Chapter 7

CONCLUSION

7.1 Conclusions

The purpose of our undertaken work of making observations is to get greater, stronger, more complete algorithms, to help persons making observations with managing the amount of greatly sized DNA produced by DNA ordering. In this thesis, we have studied DNA order, place for storing, safety and difference with artificial order. There are two moves near to trading with the growth in orders given into the public knowledge bases: 1) DNA orders place for storing, 2) DNA orders safety. These two are not in a common two-way uncommon i,e, different. A number of arguments for not storing all order facts have been put forward. We gave all attention on an algorithm for making small sized collections of assembled DNA orders. We introduced fifteen compression and two encryption algorithms.

For text compression the apparatus used by the user for making observations, namely gzip and bzip2 produced some of the worst being compression outcomes. The general approved dictionary base, LZ77, binary code, Huffman's and substitution & statistical based compression have some drawbacks. Therefore, pattern matching algorithms based on REPEAT, REVERSE & PALINDROME, GENETIC PALINDROME and their combination are proposed. If use safety directly on the formed of small units of DNA order, getting very low safety because the DNA order have within only four bases, anyone can easily read the facts by trial an error methods. Traditional symmetric keys like Data Encryption Standard (DES), advanced encryption standard (AES) and escrowed encryption standard are not good at producing an effect when the facts size is large, Huffman's code also fails badly on DNA orders both in the static & adaptive, because the chances of taking place of the four special signs are not very different, compression algorithm is offered and encryption had modified Huffman's & RSA algorithm. We have got a new, greater, stronger, more complete pattern matching, altered Huffman's & RSA algorithm which works equally well with the best existing pattern algorithm. All algorithms have been given effect to and more experimental results have been done to test useful performance of the algorithms and also to compare with other standard algorithm. The experimental results have made clear that our pattern matching, and modified Huffman's & RSA algorithm, is almost the fastest among all algorithms, when the number of designs are not very greatly sized, as an outcome of that, our short pattern matching algorithm is desired for general pattern matching application and modified Huffman's &RSA algorithm is desired for the process of encryption

These algorithms can be used by persons making observations to store and provide safety to greatly sized collections of DNA orders while, also letting them to use the facts while it is compressed. For greatly sized order groups the overhead of storing uncompressed facts, or storing compressed facts, encrypt facts and then decrypt decompressing to obtain facts, is high. These issues can be overcomed with the use of fifteen plus two algorithms.

We introduced compression algorithms, pattern & substitution repeat, reverse, palindrome & genetic palindrome & their combination. At each iteration, of the input to make certain that all done again and again making free from doubt certain conditions that are sensed in the complete getting together before substitution decisions are made. The algorithm also supports the exchange of highly frequent repetition taking place with a minimum number of times. This pattern matching and substitution technique was unable to produce better result than the marketable standard result, introduced RHUFF technique for compression as well as encryption. The sequence characteristics of the group are also kept safe and each sequence extraction without decompressing the complete group is possible. We introduced modified Huffman's and RSA technique for the process of encryption. It is right for encryption making certain safety of the DNA order. The encryption depends on level of encryption and speed, both these parameters are successfully fulfilled by our algorithm. This modified Huffman's algorithm brought under control the difficulty in using the DNA order appropriately and produced good result.

Many altercations in favour of not storing all sequence raw data have been suggested. These include" don't store raw data , just store compressed DNA sequence", "discard older DNA sequence and " store only analysis encrypt-compressed output". " Life represents order, it is not without order or random".

The cost of storing data has become a significant part of the total price in the formation and observation of DNA order. The rate of increase of DNA order is exceeding the rate of storing capacity but the storage capacity is limited. In this thesis, we present a pattern substitution and encryption-compression algorithm with small amount of support compresses the DNA orders for place of storing & make certain the safety. The output of the algorithm shows that it would be useful for searching repeated region within orders and points in the likeness between different order to make shorter only the dissimilarities. Every dissimilarity is stored as its library text record and its position on the read.

The compression results obtained from the algorithm reveal that the certain repeat, reverse, palindrome & genetic palindromes and their combination are one of the most impotant concealed regularities in DNA sequence. The library file size is too small with respect to the compressed file.

7.2 Application

These techniques are used to shorten genomic facts and on other uses as user application, text has to be make shorter because of the limitation of the network bandwidth before transmission, it also needs to be encrypted to support secrecy.

7.3 Future Work

i) The time complexity will be reduced and improve encryption level in future.

ii) In future the compression rate & ratio is improved

iii) The standard modified encryption algorithm is introduced for data security in near future.