompress	Sliding Window Algorithm	1.850
1994 & IJIPM[2]	A new challenge for Compression algorithms: gentic sequences &Grumbach	Biocompress 2	Arithmetic Coding	1.783
1997 & GIW97[3]	Fast Discerning Repeats in DNA Sequences with Compression & Rivals	Cfact	Two Pass Algorithm	2
1998 & ISMB[4]	Compression of strings with approximate repeats & Allison	ARM	Summing the probabilities over all explanatios of how the subsequence is generated	1.714
1999 & GIW99[5]	A Compression Algorithm for DNA Sequences & Its Applications in Genome Comparison& Chen	GenCompress	Hamming Distance	1.742
2002 & IJB[6]	DNACompress: fast & effective DNA sequence compression & Chen	DNACompress	Same as GenCompress but it works on larger sequences	1.7255
2007 & Proceedings of the Data Compression Conference[7]	A simple statistical algorithm for biological sequence compression &M.D.Cao	XM	Symbol by Symbol probability estimation	1.714
2010 & IJCSIT[8]	Genbit Compress Tool &Rajeswari	Genbit	Segmenting bits of four bases	1.727
2010 & JTAIT[9]	Huffbit Compress &Rajeswari	Huffbit	Extended binary tree principle	1.611
2015 & Proceedings of the World Congress on Engineering and Computer Science[10]	DNA Sequences Compression Algorithm Based on Extended-ASCII Representation	Extended ASCII DNACompress	Extended ASCII Encoding	1.7
2016 & Proceedings of World Congress on Engineering[11]	Modified DNA bit &BacemSaada	ModifiedDNABit	Extended ASCII Encoding	Less than 1.6
2016 & IRJET[12]	Reverse Sequencing based Genome sequence using Lossless Compression Algorithm & Rajesh Mukherjee	DNA Sequence Compressor	Reverse substring	It fails to achieve higher compression ratio than other methods, but it has provided very high information security.

III. PROPOSED WORK

In the present paper the author proposed to compress the DNA sequence based on palindrome technique. Palindrome means the original sequence is equal to the reverse of the sequence.For eg., when you consider a sequence S=CGTTTTGC. Here the reverse the sequence is also R=CGTTTTGC same as original. So we can store half of the string can be stored in the compressed string.

Figure 1: Palindrome Technique.



When you divide the half of the palindrome string, the reverse half string is the remaining string. The same concept will apply here.

Compression Algorithm

Algorithm PalinCompress(C,length)
C is the Ciphertext in the form of DNA Sequences and length is used to divide the DNA sequence into number of specified length strings.

```
Begin
    1. Split the DNA strand into number of substrings based on the given length
    2. Check whether each substring is palindrome or not. If the substring is palindrome then store half of the string by padding number of leftover characters to the substring.
    3. If it is not palindrome, continue with the same string.
    4. Repeat step 2 & 3 for all the substrings.
End
```

Decompression Algorithm

Algorithm PalinDeCompress(D)
D is the Decompressed DNA strand.

```
Begin
    1. Read the DNA strand until you find the digit.
    2. When the digit is found, count the number of characters equal to the digit, that are immediately presented the digit and add the same characters in reverse order from the location where the digit is found that is the digit is replaced by the first character.
End
```

For example take a DNA Strand

Best Case:
S=AGGTTGGAACGTTGCAAAAAAAAATTGGGGTT and Length=8
Divide the DNA strand into 8 length substrings then
S1=AGGTTGGA
S2=ACGTTGCA
S3=AAAAAAAA
S4=TTGGGGTT
Find the each substring is palindrome or not, if yes then store half of the string and the length of that half string. Here, all four strings are palindromes then store half of the strings and the lengths then the compressed string is
S=AGGT4ACGT4AAAA4TTGG4
Compression Ratio=Number of Sequences After Compression/Number of Sequences Before Compression
Compression Ratio=20/32=0.625

Average Case

Consider the Sequence
S=ACGTTGCAAAAACCCAGTTTTTTGATTAAAAA and take the length as 8
S1=ACGTTGCA
S2=AAAACCCA
S3=GTTTTTTG
S4= ATTAAAAA
The First and third substrings are palindromes and the remaining strings are not palindromes then the compressed string is
S=ACGT4AAAACCCAGTTT4ATTAAAAA
Compression Ratio=Number of Sequences After Compression/Number of Sequences Before Compression
Compression Ratio=26/32=0.625=0.8125

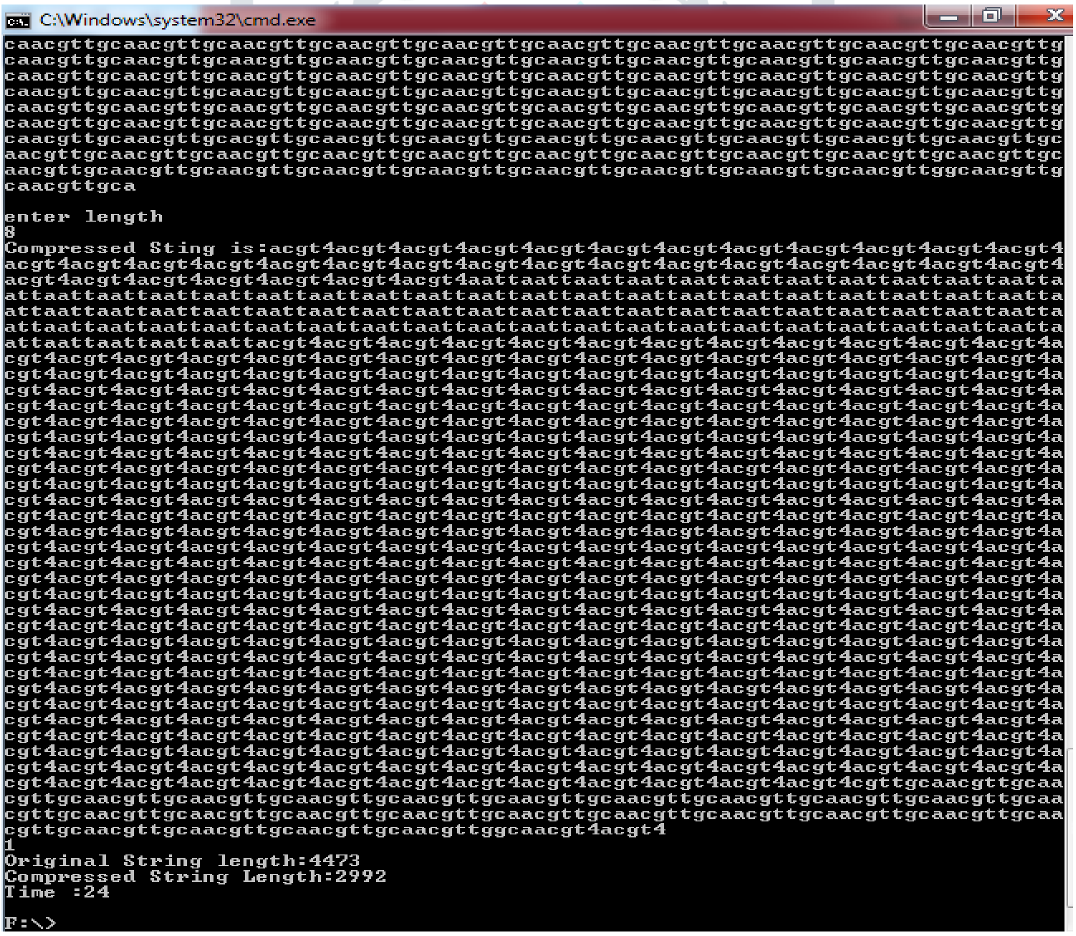
Worst Case

Consider the Sequence
S=ACGTACGTAAAAAGTAAATTATATTCCTTCCC and take the length as 8
S1=ACGTACGT
S2=AAAAAGTA
S3=AATTATAT
S4= TCCTTCCC
Here no substring is palindrome so compression ratio is 32/32=1

IV. Performance Analysis

The following screenshots are compression and decompression of DNA sequences using palindrome technique and the storage space is reduced more than 50%.

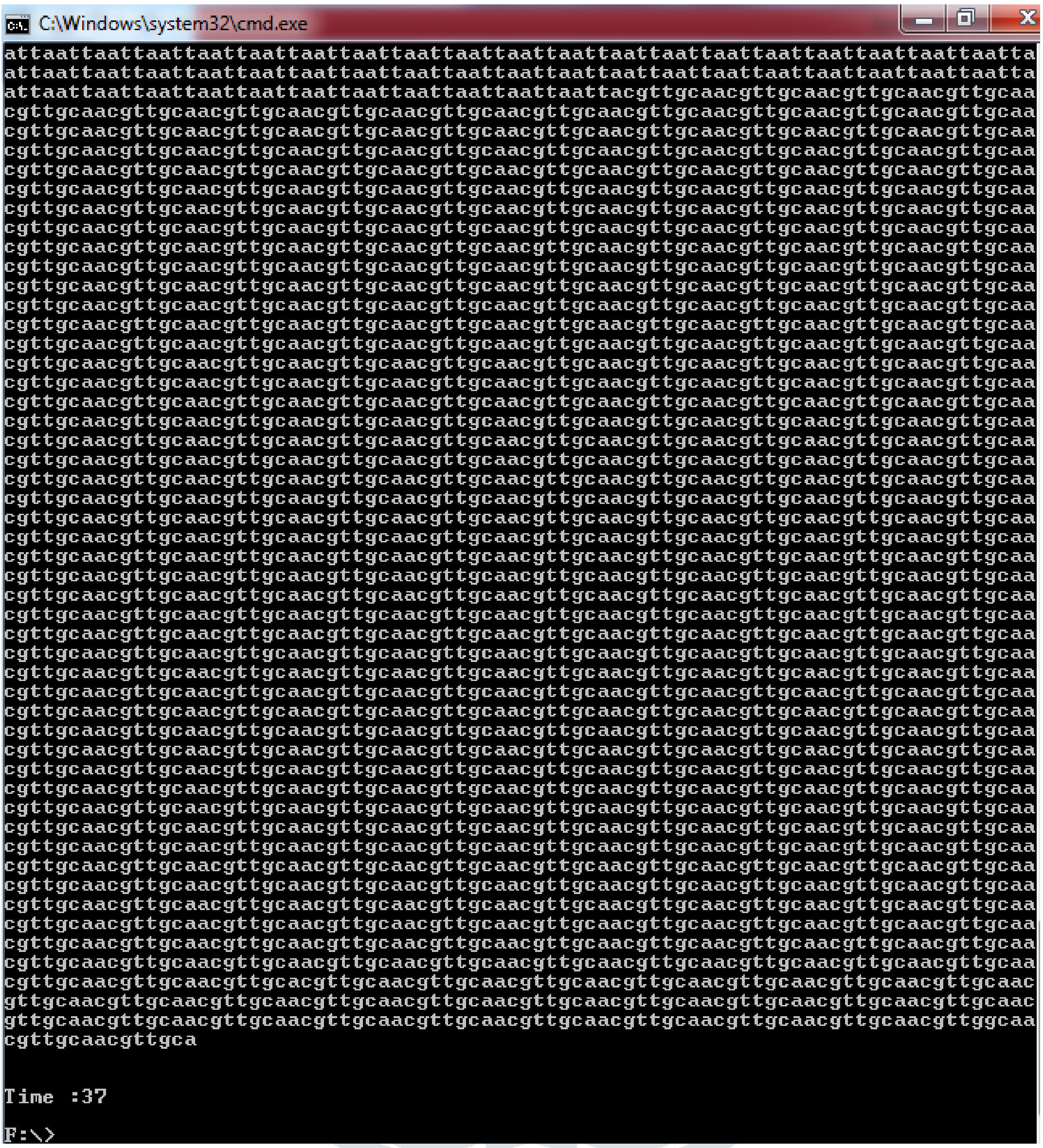
Figure 2: Sequences after Compression.



C:\Windows\system32\cmd.exe

```
caacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttg
caacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttg
caacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttg
caacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttg
caacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttg
caacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttg
caacgttgcaacgttgcacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgc
aacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgc
aacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttggcaacgttg
caacgttgca

enter length
8
Compressed Sting is:acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4
acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4
acgt4acgt4acgt4acgt4acgt4acgt4acgt4aattaattaattaattaattaattaattaattaattaattaatta
attaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaatta
attaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaatta
attaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaatta
attaattaattaattaattacgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4a
cgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4acgt4cgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttggcaacgt4acgt4
1
Original String length:4473
Compressed String Length:2992
Time :24

F:\>
```

Figure 3: Sequences after Decompression.



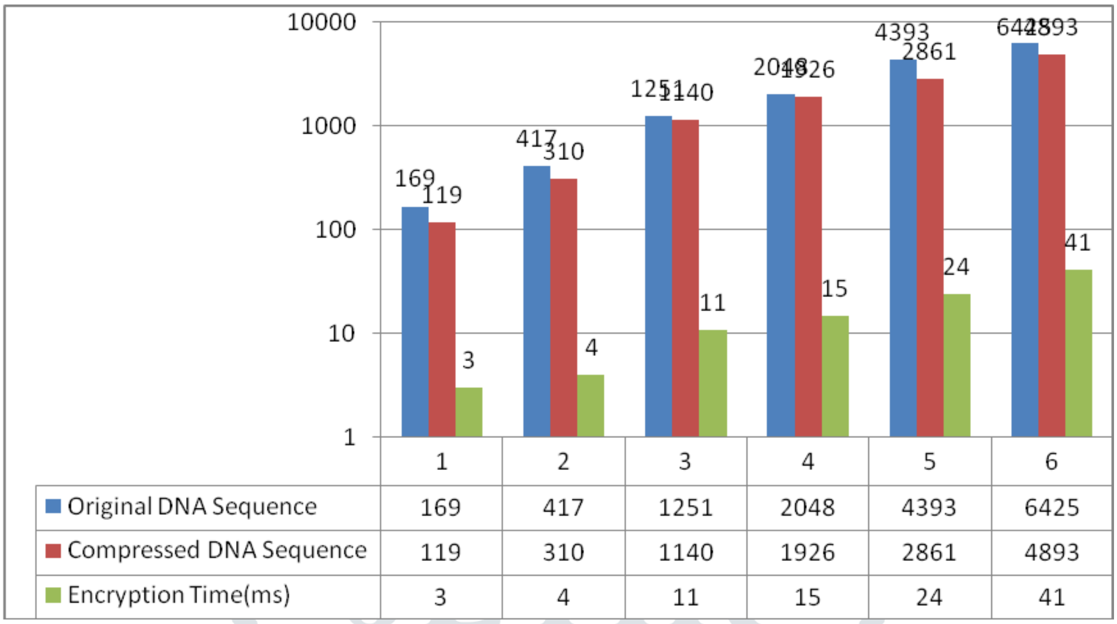
```
attaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaatta
attaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaattaatta
attaattaattaattaattaattaattaattaattaattaattaattacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaa
cgttgcaacgttgcaacgttgcacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaac
gttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaac
gttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttgcaacgttggcaa
cgttgcaacgttgca


Time :37

F:\>
```

The analysis of an algorithm measured in intel i3 processor in jdk1.8.0-131 environment with 4GB RAM in Window7 Platform for various text documents. The following table showed that the compression and Decompression time taken to compress various DNA Sequences which were stored in various text documents.

Table7.1: Performance Measurement for Compression and Decompression of various text documents.

Text Documents in terms KB	Original DNA Sequence	Compressed DNA Sequence	Compression Ratio	Compression Time(ms)	Decompression Time(ms)
1	169	119	0.7041	3	3
2	417	310	0.7434	4	4
3	1251	1140	0.9112	11	10
4	2048	1926	0.9404	15	17
5	4393	2861	0.6512	23	22
6	6425	4893	0.7615	41	38

Figure 4: Time taken to compress the sequences and the Length of ciphertext after compression.



	1	2	3	4	5	6
🟦 Original DNA Sequence	169	417	1251	2048	4393	6425
🟥 Compressed DNA Sequence	119	310	1140	1926	2861	4893
🟩 Encryption Time(ms)	3	4	11	15	24	41

V. CONCLUSION

Several methods are available to compress the DNA sequences which are stored in GeneBank. The compression algorithm which is implemented here used the palindrome technique which is meant to reduce the storage space as well as bandwidth while transmitting the data. Thus it gave rise to a lossless compression reducing more than 50% of space. Further it is easy to implement with less computational complexity.

VI. REFERENCES

[1] S. Grumbach and F. Tahi.1993, "Compression of DNA Sequences," in Proc. of the Data Compression Conf., (DCC '93), 340–350

[2] Grumbach, S., and Tahi, F.1994, A new challenge for compression algorithms: genetic sequences, Information Processing & Management, 30(6), 875–886.

[3] E. Rivals, M. Dauchet, J-P. Delahaye, et al.1997, "Fast Discerning Repeats in DNA Sequences with a Compression Algorithm," The 8th Workshop on Genome and Informatics, (GIW97), 8, 215-216.

[4] Allison, L., Edgoose, T., and Dix, T. I.1998, Compression of strings with approximate repeats, Proc. ISMB, 8–16.

[5] X. Chen, S. Kwong and M. Li.1999, "A Compression Algorithm for DNA Sequences and It's Applications in Genome Comparison," The 10th Workshop on Genome and Informatics, (GIW99), 10, 51-61.

[6] Chen, X., Li, M., Ma, B., et al.2002, DNACompress: fast and effective DNA sequence Compression, Bioinformatics,18(12), 1696–1698.

[7] M. D. Cao, T. I. Dix, L. Allison, et al.2007, "A Simple Statistical Algorithm for Biological Sequence Compression," in Proc. of the Data Compression Conf., (DCC '07), 43–52.

[8] Rajeswari, P. R., and Apparao, A.2010, Genbit Compress Tool (GBC): A Java-Based Tool To Compress DNA Sequences and Compute Compression Ratio (BITS/BASE) Of Genomes, International Journal of Computer Science and Information Technology, 2(3), 181-191.

[9] Rajeswari, P. R.,Apparao, A., and Kumar, R. K.2010, "HUFFBIT COMPRESS – Algorithm to compress DNA sequences using extended binary tree", Journal of Theoretical and Applied Information Technology, 13(2), 101-106.

[10] BacemSaada, Jing Zhang.2015, "DNA Sequences Compression Algorithm Based on Extended-ASCII Representation", Proceedings of the world congress on engineering.,2., 978-982

[11] Bacem Saada.2016, "DNA Sequences Compression techniques Based on Modified DNABIT Algorithm", Proceedings of the world congress on engineering, 978-982

[12] Rajesh Mukherjee, SubhrajyotiMandal and Bijoy Mandal.2016, "Reverse Sequencing based Genome Sequence using Lossless Compression Algorithm", International Research Journal of Engineering and Technology, 1208-1215.