SHORT TANDEM REPEATS DETECTION IN DNA SEQUENCES USING MODIFIED S-TRANSFORM

[1, 3]S. D. Sharma, [2]Rajiv Saxena, [1]S.N. Sharma, [3]A. K. Singh
[1]Department of ECE, Samrat Ashok Technological Institute, Vidisha, M.P., India
[2]JAYPEE University, AnoopShahr, U.P., India
[3]Department of ECE, JAYPEE University of Engineering & Technology, Guna, M.P., India

ABSTRACT

Identification of the short tandem repeats in DNA sequences is a challenging problem for the scientists and engineers in the current era. The detection of the short tandem repeats is also an important part of gene annotation and also it is useful to identify the various hereditary diseases and human identity, etc. The several methods have been developed to find the short tandem repeats, and these are based on string matching and signal processing. In this paper, we have proposed a signal processing method which is based on the modified S-transform for the perfect and imperfect STRs detection. The detection performance of the proposed method has been compared, and it is found that the proposed method is more sensitive than TRF and parametric spectral method for STRs detection in DNA sequences.

KEYWORDS:—*DNA sequences, short tandem repeats, Modified S-transform, Bioinformatics*

I. INTRODUCTION

Genomics signal processing is important to understand the biological functionalities in living organism. The genome contained the deoxyribose nucleic acid (DNA). The DNA consist the four nucleotides, adenine (A), thymine (T), cytosine(C), and guanine (G). The tandem repeat patterns of the nucleotide 1bp-6bp have been found in the micro satellite region of the DNA sequences [1]. Tandem repeat pattern of the nucleotides 1-6bp is called short tandem repeats (STRs) and these are classified as perfect and imperfect. The perfect STRs contain the exact copies of the repeated patterns while the imperfect STRs consists inexact copies of the repeated patterns due to insertion, deletion and substitution [2]. The number copies of the STRs increases or decreases are responsible for the changes during the formation of the protein and affect the gene regulation in coding and non-coding regions of the DNA sequences respectively. The STRs associate with diseases such as Fragile-X syndrome, Huntington's disease, Frederick's ataxia, and 40 other neurological, neurodegenerative and neuromuscular diseases [3]. STRs also find the applications in forensics analysis, DNA Fingerprinting, population studies, linkage analysis, evolutionary studies, behavior of living organism, etc[4]. Therefore, the identification of the short tandem repeats in DNA sequences is an active research area of the genomic signal processing, and several methods have been developed. Basically, the STRs identification algorithm focuses to identify its period, pattern, location and number copies. The Methods for the STRs detection in the DNA sequences can be classified as a computational method (heuristic, combinatorial and dynamic programming) and signal processing methods [2]. The signal processing methods are more sensitive to identify the imperfect STRs in the DNA sequences and the signal processing methods also have the less computational complexity than computational methods. Therefore, we are focusing on the signal processing methods. The short time periodicity transform based algorithm applied to detect the tandem repeats of the specific periodicity in the DNA sequences [5] but this algorithm faces the problem of computational complexity, and period must be known in advance. Therefore, the faster signal processing tool, discrete Fourier transform (DFT) has been introduced to detect the tandem repeat of the unknown periodicities in the DNA sequences. The DFT is also a powerful signal processing tool to identify the imperfect tandem repeats [6]. The spectral repeat finder (SRF) based on the DFT is used to detect the repetitive patterns present in the sequences and in this method, the power spectrum of a DNA sequence is used to identify the

periodicity of the repetitive patterns. The specific regions of the sequence that contain the repetitive pattern are located using a sliding window analysis, and then exact search method is used to find the repetitive pattern [7]. The modified Fourier product spectrum (MFPS) method is highly sensitive to identify the imperfect tandem repeats [8]. The quaternion periodicity transform (QPT) and exactly periodic subspace decomposition (EPSD) had been introduced to provide the solution of the problem of multiple periodicity in DFT and STPT based methods. QPT and EPSD also have the more computational complexity than DFT based methods [9, 10]. The optimized moving window spectral analysis (OMWSA) is an Autoregressive (AR) model based spectral analysis, which is used to detect the tandem repeats detection in DNA sequence and visualizes the repeating patterns into a two-dimensional plot of location and frequency (spectrogram) [11].This spectrogram shows the repeating patterns and their location without any prior knowledge. The AR model based method for the short and long tandem repeats identification in DNA sequences is more accurate and efficient than other methods [12]. DFT and AR model based methods uses the sliding window to localize the tandem repeats patterns in the DNA sequences have been reported. Therefore, DFT and AR model based methods missed the repeating patterns present in the sequence due to the fixed window length. Therefore the modified S-transform (MST) based method has been developed to provide the solution of the fixed window length and to capture the short tandem repeats present in the DNA sequences efficiently. The modified S-transform adjusts the length of the Gaussian window by varying the appropriate standard deviation with respect to the frequency of the repeating pattern of the STRs.
This paper is organized as follows in section 2. Modified S-transform, in Section 3.Algorithm for STRs identification, in section 4. The experimental results and performance comparison, and in Section 5. Conclusion and future work has been presented.

II. MODIFIED S-TRANSFORM

The S-transform of the signal x (t) is defined as

$$S(t,f,\sigma) = \int_{-\infty}^{\infty} x(\tau)\, G(t-\tau,\sigma) e^{-j2\pi f\tau} d\tau \quad (1)$$

where $G(t-\tau,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-\tau)^2}{2\sigma^2}}$, is a variable gaussian window function, σ = standard deviation, t = time, τ=central position of the window and f= frequency. The advantage of the S-transform over the STFT is that the standard deviation σ is a function of frequency f i.e. $\sigma(f) = \frac{1}{|f|}$, so the length of the Gaussian window will vary with respect to the frequency then the equation (1) becomes

$$S(t,f) = \int_{-\infty}^{\infty} x(\tau) \left\{ \frac{|f|}{\sqrt{2\pi}} e^{\frac{-(t-\tau)^2 f^2}{2}} e^{-j2\pi f\tau} \right\} d\tau \quad (2)$$

The S-transform also defined as a short time Fourier transform with variable window [13]. The S-transform provides the progressive time frequency resolution but faces the problem of poor energy concentration in the time frequency plane due to the standard deviation of the Gaussian window is inversely proportional to the frequency. For higher frequencies, the window length is small, and it contains fewer numbers of period resulting the energy concentration is poor. Similarly, for the lower frequencies, window length is large, and it contains the more number of periods within it and energy concentration is more. To improve the energy concentration in time frequency plane, Mansinha *et.al* proposed the generalized S-transform by taking$f = f/\gamma$ and a window length controls parameter (γ) gets the single value for all frequencies [14, 15, 16]. Due to the single value of γ the generalized S-transform does not able capture all the frequencies present in the signal. In this, research paper the S-transform has been modified to identify STRs presents in the DNA sequences. STRs have the periodicity from 2bp to 6bp. The modified S-transform has appropriate value of the parameter γ to control the window length for each specific frequency to improve the time and frequency resolution as well. The relation between γ and f have been established using the study of the synthetic dataset. Synthetic dataset consist the patterns of period 2bp to 10bp has been shown in Fig.1.The value of γ has been selected from 8.3 to 2.3 to vary the window length to efficiently capture the corresponding frequency f from$\frac{1}{2}$ and $\frac{1}{10}$ (period of repeated pattern 2bp to 10bp) and spectrogram of synthetic data set is shown in Fig.2.

CTCTCTCTCTCTCTCTGGGGGGGGGGGGGGGGGGGGGGGGTGATGATGATGATGATGATGAT
GATGATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGCATGCATGCATGCATGCA
TGCATGCATGCATGCATGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCC
AATCCAATCCAATCCAATCCAATCCAATCCAATCCAATCCAATCCAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCAATTCCAATTCCAATTCCAATT
CCAATTCCAATTCCAATTCCAATTCCAATTCCAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGACCCATGACCCATGACCCATGACC
CATGACCCATGACCCATGACCCATGACCCATGACCCATGACCCAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG
AATCATTGAATCATTGAATCATTGAATCATTGAATCATTGAATCATTGAATCATTGAATCATTGAATCA
TTGAATCATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTGGAGTTATTGGAGTTATTGGAGTTATT
GGAGTTATTGGAGTTATTGGAGTTATTGGAGTTATTGGAGTTATTGGAGTTATTGGAGTTCCCCCCGAG
GGGTATCGAGGGGTATCGAGGGGTATCGAGGGGTATCGAGGGGTATCGAGGGGTATCGAGGGGTATC
GAGGGGTATCGAGGGGTATCGAGGGGTATC

Fig.1. Synthetic Data Set of DNA

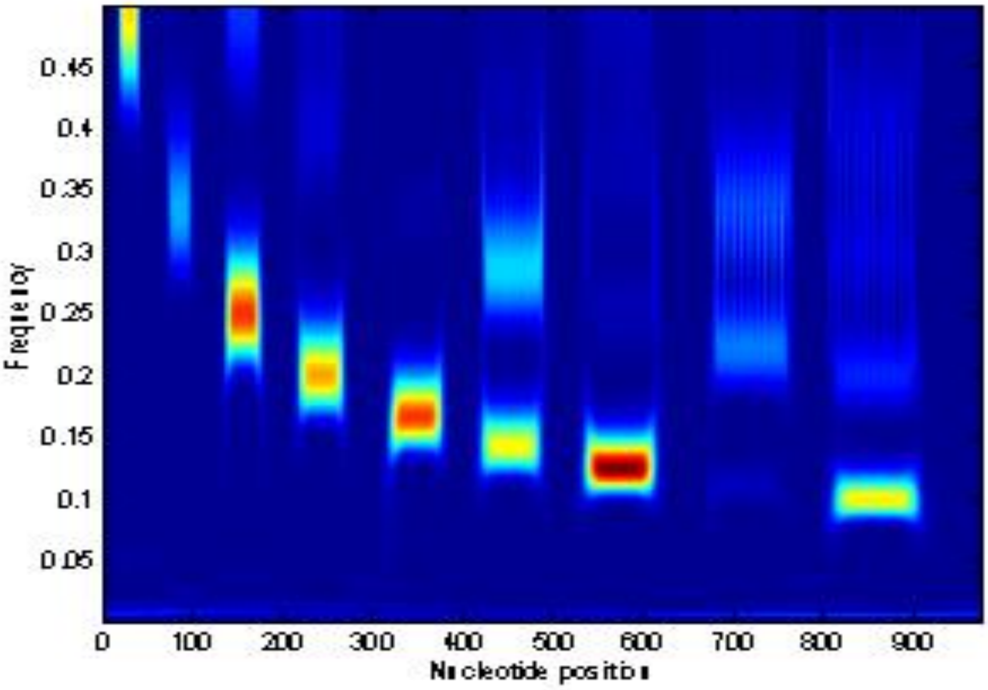

Fig.2. Spectrogram of the synthetic dataset using S-transform

Fig.2 shows the spectrogram of the synthetic dataset using S-transform. This spectrogram explains the good time resolution and poor frequency resolution because window length is inversely proportional to the frequency. For the higher frequencies, the window length is small and for lower frequencies, the window length is large. Therefore, the S-transform does able to capture all the frequencies. So to get the appropriate window length to detect the corresponding period (period =1/frequency), we have to introduce the frequency-dependent control parameter γ using simulation studies, which is given by

$$\gamma(f) = 15f + 0.8 \quad (3)$$

Using equation (3), modified S-transform (MST) for short tandem repeat detection is defined as

$$MST(t,f) = \int_{-\infty}^{\infty} x(\tau) \frac{|f|}{(15f+0.8)\sqrt{2\pi}} e^{\frac{-(t-\tau)^2 f^2}{2(15f+0.8)^2}} e^{-j2\pi\tau} d\tau \quad (4)$$

The spectrogram using modified S-transform has been shown in Fig.3and shows the detection of the period 2bp to 10bp in the time frequency plane with good time and frequency resolution.

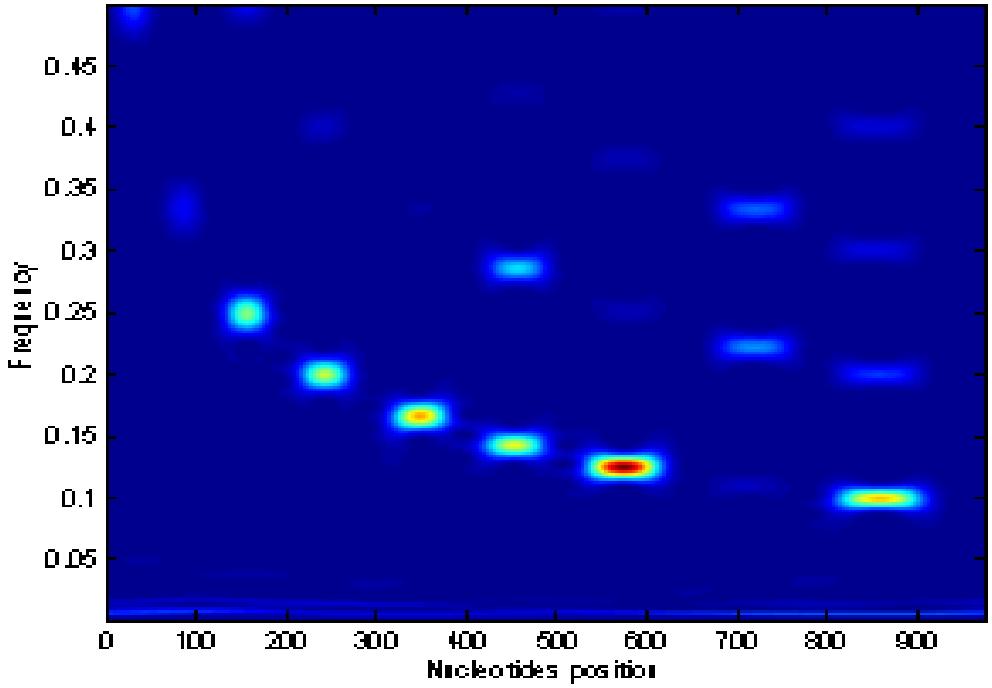

Fig.3. Spectrogram of the synthetic dataset using S-transform

So based on the above study it is found that MST may improve the detection performance of the STRs in the DNA sequences.

III. STRS DETECTION METHOD

In this research paper, we have proposed a method using MST for the STRs identification in the DNA sequences. The appismellifera microsatellite DNA sequence (acc.no.U73927) has been selected in the experiment. The proposed method has four steps –

3.1 Complex mappingof nucleotides

Compex mapping applied to converts the nucleotide sequence into complex sequence because it is usefull for the repeats detection in the DNA sequences due to its features such as A-T and C-G are complex conjugate, reflecting complementary feature of nucleotides, more accurate in gene prediction(period-3 detection). Also the complex mapping scheme reduces the 75% omputational complexity compare to the binary mapping scheme [17].For example $x = [A, T, G, C]$ converts in to complex sequence $x(n) = [1 + i, 1 - i, -1 - i, -1 + i]$.

3.2 Applicationof MST

MST is applied to the complex sequence and it is given by

$$MST(t,f) = \int_{-\infty}^{\infty} x(\tau) \frac{|f|}{(15f+0.8)\sqrt{2\pi}} e^{\frac{-(t-\tau)^2 f^2}{2(15f+0.8)^2}} e^{-j2\pi\tau} d\tau \quad (5)$$

and calculate the power spectrum of the sequence using squared magnitude of MST coeficients that is given by

$$S_{MST}(t,f) = |[MST(t,f)]|^2 \quad (6)$$

Fig.4 show spectrogram of the DNA sequences (acc.no.U73927) using proposed method i.e.MST. In this spectrogram the horizontal axis represents the location of the DNA sequences and vertical axis represents the normalized frequency and highlighted region shows the PSD corresponding to the frequency and its location.

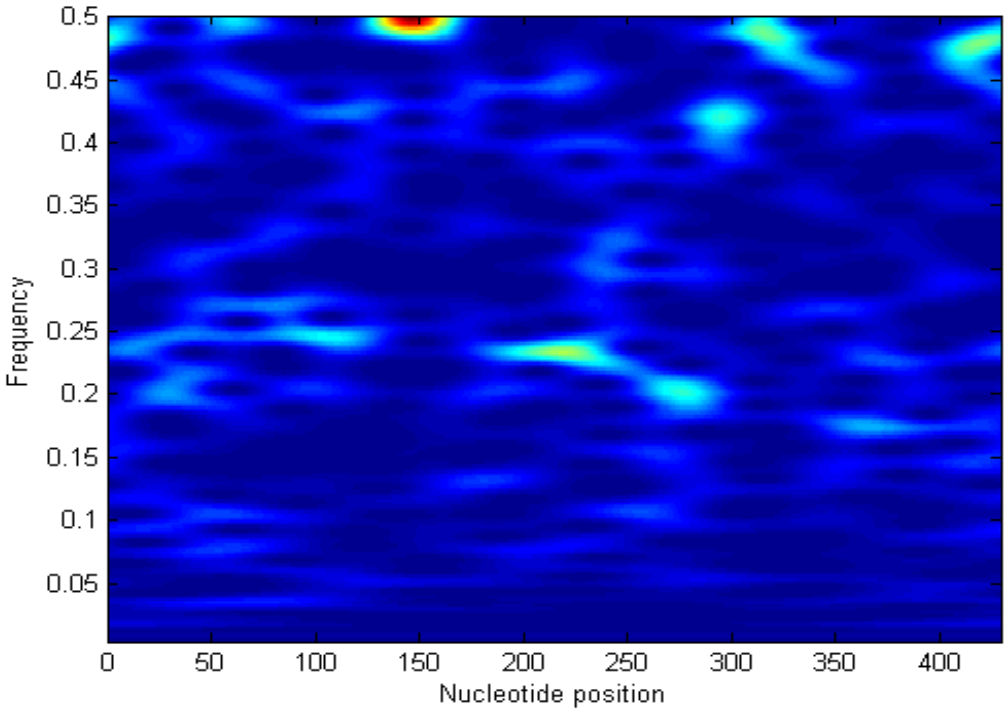

Fig.4. Spectrogram of the DNA data (Acc. no.U73927) using MST

3.3 Select the significant power spectral density (psd) and its location for the respective period using a threshold

Fig.4. shows the spectrogram and it is a visual representation of the PSD of the DNA sequences in the time-frequency plane. The highlighted region shows the corresponding STRs patterns of nucleotide of the DNA sequences. However, it is difficult to find, all the location of the STRs patterns through this spectrogram.To provide the solution for this difficulty, we have to extract PSD corresponding to the frequency 'f'(f=1/period-Lbp, where L=2,3,4,5and 6) and compare it with the threshold 'S_{Th}'at each position. The value of threshold S_{Th} is given by the following relation-

$$S_{Th} = 4 \times \text{mean}(S_{MST}(t,f)) \quad , \qquad (7)$$

The comparison of the PSD of the period-2bp, period-3bp, period-4bp, period-5bp and period-6bp with the selected threshold S_{Th} have been shown in the Fig.5(a) to Fig.5(e). The position where the power spectral density of the corresponding frequency is greater than or equal to 'S_{Th}' then the value of the power spectral density is set to be '1' otherwise '0'. After this comparison, we have got another spectrogram, which shows the location of the STRs of each frequency or period of the nucleotides in the DNA sequences in the Fig.6.

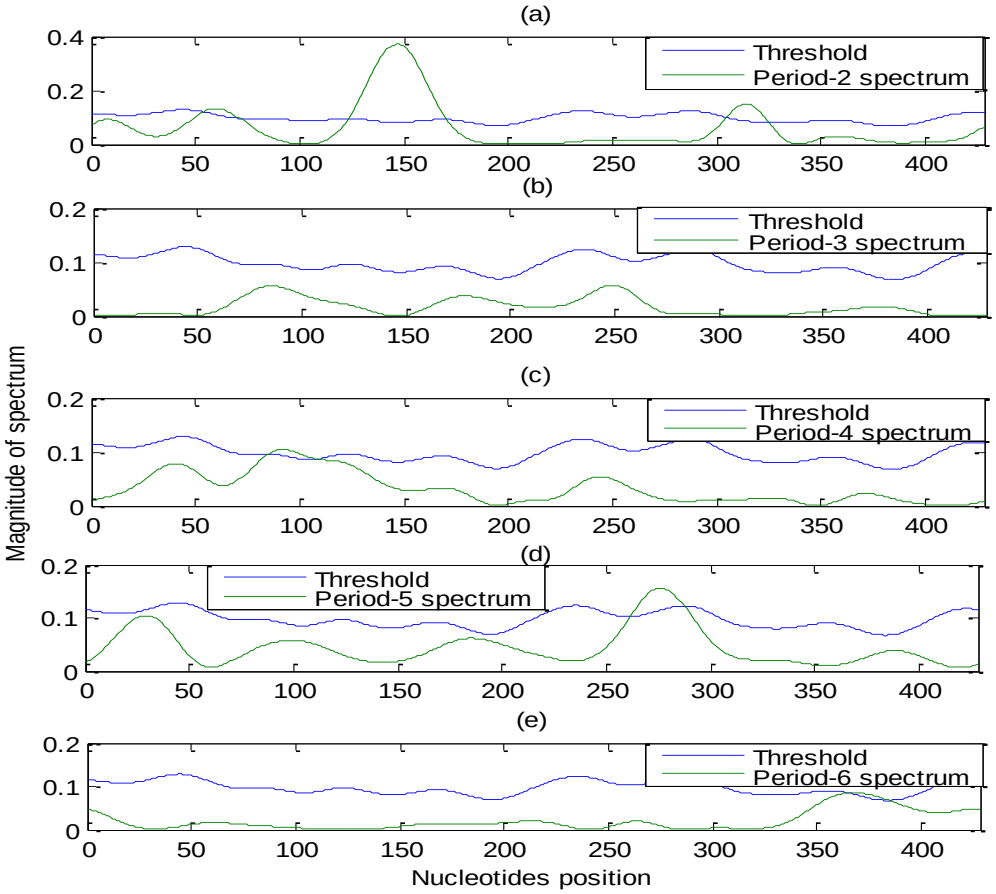

Fig.5. Power spectrum with respect to periods

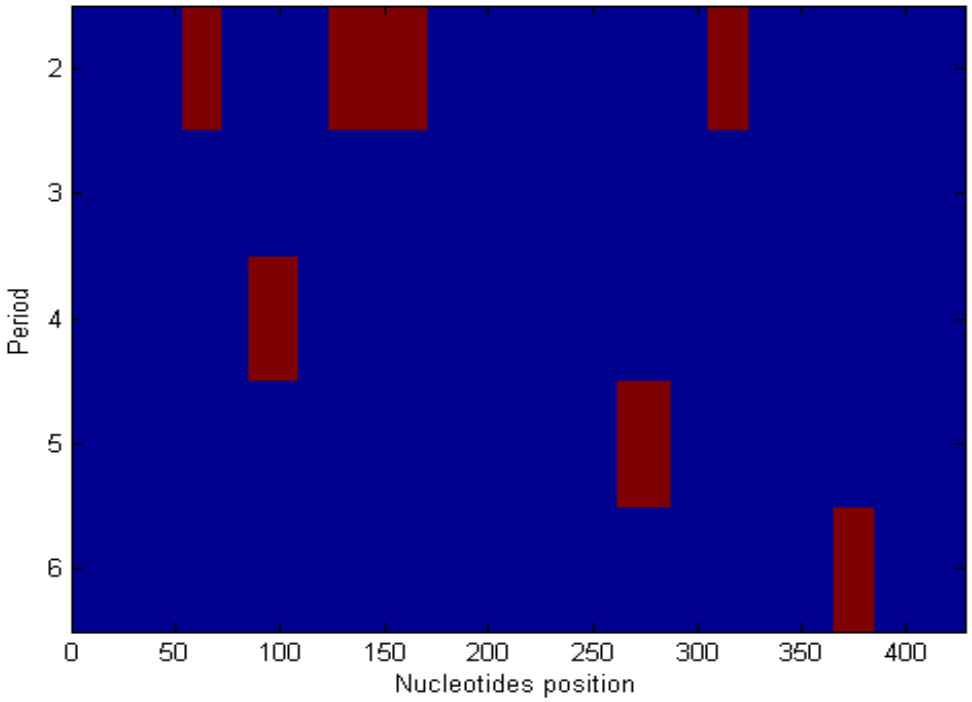

Fig.6. Time-frequency plot after verification phase for the sequence U73927

3.4 Verificationof the locationof STRS

The verification of the identified location of the STRs is important to discard the false identified location of the STRs. For the verification we have followed the procedure employed in [12,18].The STRs presents in the sequence after verification phase are tabulated in Table.1.

Table.1. STRs identified in microsatellite dataset (acc.no.U73927)

Period	Location detected using proposed algorithm	Location detected after verification	Pattern-No. of copies ($\mu = 0.6$)	No. copies (Total)
2	54-71 124-170 305-324	56-61 126-129 136-159 307-324	TC-3 TC-2 TA-12 TA-9	26
4	85-108	85-92	TTCC TTCA	2
5	262-286	262-276	TTTAA TCCAA TTCCA	3
6	340-400	------	--------	0

IV. EXPERIMENTAL RESULTS

For the experiment on real datasets, we have selected complete sequences of Apismellifera (Honeybee) from the acc. no. U73917 to U73934 which comprises the short tandem repeats[19].The performance comparison of the proposed method has been done with existing methods. The STRs detection performance of the proposed method has been shown in the Table.2 and Table.3.

Table.2: Performance comparison of proposed algorithm for the STRs detection

Data Set	Method	P	Location detected using proposed algorithm	Location detected after verification	Pattern-No. of copies ($\mu = 0.6$)
U73917	TRF	2	147-186	NA	20
	Parametric	2	144-180	145-148 149-152 155-172	CG-2 AG-2 AG-9
	Proposed	2	138-197	145-148 149-152 155-172 178-186	CG-2 AG-2 AG-9 GA-5
	TRF	6	……..	NA	….
	Parametric	6	…….	……	…
	Proposed	6	193-215	……	….
U73918	TRF	2	88-201	NA	AG-60.5
	Parametric	2	31-50 102-154 158-209	32-35 46-49 107-120 122-139 148-153 169-172 177-182 186-202	AG-2 AT-2 GA-7 AG-9 GA-3 AG-2 GA-3 GA-8.5
	Proposed	2	73-86 102-141 173-210	…. 107-120 122-139 177-182	….. GA-7 AG-9 GA-3

				186-202	GA-8.5
U73919	TRF	2		NA	
	Parametric	2	301-376	312-341 360-369	TC-15 TC-5
	Proposed	2	296-379 415-434	312-341 360-369 415-418 422-425 429-432	TC-15 TC-5 TG-2 CT-2 AC-2
	TRF	4	311-368	NA	**TCTC-14.5**
	TRF	5	…….	…….	……..
	Parametric	5	…….	……..	………..
	Proposed	5	201-229 480-512 594-635	202-211 212-221 590-604 599-608 608-617	TTGAC TTGAC TGTTA TTTGA TTTCT TGTCT TTTTT AACGA AAAGA AGATT AAAAT
U73920	TRF	2	…….	NA	…….
	Parametric	2	……	…..	…….
	Proposed	2	189-208	----	-------
	TRF	3	…….	NA	…….
	Parametric	3	……	…..	…….
	Proposed	3	245-293	245-250 250-255 262-270 271-276 279-287	AAT-2 TCT TCG TGA TAA TCA ACG-2 ACC ACC AGC
U73921	TRF	2	….	NA	……
	Parametric	2	292-344	310-331	AG-11
	Proposed	2	199-225 287-360	199-202 217-220 310-331	AG-2 AG-2 AG-11
	TRF	4	……	NA	----
	Parametric	4	……..	----	----
	Proposed	4	144-170	157-164	GAGG GATG
	TRF	6	……..	NA	
	Parametric	6	……..		
	Proposed	6	97-130	…..	…..
U73922	TRF	2	……..	NA	……
	Parametric	2	259-327	260-263 269-272 274-277 279-296 303-318	CT-2 TC-2 CT-2 CT-9 CT-8
	Proposed	2	249-334	260-263 269-272 274-277	CT-2 TC-2 CT-2

				279-296 303-318	CT-9 CT-8
	TRF	4	……	NA	……
	Parametric	4	…….	…….	…….
	Proposed	4	2-23	6-17	GCTT GCAT GCCT
	TRF	5	……..	NA	……..
	Parametric	5	……	……..	…….
	Proposed	5	51-80	…..	……
U73923	TRF	2	….	NA	----
	Parametric	2	……	……..	……..
	Proposed	2	7-52	9-12 47-50	TA-2 GA-2
	TRF	6	…….	NA	---
	Parametric	6	……	---	---
	Proposed	6	71-108	80-97	TTGGAA TTCGAT TTATAT
U73924	TRF	2	417--455	NA	TC-20
	Parametric	2	410-464	418-427 431-456	TC-5 TC-13
	Proposed	2	432-460	432-456	TC-12.5
	TRF	5	……..	NA.	…….
	Parametric	5	……..	……..	…….
	Proposed	5	308-333	312-321	CGATT CTATA
U73925	TRF	2	177--224	NA	GA-24
	Parametric	2	164-232	178-225	GA-24
	Proposed	2	161-238	178-225	GA-24
	TRF	4	…….	NA	…….
	Parametric	4	……	…….	…….
	Proposed	4	236-249	……	…….
U73926	TRF	2	66--112	NA	AG-23.5
	Parametric	2	53-124 233-266	67-112 119-122 243-258	AG-23 CG-2 TA-8
	Proposed	2	57-127 237-268	67-112 119-122 243-258	AG-23 CG-2 TA-8
U73927	TRF	2	135--159	NA	TA_12.5
	Parametric	2	121-172 293-331	126-129 136-159 297-300 307-324	TC-2 TA-12 TC-2 TA-9
	Proposed	2	54-71 124-170 305-324	56-61 126-129 136-159 307-324	TC-3 TC-2 TA-12 TA-9
	TRF	4	…….	NA	……
	Parametric	4	……..	…….	
	Proposed	4	85-108	85-92	TTCC TTCA
	TRF	5	…….	NA	……
	Parametric	5	……	……	……
	Proposed	5	262-286	262-276	TTTAA TCCAA TTCCA

	TRF	6	…….	NA	……
	Parametric	6	……	…….	…….
	Proposed	6	340-400	------	--------
U73929	TRF	2	91--120	NA	AG-15.5
	Parametric	2	86-132	86-89 92-95 96-121 129-132	AT-2 AG-2 GA-12 AG-2
	Proposed	2	86-143	86-89 92-95 96-121 129-132 136-139	AT-2 AG-2 GA-12 AG-2 AG-2
U73930	TRF	2	61--96	NA	AG-18
	Parametric	2	47-108	55-58 62-97	GA-2 AG-18
	Proposed	2	46-117	55-58 62-97	GA-2 AG-18
U73931	TRF	2	36--63	NA	GA_14
	Parametric	2	17-77	24-29 37-64 69-72	CA-3 GA-14 GA-2
	Proposed	2	29-77	37-64 69-72	GA-14 GA-2
U73932	TRF	2	115--242	NA	AG-68.5
	Parametric	2	117-169 189-242	117-130 141-162 192-195 198-201 203-228 235-238	GA-7 GA-11 TA-2 AG-2 GA-13 GA-2
	Proposed	2	107-165 190-238	117-130 141-162 192-195 198-201 203-228 235-238	GA-7 GA-11 TA-2 AG-2 GA-13 GA-2
	TRF	6	……..	NA	…….
	Parametric	6	……	…….	…….
	Proposed	6	21-47	21-38	GCCAAG GAGAAG GCGTAG
U73933	TRF	2	43--78	NA	CT-18
	Parametric	2	30-84	32-35 38-43 44-59 63-78	GC-2 AC-3 CT-8 TC-8
	Proposed	2	23-88 147-153 272-291	32-35 38-43 44-59 63-78 …. ….	GC-2 AC-3 CT-8 TC-8 …. ….
	TRF	6	….	NA	….
	Parametric	6	….	….	….
	Proposed	6	149-175	149-160	ATAAAG ATACCG
U73934	TRF	2	221--250	NA	GA-15
	Parametric	2	209-256	222-225	GA-2

			327-349	227-250 332-237	AG-12 CA-3
	Proposed	2	209-266 326-346	222-225 227-250 332-237	GA-2 AG-12 CA-3

Table.3: Performance comparison of proposed algorithm

Dataset	**Method**														
	TRF					**Parametric**					**Proposed Method**				
Period	**2**	**3**	**4**	**5**	**6**	**2**	**3**	**4**	**5**	**6**	**2**	**3**	**4**	**5**	**6**
U73917	20	-	-	-	-	13	-	-	-	-	19	-	-	-	-
U73918	60	-	-	-		36	-	-	-	-	27	-	-	-	-
U73919	-	-	14	-	-	20	-	-	-	-	26	-	-	11	-
U73920	-	-	-	-	-	-	-	-	-	-	-	12	-	-	-
U73921	-	-	-	-	-	11	-	-	-	-	15	-	2	-	-
U73922	--	-	-	-	-	23	-	-	-	-	23	-	3	-	-
U73923	-	-	-	-	-	-	-	-	-	-	4	-	-	-	3
U73924	20	-	-	-	-	18	-	-	-	-	12	-	-	2	-
U73925	24	-	-	-	-	24	-	-	-	-	24	-	-	-	-
U73926	24	-	-	-	-	34	-	-	-	-	34	-	-	-	-
U73927	13	-	-	-	-	25	-	-	-	-	26	-	2	3	-
U73929	16	-	-	-	-	18	-	-	-	-	20	-	-	-	-
U73930	18	-	-	-	-	20	-	-	-	-	20	-	-	-	-
U73931	14	-	-	-	-	19	-	-	-	-	16	-	-	-	-
U73932	68	-	-	-	-	37	-	-	-	-	37	-	-	-	3
U73933	18	-	-	-	-	21					21				2
U73934	15	-	-	-	-	17	-	-	-	-	17	-	-	-	-
TOTAL	**310**	**0**	**14**	**0**	**0**	**336**	**0**	**0**	**0**	**0**	**341**	**12**	**7**	**16**	**8**

V. CONCLUSION

The identification of the perfect and imperfect STRs in the DNA sequences have been done using proposed method and compared the detection performance with the reported method TRF and parametric spectral method. From the performance comparison, it is found that the proposed method identifies more copies of the repeated patterns of the STRs than reported methods. Therefore, it is concluded that the proposed method is more sensitive than existing methods for the STRs detection in the DNA sequences. In future a method can be develop to reduced the computational complexity and to detect the tandem repeats of the higher periods present in the DNA sequences.

REFERENCES

[1] Hani Z. Girgis and Sergey L. Sheetlin "Ms-Detector: toward a standard computational tool for DNA microsatellites detection," *Nucleic Acids Research*, 1–13, 2012**.**

[2] Atul Grover, Veenu Aishwarya and P.C. Sharma, "Searching microsatellites in DNA sequences: approaches used and tools developed," *Physiol. Mol Biol. Plants*, 18(1):11–19, 2012.

[3] Elke Schaper, Andrey V. Kajava, Alain Hauser and Maria Anisimova, Repeat or not repeat-statistical validation of tandem repeat prediction in genomic sequences, *Nucleic Acids Research*, 1–13, 2012.

[4] G. Benson, "Tandem repeats finder: A program to analyze DNA sequences," *Nucleic Acids Res.* vol. 27, no. 2, pp. 573–580, 1999.

[5] M. Buchner and S.Janjarasjitt, "Detection and visualization of tandem repeats in DNA sequences," *IEEE Trans. Signal Process*, vol. 51, no. 9, 2003.

[6] Peter G. POP, "Tandem Repeats Localization Using Spectral Techniques" *IEEE Intelligent Computer Communication and Processing*. Romania, 2007

[7] D. Sharma, B. Issac, G.P.S. Raghava, and R. Ramaswamy, "Spectral repeat finder (SRF): Identification of repetitive sequences using Fourier transforms*," Bio* informatics, vol. 20, no. 9, pp. 1405–1412, 2004.

[8] T.T. Tran, V.A. Emanuele and G.T. Zhou, "Techniques for detecting approximate tandem repeats in DNA," *in Proc. IEEE Int. Conf. Acoust.Speech Signal Process. (ICASSP),* vol. 5, May, pp. 449–452, *2004.*
[9] R. Gupta, D. Sarthi, A. Mittal, and K. Singh, "A novel signal processing measure to identify exact and inexact tandem repeat patterns in DNA sequences," *EURASIP J. Bioinf. Syst. Biol.,* vol. 2007, pp. 1–8, 2007.
[10] A.K. Brodzik "Quaternion periodicity transform: an algebraic solution to the tandem repeat detection problem". *Bio* informatics 23:694–700, 2007.
[11] Liping Du, Hongxia Zhou and Hong Yan, "OMWSA: Detection of DNA repeats using moving window spectral analysis," *Bio* informatics, vol. 23, no. 5, pp. 631–633, 2007.pp. 2280–2287, Sep. 2003.
[12] Hongxia Zhou, Liping Du, and Hong Yan, "Detection of Tandem Repeats in DNA Sequences Based on Parametric Spectral Estimation" *IEEE transactions on information technology in biomedicine*, vol. 13, no. 5, September 2009.
[13] R.G. Stockwell, L. Mansinha, R.P. Lowe Localization of the complex spectrum: the S transform, *IEEE Trans Signal Processing*, 44(4), pp: 998- 1001, 1996.
[14] CR Pinnegar, L Mansinha, "The S-transform with windows of arbitrary and varying shape." *Geophysics*. 68, 381–385 2003.
[15] L Mansinha, RG Stockwell, RP Lowe, "Pattern analysis with two-dimensional spectral localization: applications of two-dimensional S-transform." *Phys A*. 239, 286–295 (1997).
[16] Said Assous and Boualem Boashash, "Evaluation of the modified S-transform for time-frequency synchrony analysis and source localization." *EURASIP Journal on Advances in Signal Processing*, 2012:49.
[17] S. D Sharma, D.K Shakya and S. N. Sharma "Evaluation of DNA mapping schemes for exon detection," *IEEE proceeding of ICCET*, p.p.71-74, 2011.
[18] V. Boeva, M. Regnier, D. Papatsenko, and V. Makeev, Short fuzzy tandem repeats in genomic sequences, identification, and possible role in regulation of gene expression, *Bioinformatics*, vol. 22, no. 6, pp. 676–684, 2006.
[19] Micro satellite data sets: http://www.ncbi.nlm.nih.gov/

AUTHORS BIOGRAPHIES

Sunil Datt Sharma was born in Madhya Pradesh (Bhind), M.P., India, in 1980. He received the BE degree in electronics and instrumentation engineering and M.Tech. degree in embedded systems and VLSI in 2005and 2010 respectively from RGTU, Bhopal, MP, India. He is currently a lecturer in the Department of Electronics and Communication Engineering, at Samrat Ashok Technological Institute, Vidisha, MP, India and pursuing his PhD from Jaypee University of Engineering and Technology, Guna, MP, India.

Rajiv Saxena received the BE degree in electronics and telecommunication engineering and the ME degree in digital techniques and data processing. He joined IIT, Roorkee (formerly UOR, Roorkee), India, as a QIP Fellow, for his Doctoral Degree Program, and received the PhD degree in electronics and computer engineering. He is currently Director of JAYPEE University, Anoopsharhr, U.P., India. He has supervised 14 PhD candidates in the area of wireless, cellular, mobile and digital communication, digital signal processing, digital image processing, and application of DSP tools in electronic systems. He has published about 70 research articles in refereed journals of national and international repute.

Sanjeev Narayan Sharma received the BE degree in electronics and instrumentation engineering from Barkatullah University, Bhopal, MP, India, in 1991, the ME degree in measurements and instrumentation engineering from IIT, Roorkee (formerly UOR, Roorkee), India, in 1993, and the PhD degree in electronics and communication engineering from Thapar University, Patiala, India, in 2007. He is currently an associate professor and head of the Department of Electronics and Communication Engineering, Samrat Ashok Technological Institute, Vidisha, MP, India. He has 18 years of teaching experience and 20 publications in reputed journals and refereed conferences. He has supervised 02 PhD candidates in the area of Signal Processsing. His teaching and research interests include digital signal processing, genomic signal processing, fractional Fourier transform, digital filter design, and bio-signal processing.

Ashutosh Kumar Singh has completed his B.Tech in electronics and instrumentation engineering in year July 2001 from M.I.E.T., Meerut. He joined M.I.T.S., Gwalior as M.Tech student in Microwave Engineering in year 2003. After completing his M. Tech. in 2005 Mr. A K Singh has joined JUET, Guna and completed his Ph.D. in 2011. Currently he is working as an Assistant Professor in the Department of Electronics and Communication Engineering, at JUET, Guna, MP, India.