Residue Number System: An Important Application in Bioinformatics

Hassan Kehinde Bello
Federal Polytechnic, Offa
Offa, Kwara State Nigeria

Kazeem Alagbe Gbolagade
Kwara State University, Malete
Ilorin Kwara State Nigeria

ABSTRACT
This paper presents an overview and comprehensive survey of Residue Number System (RNS) and Bioinformatics. It focuses on application of RNS to Smith-Waterman Algorithm (SWA), highlights how the inherent attributes of RNS can be applied to improve performance of SWA in the analysis of Deoxyribonucleic Acid (DNA) and also observes the two principal methods of data conversion and lastly, we suggest the direction for future research.

General Terms
Algorithms, Number system, operands, Conversions

Keywords
Residue Number System, Smith-Waterman Algorithm, Deoxyribonucleic Acid, Bioinformatics.

1. INTRODUCTION
Residue Number System (RNS) is the representations of a large integer number with a set of smaller integer numbers in order to make computation fast and efficient. RNS may be traced back to 1599 years; it begins with a Chinese Scholar, Sun Tzu [1] with Chinese riddle as follows: What is the number such that when divided by 7, 5 and 3 will have the remainders of 2, 3 and 2 respectively? The procedure of obtaining the solution to the riddle is known as Chinese Remainder Theorem (CRT).

The speed of Residue number system is very high in the operations like addition, subtraction and multiplication. This is because RNS supports carry free addition; borrow free subtraction and digit to digit multiplication without partial product. These interesting properties make RNS a popular application in Digital Signal Processing (DSP), Fast Fourier Transformation (FFT), digital filtering, image processing, cryptography, etc. This is due to inherent properties like fault tolerance [2], [3]. However, RNS has some difficulties in applications involving magnitude comparisons, sign detection, overflow detection, division, reverse conversion etc.

Whenever RNS is to be applied, appropriate choice of moduli set is very important because it determines the speed of the system. The higher or more the dynamic range of the moduli set, the faster the system but the slower its reverse conversion.

The rest of the article is organized as follows: Section 2 gives the background to the paper; section 3 explains RNS in bioinformatics; section 4 discusses sequence alignment in bioinformatics; section 5 presents RNS-SWA architecture and finally, the paper concluded in section 6.

2. BACKGROUND
Residue Number System (RNS) is defined over an interval of relatively prime modulus sets $\{P_1,P_2,P_3 \ldots P_n\}$, where the greatest common divisor (gcd) between any two P_i and $P_j = 1$ i $\neq$ j i.e $\gcd(P_i, P_j)=1$. Let us represent a number X as $\{x_1, x_2, x_3, \ldots, x_n\}$. This representation is unique for any integer X in the range [0, M-1]. Where M is the product of the given moduli sets (i.e $M = P_1 \times P_2 \times P_3 \times \ldots \times P_n$.). RNS offers flexibility in digit by digit computation; this makes addition, subtraction and multiplication more efficient.
r is the RNS of a number X with respect to *m*, if and only if r is the leftover after several removal of all multiples of *m* from X, where m is the moduli.

Example 1: If X is 17 and m is 3 then the successive removal of multiple of 3 from 17 is given by

X: 17 14 11 8 5 2

After the removal of first multiple of 3 from 17, we still have more multiples, this continues till no more multiple of 3 can be found again, implies 2 is the RNS of 17 with respect to 3, since it is the leftover after successive removal of all multiples of 3 from 17

The RNS of X with respect to m is given by . $x_i = |X|_{mi}$

Example 1 above can be represented as $|17|_3 = 2$

2.1 Basic RNS arithmetic
In RNS arithmetic, addition, subtraction and multiplication are simple, depending on choice of moduli, the standard arithmetic operations can be effectively implemented in RNS. Addition / subtraction of residues is carried out by individually adding / subtracting the corresponding digits of the moduli respectively.

Example 1: if X = {2,3,5}, Y = {1,3,4} with respect to moduli 3,5,7 then,

$|X + Y|_{3,5,7} = |(2,3,5) + (1,3,4)|_{3,5,7} = (0,1,2)$

$|X - Y|_{3,5,7} = |X + \overline{Y}|_{3,5,7}$

$= |(2,3,5) + (\overline{1}, \overline{3}, \overline{4})|_{3,5,7} = (1,0,1)$

$|X \times Y|_{3,5,7} = |(2,3,5) \times (1,3,4)|_{3,5,7} = |2,9,20|_{3,5,7} = (2,4,6).$

Each digit is a small number, this makes RNS arithmetic becomes fast and simple.

Example 2: Negation, addition, subtraction and multiplication can be performed independently by operating on each digit individually. The following example in Fig 1 represents addition/subtraction/multiplication.

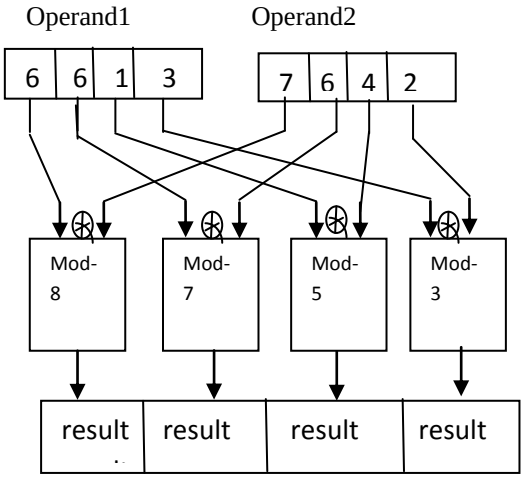

Fig 1: Operands Operation

2.2 Data conversion in residue number system

Data conversion is an important aspect in RNS. Data need to be converted from binary/decimal to RNS before any operation can be performed on them. However, the success of hardware realization depends on both data conversion and choice of moduli. Data conversion is divided into two categories: (i) Forward conversion and (ii) Reverse conversion. In this survey work, we explained each and gave a numerical example in each category.

2.2.1 Forward conversion

This is the conversion of binary/decimal number to RNS. In binary system, forward conversion can be represented as

$$|X_m| = |\sum^{N-1}_{j=0} b_j 2^j|_m \quad (1)$$

For any n-bit non negative integer X in the range

$0 < x \geq 2^n-1$, the hardware computation of forward conversion is based on Look up Table (LUT) [4].

Example, let $X = 45_{10}$ be a decimal number and we wish to determine the residue equivalent x_1, x_2. X is represented in 2's complement form as (b_N, b_{N-1}, b_{N-2}, ….b_2,b_1,b_0). N = 6 then (b_6, b_5, b_4, b_3, b_2, b_1, b_0) = 0101101_2. Let assume moduli to be (11 and 13), then the residue is given as

$$x_i = X \bmod m_i \quad i = 1,2$$

$$X_i = |b_N(m_i - (2^N \bmod m_i)) + \sum^{N-1}_{j=0} b_j(2^j \bmod m_i)|m_i \quad (2)$$

m_1=11, m_2 =13

Equation (2) will produce LUT 1 and LUT 2 below

LUT 1 for $m_i - (2^6 \bmod m_i)$

i =1	m_1 (11)	2
i =2	m_2 (13)	1

LUT 2 for $2^j \bmod m_i$

		J=5	J=4	J=3	J=2	J=1	J=0
i =1	m_1	10	5	8	4	2	1
i= 2	m_2	6	3	8	4	2	1

$x_1 = |0x2 + 1x10 + 0x5 +1x8 +1x4 +0x2 +1x1| \bmod 11 = 1$
$x_2 = |0x1 +1x6 +0x3 +1x8 + 1x4 + 0x2 +1x1| \bmod 13 = 6$
Means that the residue of 45 with moduli 11 and 13 is 1,6
Written as $45 = (1,6)_{RNS(11|13)}$ or $|45|_{11,13} = (1,6)$

2.2.2 Reverse conversion in residue number system

Reverse conversion is the conversion of residue number system to a conventional number, i.e binary/decimal numbers. The success of reverse conversion depends on forward conversion. Reverse conversion is based on two popular algorithms: Chinese Remainder Theorem (CRT) [5],[6] and Mixed Radix Conversion (MRC) [7],[8] algorithms. The use of CRT entails a large modular adder whereas MRC is a sequential process that requires a number of Look–Up Table (LUT) [9].

2.2.2.1 The Chinese Remainder Theorem (CRT)

CRT is a very useful algorithm for reverse conversion; it assumes that a number will have a unique representation in RNS if we chose appropriate moduli for the RNS. The algorithm involves computation of inverse and is given by

$$X = |\sum^{nl}_{i=1} |x_i M_i^{-1}|_{mi} M_i|_M \quad (3)$$

Equation (5) can be applied to solve the Sun Tzu riddle described in our section 1.

$x_1 = 2 \bmod 3$; $x_2 = 3 \bmod 5$; $x_3 = 2 \bmod 7$
m_1= 3, m_2=5, m_3= 7
M = 3x5x7 = 105
$M_1 = M/3 = 35$; $|M_1^{-1}|_3 = 2$
$M_2 = M/5 = 21$; $|M_2^{-1}|_5 = 1$
$M_3 = M/7 = 15$; $|M_3^{-1}|_7 = 1$

Substitute these values into equation (3)

$X = |\sum^3_{i=1} |x_i M_i^{-1}|_{mi} M_i|_M$
$X = \| x_1 M_1^{-1}|_{m1} M_1 + |x_2 M_2^{-1}|_{m2} M_2 + |x_3 M_3^{-1}|_{m3} M_3 \ |_M$
$= \| 2 \ x2|_3 \ x \ 35 \ + |3 \ x \ 1|_5 \ x \ 21 \ + |2 \ x \ 1 \ |_7 \ x \ 15 \ |_{105}$
$= \| \ 4|_3 \ x \ 35 \ + |3 \ |_5 \ x \ 21 \ + |2 \ |_7 \ x \ 15 \ |_{105}$
$= | \ 1 \ x \ 35 + 3 \ \ x \ 21 \ + 2 \ x \ 15 \ |_{105}$
$= | \ 35 \ + 63 \ + 30 \ |_{105}$
$= | \ 35 \ + 63 \ + 30 \ |_{105}$
$= |128|_{105}$
$= 23$

CRT is used in many applications like computing, cryptography, coding theorem etc. When it was discovered by Sun Tzu, a complete proof was not offered. The complete proof of the algorithm was offered by *Aryabhata*, an Indian Mathematician. A schematic diagram of CRT is depicted in Fig 2 below.

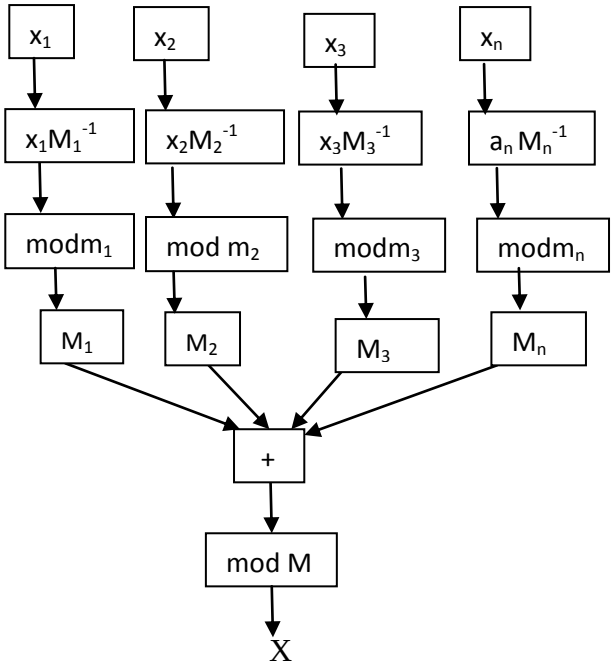

Fig.2: A Schematic diagram of CRT

2.2.2.2 The mixed radix conversion

Let the moduli set (m_1, m_2, m_3, …, m_n) has the corresponding RNS (x_1, x_2, x_3, …, x_n) and a set of digits (a_1, a_2, a_3, …, a_n) be the mixed radix digits respectively, then the corresponding decimal equivalent of the residues can be obtained using the following algorithm:

$$X = a_1 + a_2 m_1 + a_3 m_1 m_2 + \ldots.$$

The mixed radix are given by the following :

$$X = a_1 + a_2 m_1 + a_3 m_1 m_2$$

where $a_1 = x_1$

$$a_2 = |(x_2 - a_1) m_1^{-1}|_{m2} \quad (4)$$

$$a_3 = |(x_3 - a_1) m_1^{-1} - a_2 m_2^{-1}|_{m3} \ldots$$

.

.

$$a_k = |(((x_k - a_1)|m_1^{-1}|_{mk} \ldots.$$

The schematic diagram is shown in Fig 3 below

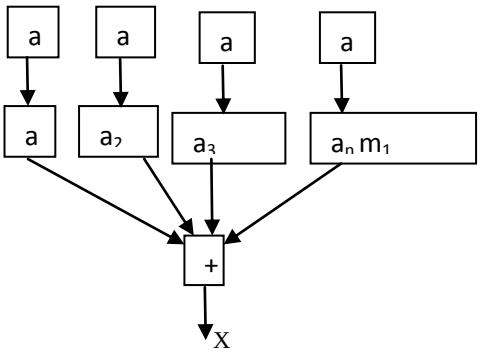

Fig. 3:. A schematic diagram of MRC

2.3 Moduli choice

The following points should be considered in the choice of RNS moduli

<> They must be relatively primed
<> The smaller the moduli, the faster the arithmetic operations
<> The higher the dynamic range of the moduli set, the faster its forward conversion and the slower its reverse conversion.
<> To avoid overflow, the dynamic range should be large enough.
<> Efficiency of the RNS moduli should be considered and high efficiency is more desirable, example the RNS (15|13|11)
- require 12 bits
- it can represent $2^{12} = 4096$, whereas only 2145 numbers are presented
- the efficiency is 52%

<> Select prime numbers in sequence until a desired dynamic range is obtained
<> Moduli numbers can be restricted to power of 2.

2.4 DNA in literature

DNA has been discussed in literature by many researchers to solve some problems on digital computer arithmetic, image processing [10], multi layer data encryption [11], digital image processing [12], computer networks [13], secure encryption techniques using DNA computations [14], Fourier transformation and many other fields that involve arithmetic algorithm. If moduli sets are properly chosen, RNS can be applied to the listed areas for a better achievement, because most computations are done in the embedded processor and secondly the absence of carry frees propagation is also a fundamental factor.

3. RNS AND BIOINFORMATICS

The availability of RNS processor makes it applicable in digital signal processing (DSP), image processing, speech processing, computer security (cryptography), computer network and many other fields that involve arithmetic algorithm.

Smith-Waterman (SWA) is a popular algorithm used in bioinformatics for the analysis of DNA. The algorithm was first proposed by Temple F. Smith and Michael S, Waterman in 1981 [29]. Like the Needleman-Wunsch algorithm of which it is a variation. The Smith–Waterman Algorithm performs local sequence alignment; that is, for determining similar regions between two strings or nucleotide or protein sequence [30].

DNA as being referred "code of life and blue print of biological life from beginning to the end" [15], is the primary genetic and hereditary material in all living organisms [16] which contains biological instructions and store biological information.

The total number of genetic sequence in an organism is known as genome. Human genome comprise of over 3 billion of DNA base pairs of which 99.9% are identical from one person to the other [16]. The residual 0.1% allows us to differentiate one person to the other, in terms of appearance, character and also allow effective human identification [17], [18], [19] through DNA typing or DNA profiling. It is made up of chemical building block called nucleotide [20] which contains nitrogen molecules Adenine, Cytosine, Guanine and Thymine associated with the following abbreviations A,C,G and T respectively [16].

The study of Deoxyribonucleic Acid (DNA) alignment is very important in bioinformatics; its analysis provides the following benefits:

1. Allows us to trace evolutionary trend.
   Example, if Bioinformatics are able to find the similarity between any two sequences, they will be able to trace and understand evolutionary trend between them [21]
2. Allows correlation of DNA information diseases. A relation between diseases and inheritance can be studied. Example genes identified to be involved in breast cancer [22].

Any research that aim to provide effective solution to the above points will require a very high speed sequence comparison.

4. SEQUENCE ALIGNMENT IN BIOINFORMATICS

Sequence alignment in bioinformatics computes similarities between any two sequences [23]. This describes the arrangement of the nucleotides in order to identify the region of similarity that may be a consequence of final structural or evolutionary relationship between them. Alignment finds level of similarity between query sequence and different database. It compares the DNA strands with each other through which we can identify genetically transmitted diseases [24].

There are many algorithms for computing sequence alignment, the most popular ones are FASTA [25], [26] and BLAST [27], [28]. The two algorithms are fast but at the expense of accuracy. SWA [29] is popularly known to be the most accurate algorithm for sequence alignment; however, the computational complexity makes the algorithm slow especially for a long sequence.

RNS has been successfully used in many applications that involve computations because of the embedded processor in RNS and the absence of carry propagation.

4.1 The Algorithm of Smith –Waterman for Local alignment
The algorithm is given by:

$$X(i,j) = Max \begin{cases} 0 \\ X(i-1,j-1) + S(a_i,b_j) \text{ match/mismatch} \\ X(i-1,j) + g \\ X(i,j-1) + g \end{cases} \quad (5)$$

X(i,0) = 0, X(0,j) = 0
where, X(i,0) = 0, X(0,j) = 0
X(i,j) => maximum similarity score between the two sequences
$S(a_i,b_j)$ => the similarity score of comparing sequence Ai to sequence Bi
d is the mismatch gap penalty in the comparison

SWA is a good algorithm used for computing homology and sequence alignment in genetic database. It compares two strings of nucleotides with high level of accuracy. However, the low speed of execution becomes its main challenge.

5. RNS-SWA ARCHITECTURE

For a Moduli set $\{2^n-1, 2^n, 2^n+1\}$, RNS-SWA architecture is shown in Fig 4 below

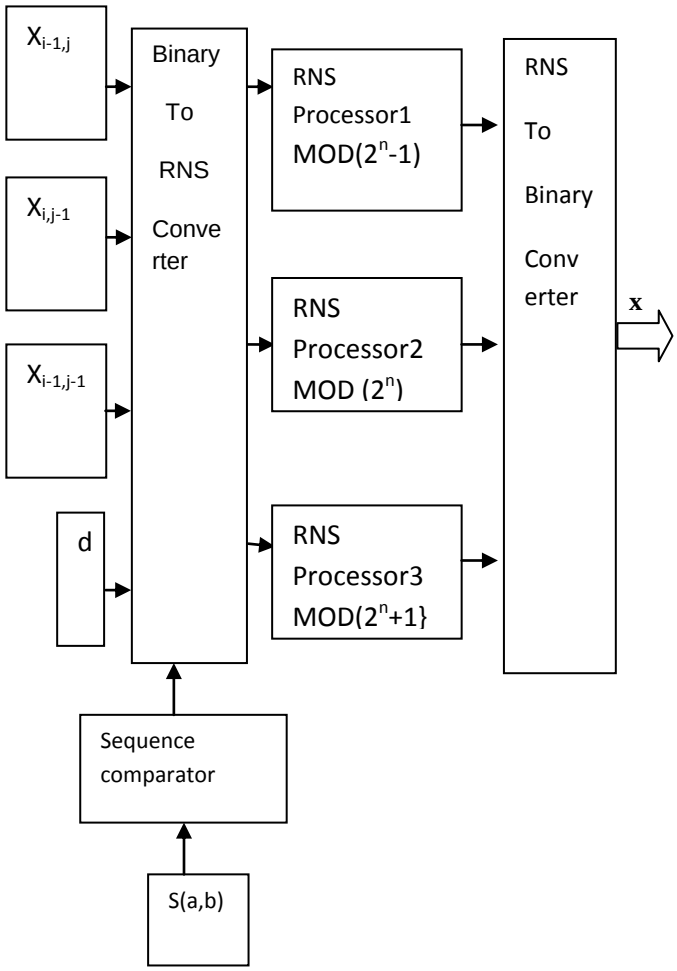

Fig. 4: Architecture of RNS-SWA

The hardware realization is based on the following

- The binary to residue converter (converter1) accepts the inputs X(i-1,j), X(I,j-1), X(i-1,j-1), d and S(i,j) and forwards to RNS processor.
- The RNS processor sends the result to converter2. Computation is done in the embedded processor; the absence of carry propagation in RNS will enable realization of high-speed and low-power consumption.
- Converter2 (residue to binary converter) then converts the latest result to binary/decimal number, X(i,j).

The above steps take the advantages of RNS arithmetic and make the speed of conversions better depending on the moduli choice.

However, in future research, we shall determine to use FPGA/VHDL to implement the RNS-SWA architecture using RNS as a tool to have a better improvement on the computational challenges facing SWA.

6. CONCLUSION

Speed and accuracy are the major challenges in the field of bioinformatics. The availability of embedded RNS processor and the absence of carry propagation in RNS brightened the possibility of realization of high-speed and low-power consumption. This paper presented a review of RNS, showed that RNS can have its way and can be applied to

bioinformatics, which goes along to reduce the computational issue in SWA. However, in future research, FPGA/VHDL could be used to implement the RNS-SWA architecture using RNS as a tool to have a better improvement on the computational challenges facing SWA.

7. REFERENCES

[1] M.A Soderstrand, W.K Jenkins, G. Julliet and F.J Taylor "Residue Number System Arithmetic", Modern Application in Digital Signal Processing 1986.

[2] Szabo N.S, Tanaka R.I " Residue number and its applications to computer technology" McGraw Hill N.Y 1967

[3] K.A Gbolagade, R. Chares, L. Sousa, S.D Cotofana " Residue –to- binary converters for the $\{2^{2n+1}-1, 2^{2n}, 2^{n}-1\}$ moduli set. 2nd IEEE International conference in adaptive science & technology. Pp 26 – 33 Accra Ghana Dec. 2009

[4] Neha Singh " An overview of Residue Number System" National seminar of devices, circuits and communication, Nov 2008

[5] S. J. Piestrak, "Design of residue generators and multi-operand modular adders using carry-save adders", IEEE Trans. Comput., vol. 423, no. 1, pp. 68-77, Jan. 1994.

[6] K. M. Elleithy, M. A. Bayoumi, "Fast and flexible architectures for RNS arithmetic decoding", IEEE Trans. Circuits Syst., vol. 39, no. 4, pp. 226-235, Apr. 1992.

[7] C. H. Huang, "A fully parallel mixed radix conversion algorithm for residue number applications", IEEE Trans. Comput., vol. 32, no. 4, pp. 398-402, Apr. 1983.

[8] H. M. Yassine, W. R Moore, "Improved mixed-rAdix conversion for residue number system architectures", IEE Proc.-G, vol. 138, no. 1, pp. 120-124, Feb. 1991.

[9] Bin Cao, Chip-Hong Chang, Thambipillai S. "Adder based residue to binary converters for a new balanced 4-moduli set" Proceedings of the 3rd International Symposium on Image and Signal Processing and analysis (2013)

[10] Gabriel K A, Emmanuel A, "application of RNS to image processing using orthogonal transformation. Conference Paper · June 2015 DOI: 0.1109/ICCSN.2015.7296177
Conference: International Conference on Communications software and Networks

[11] M. I. Youssef, A. E. Emam, M. Abd Elghany "Multi-Layer Data Encryption using Residue Number System in DNA Sequence" International Journal of Computer Applications (0975 – 8887) Volume 45– No.10, May 2012

[12] Shahram Moharrami and Davar Kheirandish Taleshmekaeil "The Application of the Residue Number System in Digital Image Processing: Propose a Scheme of Filtering in Spatial Domain "Research Journal of Applied Sciences7: 286-292; DOI: 10.3923/rjasci.2012.286.292;

[13] Zarandi A.A.E. (2017) RNS Applications in Computer Networks. In: Molahosseini A., de Sousa L., Chang CH. (eds) Embedded Systems Design with Special Arithmetic and Number Systems. Springer, Cham. DOI 10.1007/978-3-319-49742-6_14

[14] Drashti O. Vadaviya1, Purvi H. Tandel "Secure Encryption Techniques Using DNA Computation" International Journal of Modern Trends in Engineering and Research (IJMTER) Volume 2, Issue 7, [July-2015] Special Issue of ICRTET'2015.

[15] Eric B. The importance of DNA in the human cell" pc magazine, New York; 2010. Available:http://sciencing.com/importancedna-human-cell-19447.htm

[16] Hassan Kehinde Bello and Kazeem Alagbe Gbolagade "A Survey of Human Deoxyribonucleic Acid" British Journal of Applied Science & Technology. 21(5): 1-10, 2017; Article no.BJAST.32463 ISSN: 2231-0843, NLM ID: 101664541; DOI: 10.9734/BJAST/2017/32463

[17] Jennifer Romeika M, Fei Yan. Recent advances in forensic DNA analysis. J Forensic Res. 2013;S12:001. DOI: 10.4172/2157-7145.S12-001

[18] Lucia B, Pietro L. Forensic DNA and bioinformatics. Briefings in Bioinformatics. 2007;8(2):117-128. DOI: 10.1093/bib/bbm006 (Advance Access publication March 24, 2007)

[19] Adrian L, Jennifer EL. Forensic DNA profiling: State of the art. Journal of Research and Reports in Forensic Medical Science. 2014;4:25–36.

[20] Barbora K, Michael K, Norbert L, Martin F, Marc B, Daniel B, et al. Visualization of biomolecular structures. Eurographics Conference on Visualization (eurovista); 2015. DOI: 0.2312/eurovissta.2015112

[21] E.Y Baagyere, K.O Boateng, K.A Gbolagade " Bioinformatics: An important area application of Residue Number System. Journal of Engineering and Applied Sciences 6(2):174-179, 2011

[22] Miki Y., J. Swensen, D. Shattuck-Eidens, P.A. Futreal and K. Harashman et al 1994. " A strong candidate for the breast and ovarian cancer susceptibility genes BRCA1. Science 266: 66-71

[23] Kwame O and Edward Y.B. (2012) " A smith-waterman algorithm acceleration based on RNS" IJECE ISBN 1974-2166 VOL5 PP99-112

[24] Laiq Hasan Zaid Al-Ars Zubair Nawaz Koen Bertels "Hardware Implementation of the Smith-Waterman Algorithm Using Recursive Variable Expansion" Delft University of Technology Computer Engineering Laboratory Mekelweg 4, 2628 CD Delft, The Netherlands

[25]] Lipman, DJ; Pearson, WR (1985). "Rapid and sensitive protein similarity searches".Science 227 (4693): 1435–41. DOI:10.1126/science.2983426. PMID 2983426

[26] W. R. Pearson and D. J. Lipman, "Rapid and Sensitive Protein Simlarity Searches", Science, vol. 227, pp: 1435–1441, 1985.

[27] S. F. Altschul, Gish, W. Miller, W. Myers and D. J. Lipman, "A Basic Local Alignment Search Tool", Journal of Molecular Biology, vol. 215, pp: 403–410, 1990

[28] Altschul Stephen, Gish Warren, Miller Webb, Myers Eugene, Lipman David(1990) "Basic Local Alignment Search tools" Journal of molecular Biology 215(3): 403-410. DOI:10.1016/S0022-2836(05)80360-2 PMID 2231/12

[29] Smith Temple F., Waterman Michael (1981) "Identification of Common Molecular Subsequences" (PDF).Journal of Molecular biology 147: 195–197. DOI10.1016/0022-2836(81) 90087-5. PMID 7265238

[30] Gotoh O. " An improved algorithm for matching biological sequences; Journal of Molecular Biology,162(3):705-708,1982.