Variant Map System to Simulate Complex Properties of DNA Interactions Using Binary Sequences

Jeffrey Zheng, Weiqiong Zhang, Jin Luo, Wei Zhou and Ruoyu Shen

Abstract Stream cipher, DNA cryptography and DNA analysis are the most important R&D fields in both Cryptography and Bioinformatics. HC-256 is an emerged scheme as the new generation of stream ciphers for advanced network security. From a random sequencing viewpoint, both sequences of HC-256 and real DNA data may have intrinsic pseudo-random properties respectively. In a recent decade, many DNA sequencing projects are developed on cells, plants and animals over the world into huge DNA databases. Researchers notice that mammalian genomes encode thousands of large noncoding RNAs (lncRNAs), interact with chromatin regulatory complexes, and are thought to play a role in localizing these complexes to target loci across the genome. It is a challenge target using higher dimensional visualization tools to organize various complex interactive properties as visual maps. The Variant Map System VMS as an emerging scheme is systematically proposed in this chapter to apply multiple maps that uses four Meta symbols as same as DNA or RNA representations. System architecture of key components and core mechanism on the VMS are described. Key modules, equations and their I/O parameters are discussed. Applying the VM System, two sets of real DNA sequences from both sample human (noncoding DNA) and corn (coding DNA) genomes are collected in comparison with pseudo DNA sequences generated by HC-256 to show their intrinsic properties in higher levels of similar relationships among relevant DNA sequences on 2D maps. Sample 2D maps are listed and their characteristics are illustrated under controllable environment. Visual results are briefly analyzed to explore their intrinsic properties on selected genome sequences.

J. Zheng (✉)
Key Laboratory of Yunnan Software Engineering, Yunnan University, Kunming, China
e-mail: conjugatelogic@yahoo.com

W. Zhang
School of Software and Microelectronics, Peking University, Beijing, China

J. Luo
School of Life Sciences, Yunnan University, Kunming, China

W. Zhou · R. Shen
School of Software, Yunnan University, Kunming, China

J. Zheng (ed.), *Variant Construction from Theoretical Foundation to Applications*,
https://doi.org/10.1007/978-981-13-2282-2_23

Keywords Pseudo-random number generator · Stream cipher · HC-256
Binary to DNA · Pseudo DNA sequence · Large noncoding · DNA analysis
2D map · Visual distribution · Variant map system

1 Introduction

Stream ciphers [1, 2] play a key role in modern network security [3, 4] especially in multimedia network environments; its core component—pseudo random number generation mechanism [5–7]—takes the central position in modern cryptography [8, 9]. Associated with advanced development of bioinformatics, advanced DNA sequencing and analyzing techniques [10, 11] have significantly progressed over the past decade.

1.1 DNA Cryptography

DNA cryptography makes joined research in the field of DNA computing and cryptography. Scholars over the world focused on this field and different results are published such as simulating DNA evolution [12], DNA pseudorandom number generator [13–16], DNA cryptography [9, 17, 18] and so on. However in current situation, DNA cryptography is still at an earlier stage as an emerging area of advanced cryptography.

In typical results of DNA cryptography on encryption, different coding schemes could be randomly selected. E.g. the algorithm in paper [17] applies an encoding formula to express the plaintext on DNA sequence: $\{00 \rightarrow C, 01 \rightarrow T, 10 \rightarrow A, 11 \rightarrow \underline{G}\}$; however in paper [18], the same author uses the coding formula $\{00 \rightarrow A$, $01 \rightarrow T$, $10 \rightarrow C$, $11 \rightarrow G\}$ for the plaintext on DNA sequence. In encryption environment, all $4! = 24$ possible encoding methods could be equally used in different applications.

1.2 Stream Cipher HC-256

Stream ciphers are an important class of encryption algorithms. A stream cipher is a symmetric cipher which operates with a time-varying transformation on individual plaintext digits. The ECRYPT Stream Cipher Project (eSTREAM) [1] was a multi-year effort, running from 2004 to 2008, to promote the design of efficient and compact stream ciphers suitable for widespread adoption. **HC-256** is a stream cipher designed to provide bulk encryption in software at high speeds while permitting strong confidence in its security. A 128-bit variant was submitted in 2004 as an eSTREAM cipher candidate; it has been selected as one of the four final contestants in the software profile [2, 4] in 2008 as the most advanced scheme for stream cipher applications in advanced network environment.

1.3 Large Noncoding DNA and RNA

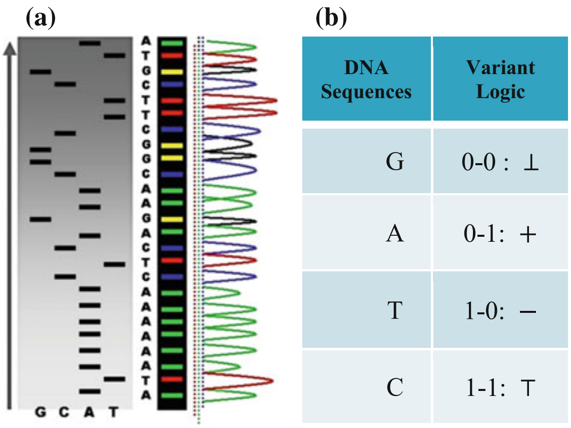
In relation to DNA analysis, visualization methods play a key role in the Human Genome Project (HGP) [19]. After HGP completed successfully, a public research consortium—the Encyclopedia of DNA Elements (ENCODE) were launched by the National Human Genome Research Institute (NHGRI) in 2003 to find all functional elements in the human genome as one of the most critical projects by NHGRI to explore genomes after HGP.

In 2012, ENCODE released a coordinated set of 30 papers published in key Journals of Nature, Genome Biology and Genome Research. These publications show that approximately 20% of noncoding DNA in the human genome is functional while an additional 60% is transcribed with no known function [20]. Much of this functional non-coding DNA is involved in the regulation of the expression of coding genes [10]. Furthermore the expression of each coding gene is controlled by multiple regulatory sites located both near and distant from the gene. These results demonstrate that gene regulation is far more complex than was previously believed [11]. Mammalian genomes encode thousands of large noncoding RNAs (lncRNAs), many of which regulate gene expression, interact with chromatin regulatory complexes, and are thought to play a role in localizing these complexes to target loci across the genome [21]. Associated with different international projects, larger numbers of Genome Databases are established and mass Genome-wide gene expression measurements are developed.

Due to huge amount of DNA sample collections and extremely difficulties to determine their variation properties in wider applications [19, 22–27], it is essential for us to extend advanced DNA analysis models, methods and tools in further extensions to explore emerging models and concepts to interpret complex interactions among complicated sets of DNA sequences in real environments.

1.4 DNA Analysis

DNA analysis plays a key role in modern genomic application [19]. The HGP is heavily relevant to advanced DNA sequencing and analysis techniques. DNA sequences are composed of four Meta symbols on $\{A, T, G, C\}$ as basic structure. Classical DNA double helix structure makes the first level of pair construction of DNA sequences with A & T and G & C complementary structures as the first level of symmetric relationships. A typical DNA sequencing result is shown in Fig. 1a. Four Meta symbols could be separated as four projective sequences.



DNA Sequences	Variant Logic
G	0-0 : ⊥
A	0-1: +
T	1-0: −
C	1-1: ⊤

Fig. 1 Modern DNA sequencing and their correspondences on Variant Logic; **a** a sample DNA sequencing and its four projection sequences; **b** four Meta DNA symbols and linkages to variant logic

In ENCODE, recent Genomic analysis results are indicated that encoded sequences have only 20% in human genomes and around 80% genomes look like useless sequences. Under further assumptions, it seems that additional symmetric properties are required to satisfy the second, third and higher levels of structural constructions to explore complex interactive properties [10, 11, 19–29].

In current situation, it is necessary for advanced researchers to shift targets in computational cell biology from directly collecting sequential data to making higher-level interpretation and exploring efficient content-based retrieval mechanism for genomes. Using higher dimensional visualization tools, their complex interactive properties could be organized as different visual maps systematically.

1.5 Variant Construction and DNA

Variant construction is a new structure composed of logic, measurement and visualization models to analyze 0–1 sequences under variant conditions. The further details of this construction can be checked on variant logic [30, 31], 2D maps [32, 33], variant pseudo-random number generator [34], DNA maps [35] and variant phase spaces [33]. Since the variant system uses another set of four Meta symbols {⊥, +, −, ⊤} to describe system, a typical correspondence shown in Fig. 1b may provides a natural mapping between DNA and variant data sequences.

Since DNA sequences are played an essential role to explore different symmetric properties based on analysis approaches, in this chapter, measurement and visual models are proposed systematically to use a fixed segment structure to measure four Meta symbols distributions in their spectrum construction. Under this construction, refined symmetric features can be identified from various polarized distributions and further symmetric properties are visualized.

1.6 Target of This Chapter

The target of this chapter is to establish the Variant Map System (VMS) as a unified framework to analyze complex DNA interactions on both artificial and natural DNA sequences. The VMS has designed to use variant logic schemes [30–35] applying multiple maps on four Meta symbols as DNA or RNA representations. System architecture of key components and core mechanism on the VMS are described. Key modules, equations and their I/O parameters are discussed. Applying the VM System, two sets of real DNA sequences from both human (noncoding DNA) and corn (coding DNA) genomes are collected in comparison with pseudo DNA sequences generated artificially by HC-256 to show their intrinsic properties in higher levels of similar relationships among DNA sequences on 2D maps. Further descriptions and discussions are provided respectively.

2 System Architecture

In this section, system architecture and their core components are discussed with the use of diagrams. The refined definitions and equations of this system are described in the next section—Variant Map System.

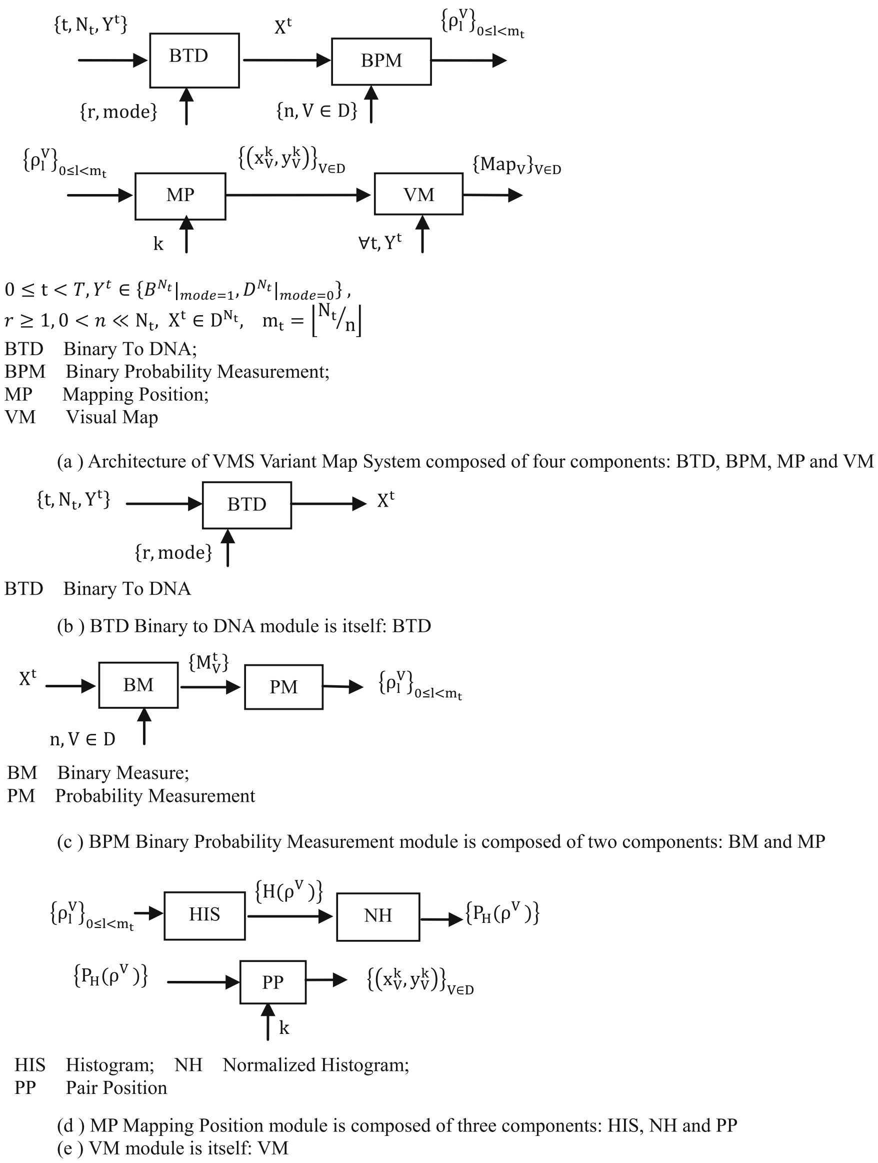
2.1 Architecture

The four components of a variant map system are the Binary To DNA (BTD), the Binary Probability Measurement (BPM), the Mapping Position (MP), and the Visual Map (VM) as shown in Fig. 2.

The architecture is shown in Fig. 2a with the key modules of the four core components being shown in Fig. 2b–e respectively.

In the first part of the system, the t-th sequence Y^t on either $\{0, 1\}$ or $\{A, G, T, C\}$ are input data to get into the BTD module. The main function of the BTM is to output a unified sequence X^t either to transfer a 0–1 sequence or to keep a DNA sequence as a pseudo or pure DNA sequence under a set of controlled parameters.

Using this unified DNA sequence, four vectors of probability measurements are created from the t-th selected DNA sequence with N_t elements as an input. Multiple segments are partitioned by a fixed number of n elements for each segment; at least m_t segments can be identified by the BPM component. Next component uses the four vectors of probability measurements and a given k value as input data, a pair of position values are created for each Meta symbol. Four pairs of values are generated by the MP component. Then, in order to process multiple selected DNA sequences, all selected sequences are processed by the VM component and each sequence may pro-



$0 \le t < T, Y^t \in \{B^{N_t}|_{mode=1}, D^{N_t}|_{mode=0}\}$,
$r \ge 1, 0 < n \ll N_t,\ X^t \in D^{N_t},\quad m_t = \lfloor N_t/n \rfloor$

BTD Binary To DNA;
BPM Binary Probability Measurement;
MP Mapping Position;
VM Visual Map

(a) Architecture of VMS Variant Map System composed of four components: BTD, BPM, MP and VM

BTD Binary To DNA

(b) BTD Binary to DNA module is itself: BTD

BM Binary Measure;
PM Probability Measurement

(c) BPM Binary Probability Measurement module is composed of two components: BM and MP

HIS Histogram; NH Normalized Histogram;
PP Pair Position

(d) MP Mapping Position module is composed of three components: HIS, NH and PP
(e) VM module is itself: VM

Fig. 2 Variant Map System VMS and key components **a** Architecture; **a** BTD component; **b** BPM component; **c** MP component; **d** VM component

vide a set of pair values to generate relevant variant maps to indicate their distribution properties respectively.

With eight parameters in an input group, there are three sets of parameters in the intermediate group and one set of parameters in the output group.

The three groups of parameters are listed as follows.

Input Group:

t An integer indicates the t-th DNA sequence selected, $0 \leq t < T$

r An integer indicates a relationship distance among elements in a binary sequence, $r \geq 1$

mode An integer indicates the mode of elements in a sequence, mode $\in \{0, 1, \ldots\}$, mode = 0 for a DNA sequence, mode = 1 for a binary sequence

N_t An integer indicates the number of elements in the t-th DNA sequence, $N_t \gg r$

Y^t An input data vector with N_t elements, $Y^t \in \left\{D^{N_t}|_{\text{mod } e=0}, B^{N_t}|_{\text{mod } e=1}\right\}$

n An integer indicates the number of elements in a segment, $n > 0$

V A symbol is selected from four DNA symbols $\{A, G, T, C\} = D$, $V \in D$

k An integer indicates the control parameter for mapping, $k > 0$.

Intermediate Group:

X^t A unified DNA vector with N_t elements, $X^t \in D^{N_t}$

$\left\{\rho_l^V\right\}$ Four sets of probability measurements with $0 \leq l < m_t$, $V \in D$

$\left\{\left(x_V^k, y_V^k\right)\right\}$ Four paired values, $k > 0$, $V \in D$

Output Group:

$\left\{\text{Map}_V\right\}$ Four 2D maps, $V \in D$

2.2 BTD Binary to DNA

The BTD component shown in Fig. 2b is composed of one module: BTD itself. Five parameters are shown as input signals and one unified vector is generated by the BTD component as the output group.

Input Group:

t An integer indicates the t-th DNA sequence selected, $0 \leq t < T$

r An integer indicates a relationship distance among elements in a binary sequence, $r \geq 1$

mode An integer indicates the mode of elements in a sequence, mode $\in \{0, 1, \ldots\}$, mode = 0 for a DNA sequence, mode = 1 for a binary sequence

N_t An integer indicates the number of elements in the t-th DNA sequence, $N_t \gg r$

Y^t An input data vector with N_t elements, $Y^t \in \left\{D^{N_t}|_{\text{mod } e=0}, B^{N_t}|_{\text{mod } e=1}\right\}$

Output Group:

X^t A unified data vector with N_t elements, $X^t \in D^{N_t}$

The BTD component uses an input vector on either binary or DNA format as input, under a set of input parameters to process transformation. The output of the BTD component is composed of a unified vector of DNA format in a given condition.

2.3 BPM Binary Probability Measurement

The BPM component shown in Fig. 2c is composed of two modules: BM Binary Measure and PM Probability Measurement. Three parameters are listed as input signals; four vectors of binary measures are outputted from the BM component as an intermediate group and four sets of probability measurements are outputted as an output group.

Input Group:

n An integer indicates the number of elements in a segment, $n > 0$
V A symbol is selected from four DNA symbols $\{A, G, T, C\} = D$, $V \in D$
X^t A DNA vector with N_t elements, $X^t \in D^{N_t}$

Intermediate Group:

$\left\{M_V^t\right\}$ Four 0–1 vectors with N_t elements, $M_V^t(I) \in \{0, 1\} = B$, $M_V^t \in B^{N_t}$, $V \in D$

Output Group:

$\left\{\rho_l^V\right\}$ Four sets of probability measurements with $0 \le l < m_t$, $V \in D$

The BPM component transforms a selected DNA sequence to generate four 0–1 vectors by BM module for the input DNA sequence. Then four probability vectors are generated by the PM module as the output of the BPM under a fixed length of segment condition.

2.4 MP Mapping Position

The MP component shown in Fig. 2d is composed of three modules: HIS Histogram, NH Normalized Histogram and PP Pair Position. Two parameters are listed as input signals; four histograms and four normalized histograms are generated from the HIS component and the NH component as intermediate groups respectively. Four paired values are generated by the PP component as the output group.

Input Group:

$\left\{\rho_l^V\right\}$ Four sets of probability measurements with $0 \le l < m_t$, $V \in D$
k An integer indicates the control parameter for mapping, $k > 0$

Intermediate Group:

$\{H(\rho^V)\}$ Four histograms for relevant probability measurements, $V \in D$

$\{P_H(\rho^V)\}$ Four normalized histograms for relevant probability measurements, $V \in D$

Output Group:

$\{(x_V^k, y_V^k)\}$ Four paired values, $k > 0$, $V \in D$

The MP component uses probability measurements as input, under a given k condition to generate each relevant histogram and its normalized distribution. The output of the MP component is composed of four paired values controlled in a given condition.

2.5 VM Visual Map

The VM component shown in Fig. 2e is composed of one module: VM Visual Map. Three parameters are input signals. Collected all selected DNA sequences, four 2D maps are generated by the VM component as the output result.

Input Group:

$\forall t$ All DNA sequences are selected, $0 \le t < T$

Y^t An input data vector with N_t elements, $Y^t \in \{D^{N_t}|_{\bmod e=0}, B^{N_t}|_{\bmod e=1}\}$

$\{(x_V^k, y_V^k)\}^t$ Four paired values for the t-th DNA sequence, $k > 0$, $V \in D$

Output Group:

$\{\mathrm{Map}_V\}$ Four 2D maps, $V \in D$

The VM component processes all selected DNA sequences as input to generate paired values for each sequence. The output of the VM component is composed of four 2D maps to show the final visual distribution for the system.

3 Variant Map System

In this section, definitions and equations are provided to describe the VMS. In addition to the initial preparation, seven core modules are involved in the BTD, BM, PM, HIS, NH, PP and VM components respectively.

3.1 Initial Preparation

Let r an input parameter make all pairs of elements with r distance in a binary sequence to be a pseudo DNA vector, mode a controlled parameter indicate various pairs of operations performed if mode $\geq$ 1. Denote $B = \{0, 1\}$ a binary base and $D = \{A, G, T, C\}$ a DNA base respectively.

3.2 BTD Module

Let Y an input sequence with N elements,$0 \leq I < N$, $Y(I) \in \{B^N|_{\text{ mod } e \geq 1}, Y(I) \in D^N|_{\text{ mod } e=0}\}$. This input vector could be expressed as follows.

$$\begin{aligned} &Y = (Y(0), \ldots, Y(I), \ldots, Y(N-1)),\ 0 \leq I < N \\ &Y(I) \in \{B^N|_{\text{mode} \geq 1},\ Y(I) \in D^N|_{\text{mode}=0}\}. \end{aligned} \tag{1}$$

Let X denote a DNA sequence with N elements, D denote a symbol set with four elements i.e. $D = \{A, G, T, C\}$. This type of a DNA sequence can be described by a four valued vector as follows:

$$\begin{aligned} &X = (X(0), \ldots, X(I), \ldots, X(N-1)), \\ &0 \leq I < N, X(I) \in D = \{A, G, T, C\}, X \in D^N \end{aligned} \tag{2}$$

From this input and associated parameters, following operations are performed. If mode = 0, for all I, $Y(I) \in D$, the output vector is equal to the input vector.

$$\forall I, X(I) = Y(I), 0 \leq I < N \tag{3}$$

If mode = 1, for all pairs of I and $I + r(\text{mod}N)$ elements of Y, $Y(I)$, $Y(I + r) \in B$, the I-th output element $X(I)$ can be determined by the corresponding conditions shown in Fig. 1b as follows.

$$\text{X(I)} = \begin{cases} \text{G}, & \text{if Y(I)} = 0\ \&\ Y(\text{I} + \text{r}) = 0 \\ \text{A}, & \text{if Y(I)} = 0\ \&\ Y(\text{I} + \text{r}) = 1 \\ \text{T}, & \text{if Y(I)} = 1\ \&\ Y(\text{I} + \text{r}) = 0 \\ \text{C}, & \text{if Y(I)} = 1\ \&\ Y(\text{I} + \text{r}) = 1 \end{cases}, \tag{4}$$

In both conditions, X will be a unified vector with four values as the output of the BTD shown in Fig. 2b.

E.g. Let a binary sequence $Y = 100111001011$, $N = 12$, three pseudo DNA sequences ($r = 1, r = 2, r = 3$) can be represented as follows.

$$Y = 100111001011$$

$$\begin{aligned} X_{r=1} &= TGACCTGATACC \\ X_{r=2} &= TAACTTAGCACT \\ X_{r=3} &= CAATTCGACATT \\ Y \in B^{12}&, X \in D^{12} \end{aligned}$$

Selecting a certain r value, a relevant pseudo DNA sequence can be generated from an input binary sequence.

3.3 BM Module

For a given I-th element, four projective operators can be defined and denoted as $\{M_A(I), M_G(I), M_T(I), M_C(I)\}$.

$$M_A(I) = \begin{cases} 1, \text{ if } X(I) = A; \\ 0, \text{ Otherwise;} \end{cases} M_G(I) = \begin{cases} 1, \text{ if } X(I) = G; \\ 0, \text{ Otherwise;} \end{cases} M_T(I) \\ = \begin{cases} 1, \text{ if } X(I) = T; \\ 0, \text{ Otherwise;} \end{cases} M_C(I) = \begin{cases} 1, \text{ if } X(I) = C; \\ 0, \text{ Otherwise;} \end{cases} \quad (5)$$

Applying the four operators to all elements, the DNA sequence X can be reorganized into the four binary sequences of 0–1 values. i.e.

$$\begin{aligned} &M_V : \{X(I)\}_{I=0}^{N-1} \rightarrow \{M_A(I), M_G(I), M_T(I), M_C(I),\}_{I=0}^{N-1}; \\ &M_V(I) \in B = \{0, 1\}, V \in D \end{aligned} \quad (6)$$

E.g. Let a DNA sequence $X = CTGATTAGCCAT$, $N = 12$, its four binary sequences can be represented as follows.

$$\begin{aligned} X &= CTGATTAGCCAT \\ M_A &= 000100100010 \\ M_G &= 001000010000 \\ M_T &= 010011000001 \\ M_C &= 100000001100 \end{aligned}$$

It is interesting to notice that the basic relationship between a DNA sequence X and its four M_V sequences are exactly same as in a modern DNA sequencing procedure to separate a selected DNA sequence into the four Meta symbol sequences shown in Fig. 1a. This correspondence could be the key feature to apply the proposed scheme naturally in simulating complex behaviors for any DNA sequence.

The projection M_V provides the essential operation in the BM component as the first module shown in Fig. 2c.

3.4 PM Module

For this set of the four binary sequences, it is convenient to partition them into m segments and each segment contained a fixed number of n elements.

For the l-th segment, let $0 \leq 1 < m, 0 \leq j < n$, the I-th position will be $I = l * n + j$, four probability measurements $\{\rho_A, \rho_G, \rho_T, \rho_C,\}$ can be defined.

$$\rho_l^V = \frac{\sum_{I=l*n}^{(l+1)*n-1} M_V(I)}{n}, V \in D, 0 \leq I < N = n * m \tag{7}$$

Under this construction, four sets of probability measurements established.

$$\rho^V : \{M_A(I), M_G(I), M_T(I), M_C(I),\}_{I=0}^{N-1} \rightarrow \quad \left\{\rho_l^A, \rho_l^G, \rho_l^T, \rho_l^C,\right\}_{l=0}^{m-1} \tag{8}$$

The probability operator ρ^V generates four probability measurement vectors in the PM component as the second module shown in Fig. 2c. After the BM and PM processes, the whole procedure of the BPM component is complete in Fig. 2c.

3.5 HIS Module

Since the BPM generates four sets of probability measurement, it is necessary to perform further operations in the MP component shown in Fig. 2d as follows.

In the HIS component as the first module in Fig. 2d, each probability sequence $\left\{\rho_l^V,\right\}_{l=0}^{m-1}, V \in D$ can be calculated from n positions, at most $n+1$ distinguished values identified in a vector. Under this organization, a histogram distribution can be established.

Let $H(.)$ be a histogram operator, for each position, it satisfies following relation,

$$H\left(\rho_l^V\right) = \begin{cases} 1, \text{ if } \rho_l^V = \frac{i}{n}, & V \in D; \\ 0, \text{ Otherwise}, & 0 \leq i \leq n. \end{cases} \tag{9}$$

Collecting all possible values, a histogram distribution can be established,

$$H\left(\rho^V\right) = \sum_{l=0}^{m-1} H\left(\rho_l^V\right) \tag{10}$$

The histogram $H\left(\rho^V\right)$ is the output of the HIS module. Four histograms are generated after HIS process. Further normalized process will be performed in the NH component as the second module in Fig. 2d.

3.6 NH Module

Under this construction, a normalized histogram can be defined as

$$P_H\left(\rho^V\right) = H\left(\rho^V\right)/m \tag{11}$$

After the NH component processed, its output provides the *PP* component for further operations as the third module in Fig. 2d.

3.7 PP Module

Relevant probability vectors have $(n+1)$ distinguished values; four sets of normalized vectors can be organized as a linear order as follows,

$$p_i^V = \sum_{l=0}^{m-1} H\left(\rho_l^V | \rho_l^V = \frac{i}{n}\right)/m,\ 0 \le i \le n \tag{12}$$

Under this condition, four linear sets of probability vectors are established,

$$\begin{aligned} &P_H\left(\rho^V\right) = \left\{p_i^A, p_i^G, p_i^T, p_i^C,\right\}_{i=0}^n, \\ &p_i^V \in [0, 1],\ V \in D, 0 \le i \le n \end{aligned} \tag{13}$$

For four vectors, their components can be normalized respectively,

$$\sum_{i=0}^{n} p_i^V = 1,\ V \in D \tag{14}$$

Four sets of probability vectors are composed of a complete partition on their measurements.

Using this set of measurements, two mapping functions can be established to calculate a pair of values to map analyzed DNA sequence into a 2D map as follows.

Let $y = F(P, V, k)$ and $x = F(P, V, 1/k)$ or $\left(x_V^k, y_V^k\right)$ be a pair of values defined by following equations,

$$\begin{aligned} &y_V^k = F(P, V, k) = \left(\sum_{i=0}^{n} \sqrt[k]{p_i^V}\right)^k \ \& \\ &x_V^k = F(P, V, 1/k) = \sqrt[k]{\sum_{i=0}^{n} \left(p_i^V\right)^k},\ V \in D \end{aligned} \tag{15}$$

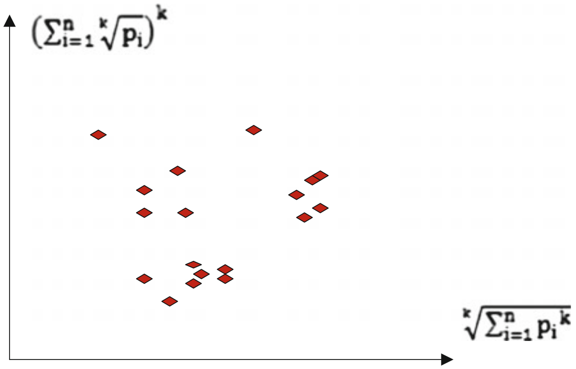


Fig. 3 A sample 2D map of VM on multiple sequences

In the *PP* component, four paired values are generated and each pair indicates a specific position on a 2D map for the selected DNA sequence. The core operations of three key components: BTD, BPM and MP for a selected sequence are performed in Fig. 2b–d.

3.8 VM Module

Since only one point of a 2D map is determined for a selected DNA sequence, it is essential to apply relative larger number of DNA sequences as inputs to generate visible distributions. This type of operations will be performed in the VM component shown in Fig. 2e.

In a general condition, the VM component processes a selected data set $\{Y^t\}_{t=0}^{T-1}$ composed of T sequences, the t-th sequence with N_t elements can be expressed by $Y^t = (Y^t(0), \ldots, Y^t(I), \ldots, Y^t(N_t - 1))$, $Y^t \in Y(I) \in \{B^{N_t}|_{\text{mode} \geq 1}, Y(I) \in D^{N_t}|_{\text{mode}=0}\}$. Each sequence can be processed to apply the same procedures of the BTD, BPM and MP components. Since for each segment, its length n will be fixed for all selected sequences, it is essential to make number of segments be $m^t = \lfloor N_t/n \rfloor$ in convention to match each sequence. Under this expression, the last module VM collects all T pairs of positions on relevant 2D visual maps as follows,

$$\text{VM} : \{X^t\}_{t=0}^{T-1} \to \left\{\left(x_V^k, y_V^k\right)^t\right\}_{t=0}^{T-1} \to \{\text{MAP}_V\}, V \in D \tag{16}$$

A sample 2D map of VM is shown in Fig. 3; this provides an assistant illustration for this type of visual maps on a case of multiple sequences.

Under this construction, a total number of T DNA sequences are transformed as T visual points on four 2D visual maps that would be help analyzers to explore their intrinsic symmetry properties among four binary sequences.

4 Sample Results on 2D Maps

Two types of data sets are selected for comparison. The first type of data sets are real DNA data sequences collected from both human and plan genomes to illustrate their differences on 2D maps. The second type of data set is collected from the Stream Cipher HC-256 to generate a pseudo random binary sequence under a certain condition.

4.1 DNA Data Resources

It is important to use some real DNA sequences to illustrate various test results of the VMS. Two sets of DNA sequences are selected and relevant resource features are described as follows.

The first data set originally comes from the human genome assembly version 37 and was taken from the reference sequences of 13 anonymous volunteers from Buffalo, New York. Hi-C technology [5] used to analyze chromatin interaction role in genome. From a genomic analysis viewpoint, this set of data may contain more complex secondary or higher level structures. A special structure nearly the GRCh37 DNA sequence has been identified to explore their spatial characteristics. After positive and negative sequencing, each data file contain 2700 DNA sequences and each sequence has around 500 elements stored in two files *left* and *right* respectively.

The second DNA data set are selected from some plant gene database for comparison. One set of DNA sequences of Corn genomes are stored in file 201–500 that contains 2700 DNA sequences and each sequence has around 200–600 elements. It may be ordinary single sequences without complex secondary structures.

4.2 Pseudo DNA Data Resources

The Stream Cipher HC-256 has being used to generate a binary sequence on a total length of 2700 × 500 bits in the file *hc256* that has been partitioned as 2700 subsequences and each sub-sequence in 500 bits.

Using the VMS in various parameters, three sets of pseudo DNA sequences are generated and their 2D maps are illustrated, analyzed and compared in following subsections.

4.3 Sample Results

Using the three files of DNA sequences and one pseudo binary sequence in three parameters, six sets of 2D maps are listed in Figs. 4, 5, 6, 7, 8 and 9 under different conditions to illustrate their spatial distributions using the VMS in a controllable environment.

In Fig. 4, three groups of eighteen 2D maps are shown in the range of $n = 3 \sim 50$, $k = 7$, $N \cong 200 \sim 600$, $T = 2700$ for comparison; (a1–a6) six Map_A maps for the file *Right*; (b1–b6) six Map_G maps for the file 201–500; (c1–c6) six Map_A maps for the file *hc256* respectively.

In Fig. 5, four groups of sixteen 2D maps for the file *right* are listed in the range of $n = 15, k = \{2, 3, 4, 7\}, N \cong 500, T = 2700$; (a) group (a1–a4) four Map_A maps; (b) group (b1–b4) four Map_T maps; (c) group (c1–c4) four Map_G maps; (d) group (d1–d4) four Map_C maps.

In Fig. 6, four groups of sixteen 2D maps for the file *hc256* are listed in the range of $n = 12, k = \{2, 3, 4, 7\}, N \cong 500, T = 2700, r = 1, \mathrm{mod}e = 1$; (a) group (a1–a4) four Map_A maps; (b) group (b1–b4) four Map_T maps; (c) group (c1–c4) four Map_G maps; (d) group (d1–d4) four Map_C maps.

In Fig. 7, four groups of sixteen 2D maps for the file *right* are selected in the range of$n = 15, k = \{2, 3, 4, 7\}, N \cong 500, T = 2700$; (a) group (a1–a4) four Map_A maps; (b) group (b1–b4) four Map_T maps; (c) group (c1–c4) four Map_G maps; (d) group (d1–d4) four Map_C maps.

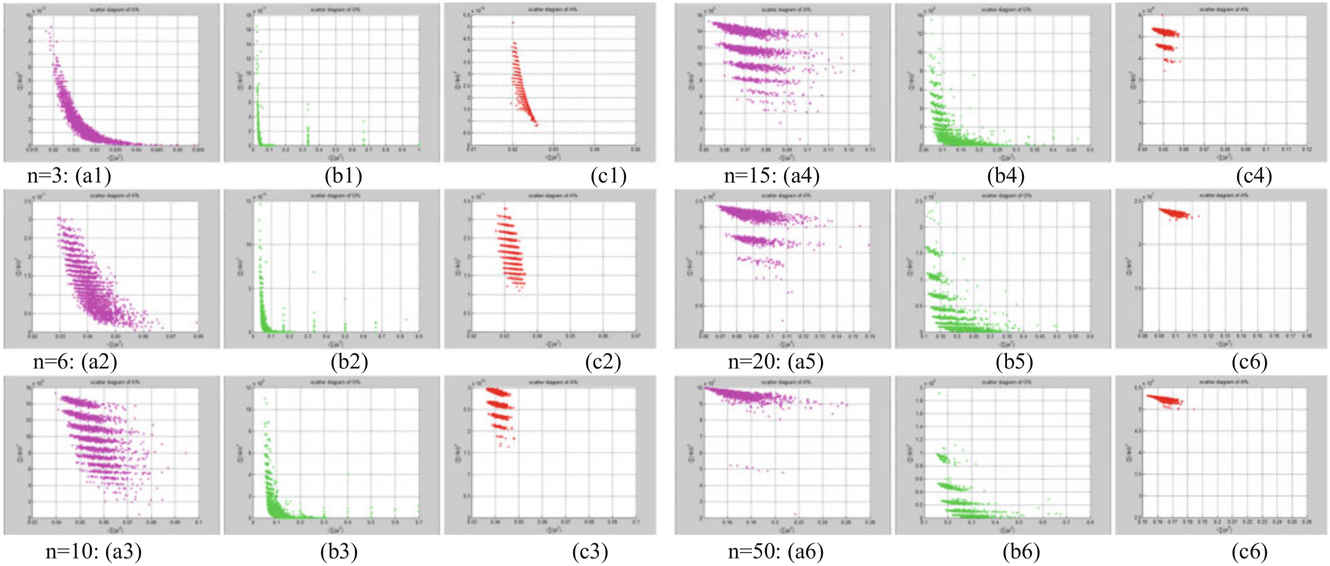
In Fig. 8, three groups of twelve 2D maps for the file *hc256* are compared in the range of $n = 12, k = 7, N \cong 500, T = 2700, r = \{1, 2, 3\}, \mathrm{mod}e = 1$; (a) group (a1–a4) four Map_V maps $r = 1$; (b) group (b1–b4) four Map_V maps $r = 2$; (c) group (c1–c4) four Map_V maps $r = 3$.

In Fig. 9, three groups of twelve 2D maps for two files *right* and *hc256* are compared in the range of $k = 7, N \cong 500, T = 2700$; (a) the file *right* $n = 15$, mode = 0; (b) the file *hc256* $n = 12$, mode = 1, $r = 1$; (c) the file *hc256* $n = 12$, mode = 1, $r = 3$; (a1–c1) Map_A maps; (a2–c2) Map_T maps; (a3–c3) Map_G maps; (a4–c4) Map_C maps.

4.4 Result Analysis of 2D Maps

Six groups of 2D maps contain different information, it is necessary to make a brief discussion on their important issues as follows.

The first group of results shown in Fig. 4 presents three sets of eighteen 2D maps from three data files: *right,* 201–500 and *hc256* undertaken various lengths of basic segment from 3 to 50 to illustrate their variations respectively. Six 2D maps of each group in Fig. 4 (a1–a6) show significant trace on their visual distributions; the numbers of main visible clusters identified are decreased when the length of segment has being increased e.g. (a3–a6). However lesser length of segment does not pro-

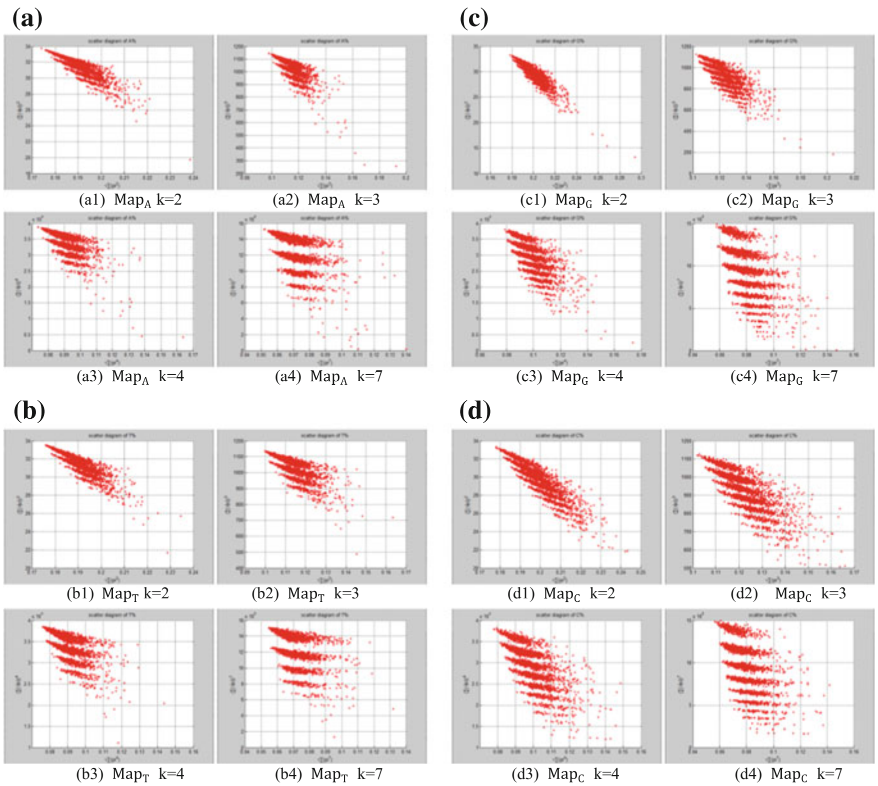


n=3: (a1) (b1) (c1) n=15: (a4) (b4) (c4)

n=6: (a2) (b2) (c2) n=20: (a5) (b5) (c6)

n=10: (a3) (b3) (c3) n=50: (a6) (b6) (c6)

Fig. 4 Three groups of eighteen 2D maps in the range of *n=3~50, k=7, N≅200~600, T=2700*; (a1–a6) Map$_A$ for the file *Right*; (b1–b6) Map$_G$ for the file 201–500; (c1–c6) Map$_A$ for the file *hc256* mode $= 1, r = 1$



(a1) Map$_A$ k=2 (a2) Map$_A$ k=3 (c1) Map$_G$ k=2 (c2) Map$_G$ k=3

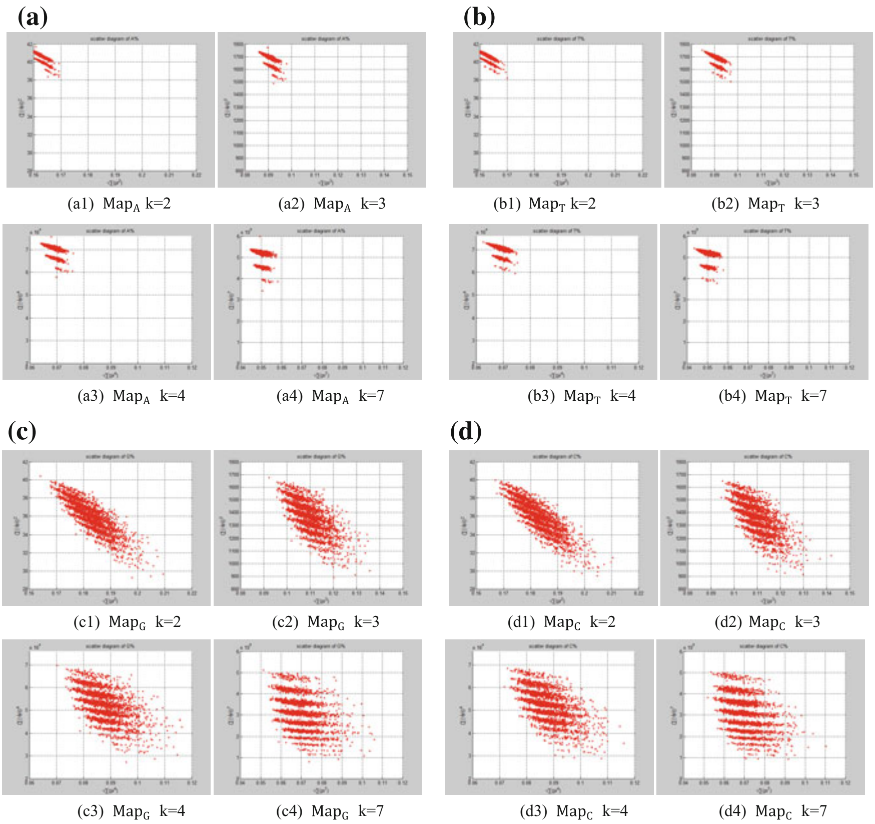
(a3) Map$_A$ k=4 (a4) Map$_A$ k=7 (c3) Map$_G$ k=4 (c4) Map$_G$ k=7

(b1) Map$_T$ k=2 (b2) Map$_T$ k=3 (d1) Map$_C$ k=2 (d2) Map$_C$ k=3

(b3) Map$_T$ k=4 (b4) Map$_T$ k=7 (d3) Map$_C$ k=4 (d4) Map$_C$ k=7

Fig. 5 Four groups of sixteen 2D maps in the range of $n = 15, k = \{2, 3, 4, 7\}, N \cong 500, T = 2700$; **a** group (a1–a4) four Map$_A$ maps; **b** group (b1–b4) four Map$_T$ maps; **c** (c1–c4) four Map$_G$ maps; **d** (d1–d4) four Map$_C$ maps for the file *right*

vide refined visual distinctions with larger region in fuzzy areas e.g. (a1–a2). From a structural viewpoint, middle ranged numbers of length provide better clustering results e.g. (a3–a5) for further analysis targets. To check another six 2D maps of Fig. 4 (b1–b6) for the file 201–500, significantly different visual distributions can be observed than (a1–a6); the numbers of main visible clusters identified are decreased when the length of segment has being increased less significantly e.g. (b4–b6). However lesser length of segment does not provide refined visual distinctions with wider regions in fuzzy areas e.g. (b1–b3). In general, middle ranged numbers of length still provide better clustering effects e.g. (b4–b6) for further analysis purpose. To check six 2D maps of Fig. 4 (c1–c6) for the file *hc256 r=1*, similar visual distributions can be observed than (a1–a6) and significantly differences are observed than (b1–b6); the numbers of main visible clusters identified are decreased when the length of segment has being increased less significantly e.g. (c3–c6). However lesser length of segment does provide refined visual distinctions with regions in fuzzy areas e.g. (b1). In general, middle ranged numbers of length still provide better clustering effects



(a1) Map$_A$ k=2 (a2) Map$_A$ k=3 (b1) Map$_T$ k=2 (b2) Map$_T$ k=3

(a3) Map$_A$ k=4 (a4) Map$_A$ k=7 (b3) Map$_T$ k=4 (b4) Map$_T$ k=7

(c1) Map$_G$ k=2 (c2) Map$_G$ k=3 (d1) Map$_C$ k=2 (d2) Map$_C$ k=3

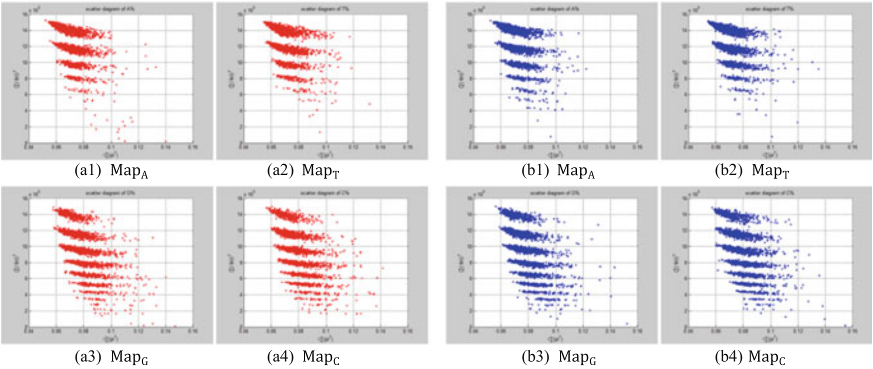
(c3) Map$_G$ k=4 (c4) Map$_G$ k=7 (d3) Map$_C$ k=4 (d4) Map$_C$ k=7

Fig. 6 Four groups of sixteen 2D maps in the range of $n = 12, k = \{2, 3, 4, 7\}, N \cong 500, T = 2700$ for the file *hc256*, $r = 1$, mode = 1; **a** group (a1–a4) four Map$_A$ maps; **b** group (b1–b4) four Map$_T$ maps; **c** (c1–c4) four Map$_G$ maps; **d** (d1–d4) four Map$_C$ maps

e.g. (c2–c4) for further analysis purpose. From their distributions, groups (a) and (c) have shared much stronger similar properties than group (b).

It is interesting to observe different maps when control parameter k changed. Four groups of sixteen 2D maps for the file *right* are shown in Fig. 5 on the range of $n = 15, k = \{2, 3, 4, 7\}, N \cong 500, T = 2700$; four groups in (a)–(d) provide four maps to share the same other parameters with different k values. Checking visible clusters in different maps, it is important to notice nearly same numbers of clusters identified in the same group, but different groups may contain significantly different numbers. Lesser k value (e.g. $k = 2$) makes a tighter distribution and larger k value (e.g. $k = 7$) takes better separation on the maps. Through $k = 7$ maps provide better separation effects, it is easy to observe their y axis values already in 10^8 range.

Four groups of sixteen 2D maps for the file *hc256* are shown in Fig. 6 in the range of $n = 12, k = \{2, 3, 4, 7\}, N \cong 500, T = 2700, r = 1$. This group of 2D maps

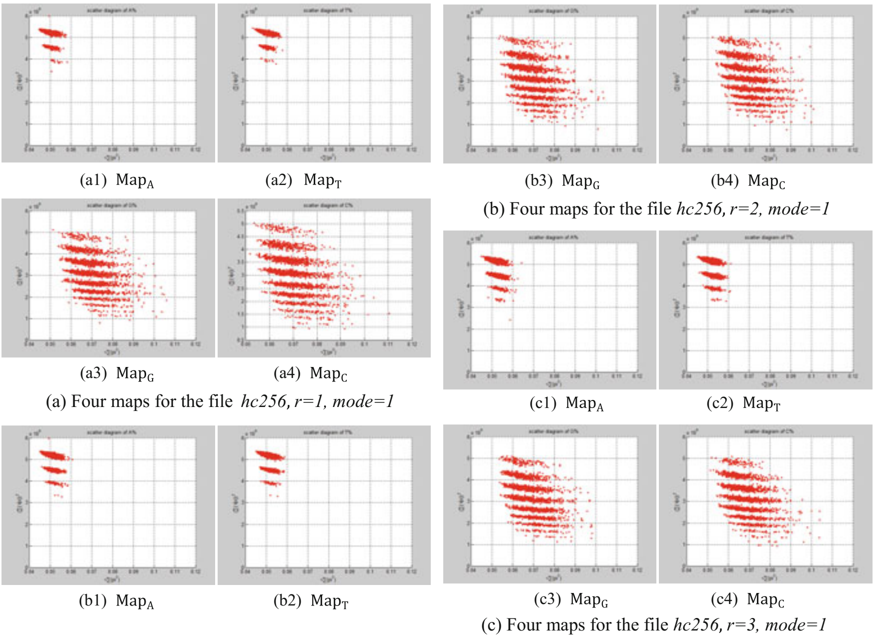


(a1) Map_A (a2) Map_T (b1) Map_A (b2) Map_T

(a3) Map_G (a4) Map_C (b3) Map_G (b4) Map_C

(a) Four maps for the file *left* (b) Four maps for the file *right*

Fig. 7 Two groups of eight 2D maps in the range of $n = 15, k = 7, N \cong 200 \sim 600, T = 2700$; **a** group (a1–a4) four Map_V maps for the file *left*; (b) group (b1–b4) four Map_V maps for the file *right*



(a1) Map_A (a2) Map_T (b3) Map_G (b4) Map_C

(b) Four maps for the file *hc256, r=2, mode=1*

(a3) Map_G (a4) Map_C

(a) Four maps for the file *hc256, r=1, mode=1*

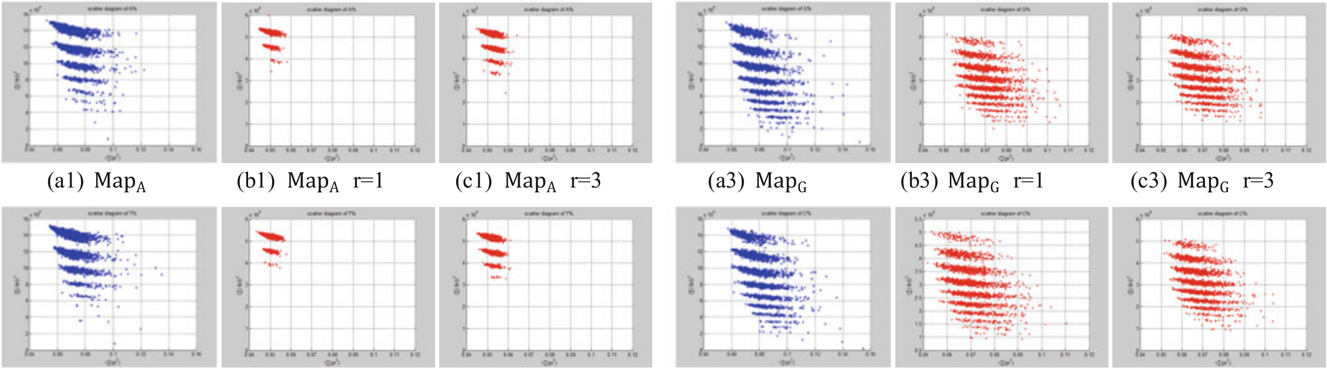
(c1) Map_A (c2) Map_T

(b1) Map_A (b2) Map_T (c3) Map_G (c4) Map_C

(c) Four maps for the file *hc256, r=3, mode=1*

Fig. 8 Three groups of twelve 2D maps in the range of $n=12$, $k=7$, $N=500$, $T=2700$ for the file *hc256*, $r=\{1,2,3\}$, $mode=1$; **a** group (a1–a4) four Map_V maps $r=1$; **b** group (b1–b4) four Map_V maps $r=2$; **c** group (c1–c4) four Map_V maps $r=3$

can be compared with 2D maps in Fig. 5. Under the same parameters, similar visible effects and feature clustering properties could be observed if various k values are selected.



(a1) Map_A (b1) Map_A r=1 (c1) Map_A r=3 (a3) Map_G (b3) Map_G r=1 (c3) Map_G r=3

(a2) Map_T (b2) Map_T r=1 (c2) Map_T r=3 (a4) Map_C (b4) Map_C r=1 (c4) Map_C r=3

(a1)-(a4) Four maps for the file *right, n=15, mode=0*
(b1)-(b4) Four maps for the file *hc256, n=12, r=1, mode=1*
(c1)-(c4) Four maps for the file *hc256, n=12, r=3, mode=1*

Fig. 9 Three groups of twelve maps in the ranges: $N=500$, $T=2700$, $k=7$; **a** Real DNA data; (a1–4) DNA sequences from the file *right*; (**b–c**) Simulation data; (b1–4) Binary sequences from the file *hc256*, $r=1$; (c1–4) Binary sequences from the file *hc256*, $r=3$

Using a set of selected parameters, two groups of eight 2D maps are compared in Fig. 7 for two files: *left, right* to explore higher levels of symmetric properties for secondary or higher levels of structures potentially contained in DNA sequences. Selected parameters are in the range of $n = 15, k = 7, N \cong 500, T = 2700$. Group (a) provides four Map$_V$ maps (a1–a4) for the file *left*; group (b) uses four Map$_V$ maps (b1–b4) for the file *right*.

In convenient description, let~be a similar operator, for groups (a) and (b), four pairs of {(a1)~(b1), (a2)~(b2), (a3)~(b3), (a4)~(b4)} maps i.e. (*left*-A~*right*-A, *left*-T~*right*-T, *left*-G~*right*-G, *left*-C~*right*-C) have a stronger similar distribution between *left* & *right*. In addition, only two clustering classes could be significantly identified as {(a1)~(a2)~(b1)~(b2), (a3)~(a4)~(b3)~(b4)} i.e. (*left*-A~*right*-A~*left*-T~*right*-T, *left*-G~*right*-G~*left*-C~*right*-C) respectively. This type of similar clustering distributions may strongly indicate eight maps with intrinsically higher levels of DNA sequences with extra A–T and G–C pairs of symmetric relationships between two files: *left* & *right*.

Using a set of selected parameters, three groups of twelve 2D maps are listed in Fig. 8 for the file *hc256, r={1,2,3}* to explore properties for their higher levels of structures potentially contained in pseudo DNA sequences. Selected parameters are in the range of $n = 12, k = 7, N \cong 500, T = 2700$. Group (a) provides four Map$_V$ maps (a1–a4) for *r=1*; group (b) uses four Map$_V$ maps (b1–b4) for *r=2* (c) uses four Map$_V$ maps (c1–c4) for *r=3*. Using a similar operator, for groups (a–c), four pairs of {(a1)~(b1)~(c1), (a2)~(b2)~(c2), (a3)~(b3)~(c3), (a4)~(b4)~(c4)} maps i.e. (A(*r=1*)~A(*r=2*)~A(*r=3*), ..., C(*r=1*)~C(*r=2*)~C(*r=3*)) have a stronger similar distribution among *r={1,2,3}*. In addition, only two clustering classes could be significantly identified as {(a1)~(a2)~(b1)~(b2)~(c1)~(c2), (a3)~(a4)~(b3)~(b4)~(c3~c4)} i.e. three maps are shown in (A~T, G~C) respectively.

In a convenient comparison, using a set of selected parameters, three groups of twelve 2D maps are compared in Fig. 9 for the files: *right* and *hc256, r={1,3}* to check their distribution properties contained in both DNA and created pseudo DNA sequences. Group (a) provides four Map$_V$ maps (a1–a4) for the file *right*; groups (b) and (c) provide four Map$_V$ maps (b1–b4) for *hc256, r=1* (c) and (c1–c4) for *hc256, r=3*.

Using a weak similar operator $\simeq$, for groups (a–c), four pairs of {(a1)$\simeq$(b1)~(c1), (a2)$\simeq$(b2)~(c2), (a3)$\sim$(b3)~(c3), (a4)$\sim$(b4)~(c4)} maps have a stronger similar distribution between *r={1,3}* and a weak similar distribution on A and T cases. In addition, only two clustering classes could be significantly identified as {(a1)~(a2)$\simeq$(b1)~(b2)~(c1)~(c2), (a3)~(a4)~(b3)~(b4)~(c3)~(c4)} i.e. three maps are strongly shown in relationships among (A~|$\simeq$T, G~C) for different cases respectively.

In addition, this set of results illustrates directly visual comparisons with stronger similarity between DNA and pseudo DNA on VMS maps, their similarly clustering distributions may indicate those maps with comparable mechanism to express real DNA sequences with extra A–T and G–C pairs of symmetric relationships in their higher levels of relationships applying the Stream Cipher mechanism.

5 Conclusion

This chapter proposes architecture to support the Variant Map System. Using a binary random sequence as input, a set of special pseudo DNA sequences can be generated. Under variant measures, probability measurement and normalized histogram, a pair of values can be determined by a series of controlled parameters. Collecting relevant pairs on multiple DNA sequences, four 2D maps can be generated.

The main results of this chapter provide the VMS architecture description in diagrams, main components, modules, expressions and important equations for the VMS. Core models and diagrams, sample results are illustrated to apply two types of data sets selected from real DNA sequences and generated from the pseudo random sequences from the Stream Cipher HC-256 for comparison under the VMS testing. After proper set of parameters selected, suitable visual distributions could be observed using the VMS. Results in Figs. 4, 5, 6, 7, 8 and 9 provide useful evidences systematically to support proposed VMS useful in checking higher levels of symmetric/similar properties among complex DNA sequences in both natural and artificial environment.

This construction could provide useful insights to spatial information on complex DNA expressions especially on large encoding RNA/DNA construction via 2D maps to explore higher levels of complex interactive environments in near future.

Acknowledgements Thanks to the school of software Yunnan University, to the key laboratory of Yunnan software engineering and the key laboratory for Conservation and Utilization of Bioresource for excellent working environment, to the Yunnan Advanced Overseas Scholar Project (W8110305), the Key R&D project of Yunnan Higher Education Bureau (K1059178) and National Science Foundation of China (61362014) for financial supports to this project. This work was supported by the Key Project on Electric Information and Next Generation IT Technology of Yunnan (2018ZI002), NSF of China (61362014) and Yunnan Advanced Overseas Scholar Project.

References

1. ESTREAM project, http://en.wikipedia.org/wiki/ESTREAM
2. H.J. Wu, Stream Cipher HC-256, ESTREAM 2004, http://www.ecrypt.eu.org/stream/p3ciphers/hc/hc256_p3.pdf
3. M. Santha, U.V. Vazirani, Generating Quasi-Random Sequences from slightly random sources. J. Comput. Syst. Sci. **33**, 75–87 (1986)
4. G. Paul, S. Maitra. *RC4 Stream Cipher and Its Variants* (CRC Press, 2011)
5. M. Gude, Concept for a high-performance random number generator based on physical random noise. Frequenz **39**, 187–190 (1985)
6. D. Eastlake, S.D. Crocker, J.I. Schiller, Randomness requirements for security, in *RFC 1750* (Internet Engineering Task Force, 1994)
7. C. Plumb, Truly random numbers. Dr. Dobbs J. **19**(13), 113–115 (1994)
8. G.B. Agnew, Random source for cryptographic systems, in *Advanced in Cryptology—EUROCRYPT'87 Proceedings* (Springer, 1988), pp. 77–81
9. A. Gehani, T. LaBean, J. Reif, DNA-based Cryptography, DIMACS Series in Discrete Mathematica and Theoretical Computer Science, 54, (2000) 233–249. http://www.cs.duke.edu/~reif/paper/DNAcrypt/DNA5.DNAcrypt.pdf

10. B.E. Bernstein, E. Birney, I. Dunham et al., An integrated encyclopedia of DNA elements in the human genome. Nature **489**(7414), 57–74 (2012). https://doi.org/10.1038/nature11247. PMID22955616
11. E. Pennisi, Genomics. ENCODE project writes eulogy for junk DNA. Science **337** (6099), 1159, 1161, (2012). https://doi.org/10.1126/science.337.6099.1159.PMID22955811
12. Michael Schoöniger, Arndt von Haeseler, Simulating efficiently the evolution of DNA sequences. Comput. Appl. Biosci. **11**(1), 111–115 (1995). https://doi.org/10.1093/bioinformatics/11.1.111
13. F. Piva, G. Principato, RANDNA: a random DNA sequence generator, Silico Biol. **6**, 0024 (2006). http://www.bioinfo.de/isb/2006060024/
14. C.M. Gearheart, B. Arazi, E.C. Rouchka, DNA-based random number generation in security circuitry. Biosystems **100**(3), 208–214 (2010)
15. O. O. Babatunde, On pseudorandom number generation from programmable and computable biomolecules: deoxyribonucleic (DNA) as a novel pseudorandom number generator. World Applied Programming, **1**(3), 215–227 (2011)
16. G.C. Sirakoulis, Hybrid DNA cellular automata for pseudorandom number generation, in *International Conference on High Performance Computing and Simulation (HPCS)* (2012)
17. Y. Zhang, Y. Zhu, Z. Wang, R.O. Sinnott, Index-based symmetric DNA encryption algorithm, in *4-th International Congress on Image and Signal Processing (CSIP) 2011*. http://dtl.unimelb.edu.au/researchfile287042.pdf
18. Y. Zhang, L.H.B. Fu, in *Research on DNA Cryptography, Applied Cryptography and Network Security*, ed by J. Sen (InTech Press, 2012), pp. 357–376. http://www.intechopen.com/books/applied-cryptography-and-network-security/research-on-dna-cryptography
19. Erez Lieberman-Aiden et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science **326**, 289–293 (2009). https://doi.org/10.1126/science.1181369
20. M.B. Gerstein, A. Kundaje, M. Hariharan et al. (2012) Architecture of the human regulatory network derived from ENCODE data, Nature 489 91–100, 2012. https://doi.org/10.1038/nature11245
21. J.M. Engreitz, A. Pandya-Jones, P. McDonel et al., Large noncoding RNAs can localize to regulatory DNA targets by exploriting the 3D architecture of the genome, in *Proceedings of The Biology of Genomes* (Cold Spring Harbor Laboratory Press, 2013), p. 122
22. K. Sakamoto, Molecular computation by DNA hairpin formation. Science **283**, 1223–1227 (2000)
23. A. Arneodo, C. Vaillant, et al., Multi-scale coding of genomic information: From DNA sequence to genome structure and function. Phys. Rep. **498**(2), 45–188 (2011)
24. S. Engela, A. Alemany, NuriaForns, folding and unfolding of a triple-branch DNA molecule with four conformational states. Phil. Mag. **91**(13), 2049–2065 (2011)
25. J.M. Urquiza, I. Rojas, et al., Method for prediction of protein–protein interactions in yeast using genomics/proteomics information and feature selection, Neurocomputing **74**(16), 2683–2690 (2011)
26. H. Zhang, X. Liu, A CLIQUE algorithm using DNA computing techniques based on closed-circle DNA sequences. Biosystems **105**(1), 73–82 (2011)
27. B. Banfai, H. Jia, J. Khatun et al., Long noncoding RNAs are rarely translated in two human cell lines. Genome Research, Cold Spring Harbor Laboratory Press **22**, 1646–1657 (2012). https://doi.org/10.1101/gr.134767.111
28. W. Ford, Doolittle, is junk DNA bunk? A critique of ENCODE, in *Proceedings of the National Academy of Sciences* (2013)
29. J. Wang, M. Yan. *Numerical Methods in Bioinformatics* (Science Press, 2013)
30. J.Z.J. Zheng, C.H. Zheng, A framework to express variant and invariant functional spaces for binary logic. Front. Electr. Electron. Eng. China. **5**(2), 163–172 (2010). http://www.springerlink.com/content/91474403127n446u/.

31. W.Z. Yang, J. Zheng, Pseudo-random number generator based on variant logic model, in *ChinaCom 2012 Conference Proceedings* (2012)
32. J. Zheng, C. Zheng, T. Kunii, A framework of variant logic construction for cellular automata, in *Cellular Automata—Innovative Modelling for Science and Engineering*, ed by A. Salcido (InTech Press, 2011), pp. 325–352. http://www.intechopen.com/chapters/20706
33. J. Zheng, C. Zheng, T. Kunii, Interactive maps on variant phase spaces—from measurements—micro ensembles to ensemble matrices on statistical mechanics of particle models, in *Emerging Application of Cellular Automata*, ed by A. Salcido (InTech Press, 2013), pp. 113–196. http://dx.doi.org/10.5772/51635
34. W.Z. Yang, J. Zheng, Variant pseudo-random number generator, *Hakin9 Extra*, **6**(13), 28–31 (2012). http://hakin9.org/hakin9-extra-62012/
35. W.Q. Zhang, J. Zheng, Randomness Measurement of Pseudorandom Sequence Using different Generation Mechanisms and DNA Sequence. J. Chengdu Univ. Inf. Technol. **27**(6), 548–555 (2012)
36. N.A. Tchurikov, O.V. Kretova, D.M. Fedoseeva et al., DNA double-strand breaks coupled with parp1 and hnrnpa2b1 binding sites flank coordinately expressed domains in human chromosomes. PLoS Genet. **9**(4), e1003429 (2013). https://doi.org/10.1371/journal.pgen.1003429