

Polarization Analysis of Gene Sequence Structures: Mapping of Extreme Local Polarization States

Dmitry A. Zimnyakov^{1,2}, Marina V. Alonova^{1*}, Anatoly V. Skripal³, Sergey S. Zaitsev⁴,
and Valentina A. Feodorova⁴

¹Yury Gagarin State Technical University of Saratov, 77 Polytechnicheskaya str., Saratov 410054, Russia

²Institute for Precision Mechanics and Control Problems of the Russian Academy of Sciences (IPTMU RAS),
24 Rabochaya str., Saratov 410028, Russia

³Saratov State University, 83 Astrakhanskaya str., Saratov 410012, Russia

⁴Federal Research Center for Virology and Microbiology (Saratov Branch), 53 Strelkovaya Divisiya str., building 6,
Saratov 410028, Russia

* e-mail: alonova_marina@mail.ru

Abstract. A method for visualization and identification of nucleotide sequences is proposed based on the synthesis of phase screens displaying the structure of the analyzed sequences and reconstruction of binary maps of the extreme values of the local Stokes vector components in the diffraction zone. This diffraction zone is formed due to reading out the phase screen by a coherent collimated beam with two orthogonally polarized (x - y) components. With different phase delays of the x - y components of the readout beam introduced by the phase screen elements, this causes a variety of local polarization states in the diffraction zone. The discrimination level for the local component of the Stokes vector chosen for the binary mapping is established near the extreme value for this component. Computer verification of the proposed method using nucleotide sequences for various strains of the model African swine fever (ASF) virus as the test objects shows its high efficiency in the detection of differences between two compared sequences corresponding to the same type of infectious agent. Analysis of the model data for the strains under study made it possible to establish a power-law character of correlation coefficients of the synthesized binary distributions of extreme local polarization states depending on the discrimination threshold detuning from the maximum value of the Stokes vector component used for the mapping. © 2022 Journal of Biomedical Photonics & Engineering.

Keywords: gene sequences; nucleotide triplets; phase screen; diffraction; Stokes vector components; extreme states; binary distributions.

Paper #3526 received 12 Sep 2022; revised manuscript received 24 Oct 2022; accepted for publication 28 Oct 2022; published online 3 Dec 2022. doi: [10.18287/JBPE22.08.040302](https://doi.org/10.18287/JBPE22.08.040302).

1 Introduction

Improvement of tools and methods for rapid diagnosis of viral infectious disease (VID) agents requires development and application of the latest approaches based on advances in molecular microbiology and bioinformatics. In recent years, emergence of new infections and more frequent outbreaks of VPDs with high epidemic or even pandemic potential associated with identification of mutant variants of viruses require a detailed study of the biodiversity of the relevant

pathogens and search for alternative technologies needed for rapid and highly accurate analysis of polymorphism of both individual targeted genes and entire genomes of VPD pathogens [1]. The process of studying the genetic polymorphism of biological objects includes two essential stages; at the first stage, the relevant DNA is sequenced using one of two major well-established techniques known as next-generation sequencing (NGS) technology: short-read sequencing and long-read sequencing [2]. As a result, the primary structure of a linear biomolecule of a certain length is determined

strongly indicating positions of four basic nucleotides (adenine, cytosine, thymine and guanine, assigned by the symbols A, C, T, G, respectively) in the target DNA. At the second essential stage, the structure of the obtained symbol sequence is analyzed in order to identify the most significant features that characterize the DNA under study. It should be noted that this analysis is no less important than the initial DNA sequencing and plays a decisive role in various bioinformatics applications. Typically, the analysis of the structure of symbol series obtained by sequencing of the microbial DNA and RNA is performed using software methods based on the statistical analysis of nucleotide-corresponding symbols and their groups in the studied series [3]. At the same time, new instrumental and instrumental-software (hybrid) approaches based on the principles of coherent-optical and polarization analysis of quasi-random two-dimensional structures can also be applied to solve such problems.

In Refs. [4, 5], a method for imaging and analysis of gene sequences based on computer synthesis of GB (Gene-Based) speckle patterns, which carry information regarding unique combinations of three of the four basic nucleotide bases (adenine (A), guanine (G), cytosine (C), thymine (T)), is considered. Thus, a fragment of a linear nucleotide sequence (for example, ATG, GGA, TGT), containing N^2 triplets, is transformed into a matrix of $N \times N$ elements, each of which is assigned an integer value in the range from 0 to 63, depending on the combination of nucleotides in the triplet. In particular, transformation of the alphabetic code of the triplet $X_{i,j}$ ($0 \leq i \leq N, 0 \leq j \leq N$) into the corresponding value of the matrix element can be performed according to the following algorithm [4, 5]: $X_{i,j} = 16\Xi_1 + 4\Xi_2 + \Xi_3 - 21$, where each of the three factors Ξ_1, Ξ_2, Ξ_3 (indices “1–3” correspond to the position of the nucleotide in the triplet) takes the value from 1 to 4 ($A \leftrightarrow 1, C \leftrightarrow 2, G \leftrightarrow 3, T \leftrightarrow 4$). Thus, the maximum value occurs when the triplet “TTT” appears in the sequence, and the minimum (zero) value occurs when the triplet “AAA” appears.

The matrix generated in this way can be used as a spatial phase modulator for a coherent light beam with the depth of phase modulation of each element equal to $K_\varphi X_{i,j}$ (K_φ is a scaling factor). As a result of diffraction of a coherent readout beam on the synthesized phase modulator, a GB speckle-structure is formed in the far diffraction zone (for example, in the focal plane of the Fourier-transforming lens).

Stochasticity of the distribution of the diffracted light field is due to the close-to-random character of the $X_{i,j}$ – distribution; accordingly, at sufficiently large values of the scaling factor $K_\varphi > \pi$ the formed GB speckle-structure will tend to the so-called developed speckle field [6]. The developed speckle patterns are characterized by the Gaussian probability density functions of the real and imaginary components of the

complex amplitude with zero mean values. In turn, this leads to a negative exponential form of intensity probability density functions. At the same time, for each nucleotide sequence transformed into a stochastic phase matrix, a unique GB speckle microscopic structure is formed, which can be used to identify a nucleotide sequence for a particular gene. In Refs. [4, 5], it was proposed to search for mutation-driven differences in gene sequences corresponding to biological objects of one pathogenic microorganism, *Chlamydia trachomatis*, based on the correlation analysis of GB speckle structures synthesized in this way.

The concept of visualization and recognition of differences in nucleotide sequences (polymorphism) using synthesized GB speckles seems quite original; however, it has certain drawbacks that limit the possibilities of its instrumental implementation and practical application. One of the most significant drawbacks is a rather weak effect of decorrelation of spatial intensity distribution of the synthesized GB speckle structure with respect to the initial distribution when a small number of triplets (1-2) in the analyzed target gene sequence is substituted. In addition, initial consideration of this concept within the framework of scalar diffraction theory greatly limits the functionality that can be significantly expanded by applying the principles of polarization optics.

The aim of this work was to further develop the methodology for coherent-optical imaging and analysis of nucleotide sequences of infectious agents based on the principles of local stochastic transformations of the polarization state of the reading coherent beam by a synthesized spatial modulator of local phase shifts of orthogonally polarized components of the reading beam and identification in the diffracted field of the regions corresponding to extreme values of the local components of the Stokes vector. It is expected that this approach should be characterized by substantially higher sensitivity to local changes in the structure of the analyzed triplet sequences in comparison with the correlation analysis of “scalar” GB speckles. Such expectations are stipulated by the fact that extreme states of the normalized local components of the Stokes vector (e.g., minima of the first component s^0 or absolute values of the fourth component s^3 close to 1, which determine contribution of the circular polarization to the local polarization state) are achieved at a strictly definite phase and amplitude relations between the x–y components of the diffracted light field at a given point. Changing the triplet in the original sequence of the target gene by replacing at least one nucleotide should lead to violation of these phase relations and, consequently, to the shifts in the zones of extreme states in the observation plane, which can be identified by the correlation analysis of binarized distributions of extreme states.

2 Polarization Visualization of the Structure of Nucleotide Sequences

The proposed principle of transforming the sequence containing N^2 triplets of nucleotides into a phase screen is fundamentally different from the principle considered in Ref. [4]. Each triplet is associated with the 2×2 submatrix $\{a_{l,o}\}$; the matrix elements and base nucleotides are assigned to each other as $a_{0,0} \leftrightarrow A$, $a_{0,1} \leftrightarrow G$, $a_{1,0} \leftrightarrow C$, $a_{1,1} \leftrightarrow T$. Note that this assignment is conditional and does not affect the subsequent analysis; the only condition is that the transformation rule is constant for all triplets in the nucleotide sequence of a particular gene. The values of submatrix elements determine the number of corresponding nucleotides in the triplet; thus, the sum of all submatrix elements always equals 3. Accordingly, the minimum number of null elements in the submatrix is 1, and the maximum number is 3. The phase screen matrix $D_{i,t}$ of the size $2N \times 2N$ is synthesized by sequentially combining all N^2 submatrices.

Consider the process of diffraction on a synthesized phase screen of a linearly polarized coherent light beam; the polarization plane of the beam forms the angle $\pi/4$ with the sides of the screen. The diffracted light field (formed, for example, in the focal plane of the Fourier-transforming lens ("frequency plane")) is a superposition of linearly polarized components with orthogonal polarization directions, where the spatial distribution of each can be calculated using identical discrete Fourier transforms (see, for example, Ref. [7]):

$$E_{k,m}^{x,y} = \frac{1}{4N^2} \times \sum_{i=-N}^{N-1} \sum_{t=-N}^{N-1} \exp[-j \cdot scale \cdot \{(\pi/N)(k \cdot i + m \cdot t) - f_{i,t}^{x,y}\}] \quad (1)$$

In the Eq. (1), the index "y" corresponds to the direction along the columns of the synthesized phase screen and, respectively, "x" along the rows. The amplitudes of the x- and y-components of the incident readout beam on the screen are assumed to be 1. The scaling parameter *scale* is used to select the analyzed region of the diffracted field. In the case of the instrumental implementation of the scheme under consideration, the following relationship takes place between the dimensional scales in the input (i, t) and output (k, m) planes: $\Delta_{k,m} = F\lambda/\Delta_{i,t}$. Here, F is the focal length of the used Fourier-transforming lens, λ is the wavelength of laser radiation, and $\Delta_{k,m}, \Delta_{i,t}$ determine the dimensional separation between adjacent points in the corresponding systems of discrete coordinates.

The terms $\varphi_{i,t}^x$ are defined as $\varphi_{i,t}^x = K_\varphi D_{i,t}$, and the terms $\varphi_{i,t}^y$ are defined as $\varphi_{i,t}^y = \varphi_{i,t}^x + \Delta\Phi$ ($\Delta\Phi$ is the

constant phase shift between the x- and y-components of the readout beam passing through the screen).

For each pair of the calculated values of complex amplitudes of diffracted components $E_{k,m}^x, E_{k,m}^y$ ($0 \leq k < 2N, 0 \leq m < 2N$), we calculate normalized local values of the Stokes vector components as:

$$\begin{cases} s_{k,m}^0 = (|E_{k,m}^x|^2 + |E_{k,m}^y|^2)/2; \\ s_{k,m}^1 = (|E_{k,m}^x|^2 - |E_{k,m}^y|^2)/2s_{k,m}^0; \\ s_{k,m}^2 = 2|E_{k,m}^x||E_{k,m}^y|\cos(\delta_{k,m})/2s_{k,m}^0; \\ s_{k,m}^3 = 2|E_{k,m}^x||E_{k,m}^y|\sin(\delta_{k,m})/2s_{k,m}^0, \end{cases} \quad (2)$$

where $\delta_{k,m}$ is the phase difference of x- and y-components of the diffracted field at the corresponding point of the diffraction plane.

3 Model Polarization Images of the Nucleotide Sequences of the p72 Target Gene (by the Example of the African Swine Fever Pathogen; Strains "HuB20", "Zaire", "Ulyanovsk 19/WB-5699")

As an example, let us consider the results of the computer experiment on polarization imaging of the nucleotide sequences of the target gene p72 of three different strains of the model African swine fever virus (ASF) ("HuB20" (NCBI GenBank access number: MW521382.1), Zaire (NCBI GenBank access number MW296952.1), Ulyanovsk 19/WB-5699 (NCBI GenBank access number MW306192.1) [8]). ASF is known to be a contagious viral disease of animals (pigs and boars) with a lethal outcome of up to 100% [9]. The initial fragment of the nucleotide sequence from the start codon for the "HuB20" strain is ATGGCATCAGGAGGAGGAGCTTTT..., and the corresponding fragment of the first line of the synthesized phase matrix is

$$\begin{matrix} 1 & 1 & 1 & 1 & 1 & 0 & 1 & 2 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 3 \dots \end{matrix} \quad (3)$$

The differences in the structures of nucleotide triplet (codon) sequences within the analyzed fragments of 625 triplets for the strains under consideration are presented in Table 1. Note that the sequence of the "Zaire" strain differs significantly from those of "HuB20" and "Ulyanovsk" (30 and 29 triplets, respectively), while the difference between the "HuB20" and "Ulyanovsk" sequences is minimal (1 triplet).

For simulation purposes, the value of the scaling factor K_φ was taken as π ; it can be shown that for the case of equal probability of the finding values 0, 1, 2, and 3 at a random choice of the matrix element $D_{i,t}$, the value

of $K_\varphi = \pi$ gives zero value of the element $s_{0,0}^0$ (i.e., suppression of the non-diffracted component of the readout beam passing through the modulator in the near-axis region of the observation plane). Indeed, in the case of the equal probability of finding matrix elements ($D_{i,t}$), the real part of the on-axis amplitude of the x-component of the diffracted field can be represented as follows:

$$\begin{aligned} \operatorname{Re}(E_{0,0}^x) &\propto (1-4\rho) + \rho + \rho \cos(K_\varphi) + \\ &+ \rho \cos(2K_\varphi) + \rho \cos(3K_\varphi) = (1-3\rho) + \\ &+ \rho(\cos(K_\varphi) + \cos(2K_\varphi) + \cos(3K_\varphi)). \end{aligned} \quad (4)$$

Here, $\rho = 0.25$ is the probability of finding one of the elements 0, 1, 2, and 3 at random choice. Accordingly, the imaginary part of $E_{0,0}^x$ can be written as:

$$\operatorname{Im}(E_{0,0}^x) \propto \rho(\sin(K_\varphi) + \sin(2K_\varphi) + \sin(3K_\varphi)), \quad (5)$$

and the on-axis intensity of the x-component is:

$$\begin{aligned} I_{0,0}^x &\propto [\operatorname{Re}(E_{0,0}^x)]^2 + [\operatorname{Im}(E_{0,0}^x)]^2 \propto 1-6\rho + \\ &+ 12\rho^2 + 2\rho(1-\rho)\cos K_\varphi + \\ &+ 2\rho(1-2\rho)\cos 2K_\varphi + 2\rho(1-3\rho)\cos 3K_\varphi. \end{aligned} \quad (6)$$

It can be seen that substituting $K_\varphi = \pi$ for $\rho = 0.25$ gives a zero value of $I_{0,0}^x$. A similar consideration can be carried out for the y-component under the condition $\Delta\Phi = \text{const}$.

Significantly non-zero values of intensity of the non-diffracted component indicate, in particular, the dominance of some of the four nucleotides in the original sequence. The value of the phase shift $\Delta\Phi$ was assumed to be $\pi/2$. This choice is because in this case the local polarization states of the boundary field (directly behind the phase screen) will correspond to either the right circular polarization (in the cases if $D_{i,t} = 0$ or 2) or the left circular polarization (in the cases if $D_{i,t} = 1$ or 3). This causes a broader variety of local polarization states in the (k, m) plane and, accordingly, a greater probability of the appearance of states with $s_{k,m}^3$ close to the extreme values of 1 or -1 , compared, for example, with the case of $\Delta\Phi = \pi$.

Based on the size of the generated phase matrices $D_{i,t}$ (50×50), the N value in the simulation was taken to be 100, and the indices k and m varied from 0 to 99; the maximum value $scale$ should not exceed 0.5 to avoid the effect of a spatial frequency overlap (aliasing) [10].

Fig. 1 shows the synthesized spatial distributions of $\lg(s_{k,m}^0)$ as grayscale maps in the near-axis region of the frequency plane ($scale = 0.1$) for three strains under consideration (the logarithmic scale of representation chosen due to very wide intervals of intensity changes of synthesized intensity distributions within the near-axis region). Note the pronounced speckle modulation of the diffracted field. The phase differences of the boundary field (immediately behind the phase screen) for different elements of the phase matrix take only three possible values ($-\pi, 0, \pi$) for the used phase modulation algorithm, though stochasticity of spatial distributions of the elements of the synthesized phase screens leads to speckle formation in the diffracted zone. Substantially nonzero values of the diffracted field intensity at the beam axis (the central region of the synthesized distributions) indicate that the nucleotide distributions in the analyzed sequences are substantially non-uniform.

Table 1 Differences in triplet sequences for the analyzed strains.

# in the sequence (from the start codon)	Composition of triplets		
	HuB20	Ulyanovsk	Zaire
26	CTG	CTG	CTT
30	ATC	ATC	ATT
44	GAT	GAT	GAC
107	CAT	CAT	CAC
115	TCA	TCA	TCG
126	TCC	TCC	GCC
132	GGG	GGG	GGT
149	CCC	CCC	CCT
160	CCT	CCT	CCC
164	CCC	CCC	CCT
200	CTA	CTA	CTG
228	GTT	GTT	GTC
268	TGT	TGT	TGC
275	TTT	TTT	TTC
278	CAG	CAG	CAA
298	ATC	ATC	ATT
333	TTG	TTG	CTG
338	AAT	AAT	AAC
347	TCA	TCA	TCG
376	GTA	GTA	ATA
379	TCA	TCA	TCG
383	GCT	GCT	CCT
432	CGC	CGC	CGA
434	TCG	TCG	TCC
483	CAA	CAA	CAG
490	TTT	TTC	TTC
500	CCC	CCC	CCT
523	TCT	TCT	TCG
552	AAA	AAA	AAG
570	GCG	GCG	GCA

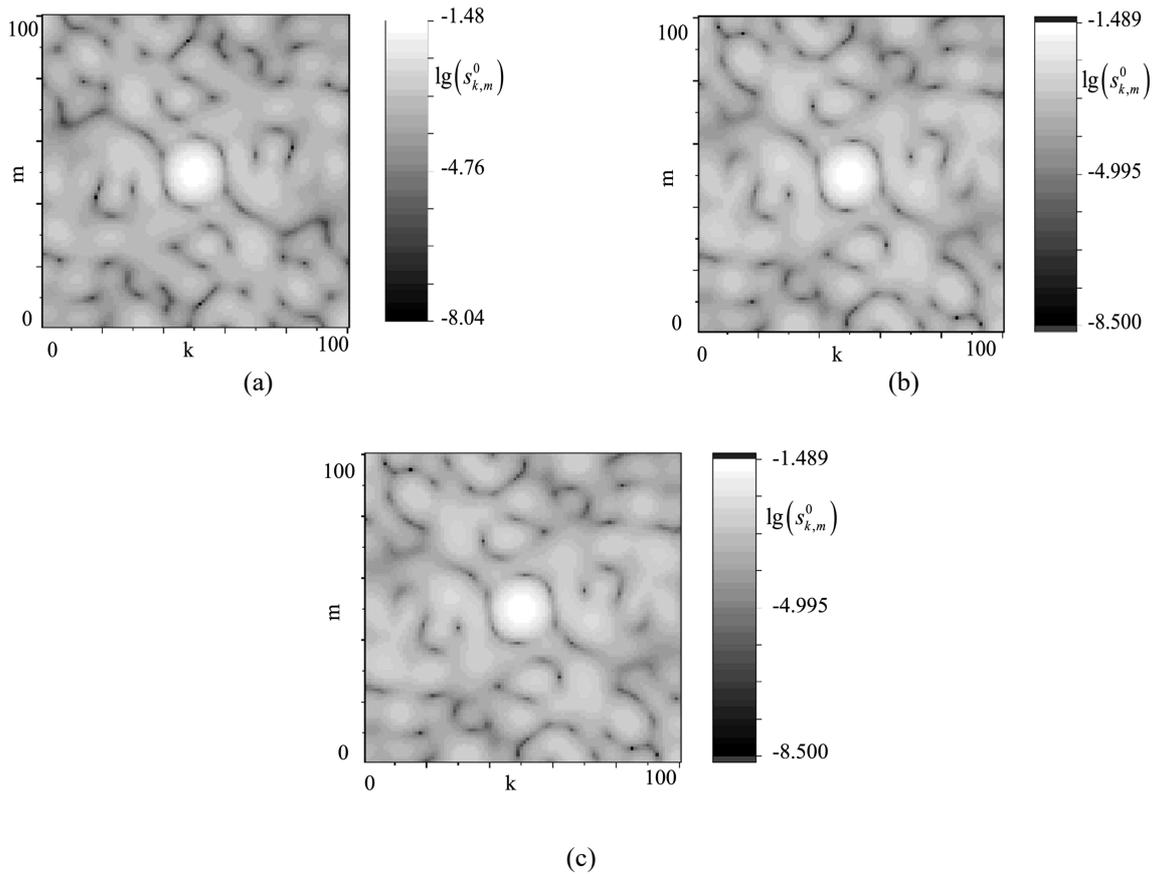


Fig. 1 Model distributions of local values of the logarithm of the first component of the Stokes vector in the frequency plane ($scale = 0.1$) for target genes of the ASF virus strains: (a) “HuB20”; (b) “Ulyanovsk”; and (c) “Zaire”.

Visually, distributions of $lg(s_{k,m}^0)$ appear to be insignificantly different from each other, but spatial distributions of the special points (in particular, for which the $lg(s_{k,m}^0)$ values are less than some threshold value—conditional points and “zero amplitude” lines) are expected to show significant differences from strain to strain.

Fig. 2 shows similar distributions of the components $s_{k,m}^3$ characterizing contribution of the left ($s_{k,m}^3 = -1$) or right ($s_{k,m}^3 = 1$) circular component to the local polarization state at the point (k, m) . Antisymmetric character of these distributions ($s_{k-50,m-50}^3 = -s_{50-k,50-m}^3$), caused by the applied algorithm of phase modulation of the x- and y-components of the readout beam passing through the synthesized screen, is remarkable. Uniqueness of $s_{k,m}^3$ distributions for each analyzed strain should also be noted. The next section is devoted to the analysis of such uniqueness.

4 Mutational Changes in Nucleotide Sequences and Variability of the Binary Maps of Extreme Polarization States

The obtained model for spatial distributions of the local values of the Stokes vector component $\{s_{k,m}^3\}$ can be transformed into binary forms by introducing a certain threshold value of s_{th}^3 and subsequent binarization according to the rule:

$$\begin{cases} s_{k,m}^3 \geq s_{th}^3 \rightarrow \tilde{s}_{k,m}^3 = 1(0); \\ s_{k,m}^3 < s_{th}^3 \rightarrow \tilde{s}_{k,m}^3 = 0(1). \end{cases} \quad (7)$$

Binary $\tilde{s}_{k,m}^3$ distributions for the target genes of the strains “HuB20”, “Zaire”, and “Ulyanovsk” at $s_{th}^3 = -0.99$ (selection of the near-circular left local polarization states) are shown in Fig. 3. The maximum allowable value of $scale$, equal to 0.5, was used in the modeling. For quantitative comparison of the limit state maps synthesized in this way, a correlation coefficient is defined as:

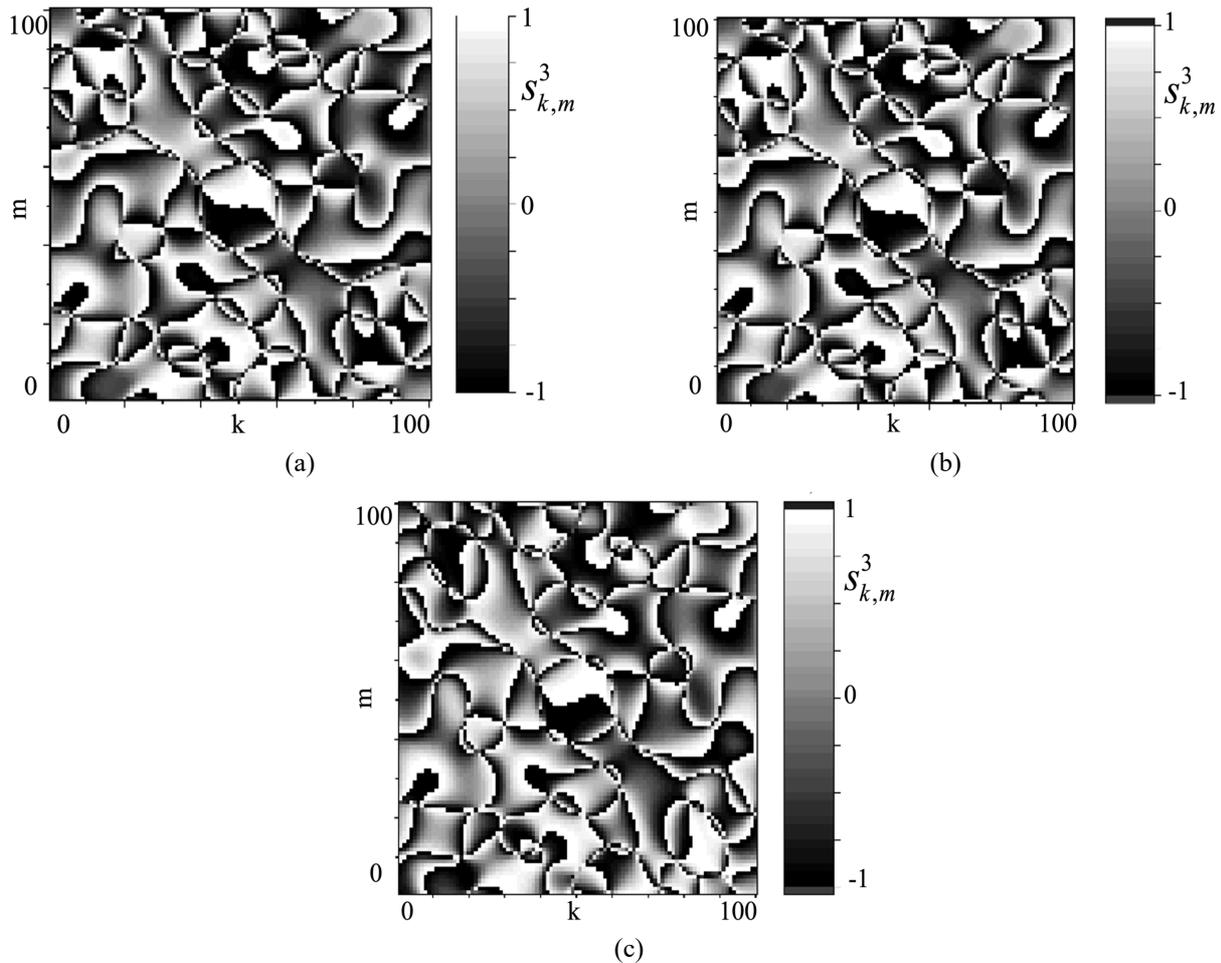


Fig. 2 Model distributions of the local values of the third component of the Stokes vector in the frequency plane (*scale* = 0.1) for the target genes of the ASF virus strains: (a) “HuB20”, (b) “Ulyanovsk”, and (c) “Zaire”.

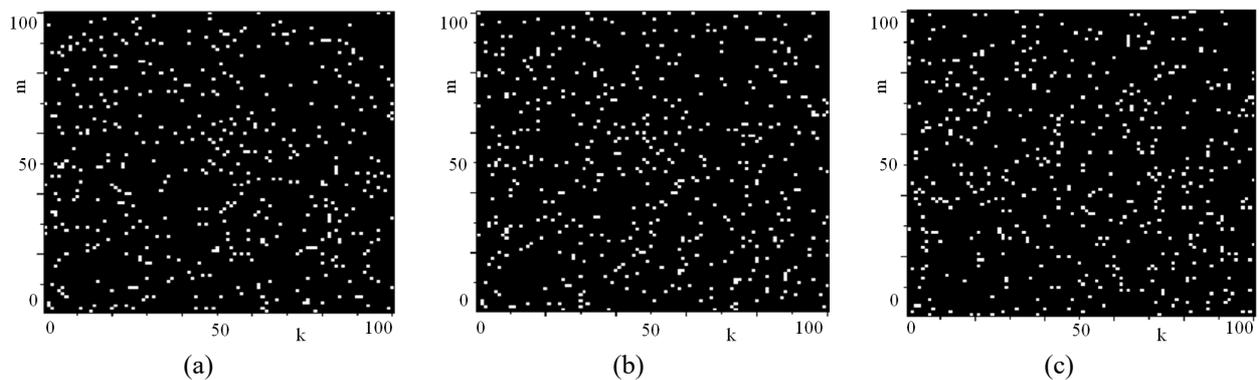


Fig. 3 “Panoramic” binary maps for the polarization extreme states (left elliptical polarization with $s_{k,m}^3 < -0.99$, *scale* = 0.5) for the target genes of ASF virus strains: (a) “HuB20”; (b) “Ulyanovsk”; and (c) “Zaire”.

$$R_{1,2}^3 = \frac{\sum_{k,m} \tilde{s}_{k,m}^{1,3} \tilde{s}_{k,m}^{2,3}}{\sum_{k,m} (\tilde{s}_{k,m}^{1,3})^2}, \tag{8}$$

where indices “1” and “2” correspond to the nucleotide sequences for different strains. Accordingly, for the

“HuB20” and “Ulyanovsk” strains (Fig. 5 (a, b)) $R_{H,U}^3$ is 0.472, and for the “HuB20” and “Zaire” strains $R_{H,K}^3 \approx 0.099$.

It is of interest to analyze sensitivity of the correlation coefficient to the binary distributions of extreme states $\{\tilde{s}_{k,m}^3\}$ introduced in this way to the changes in the structure of nucleotide sequences caused by random

substitutions to a part of the triplets – one nucleotide for another. This analysis was performed for an original nucleotide sequence corresponding to the “HuB20” strain by the random changes in one, two, or more triplets providing that the sum of the elements of the substituent submatrix is equal to 3. The average values of the correlation coefficient were calculated as

$$\langle R_{1,2}^3 \rangle = \left\langle \frac{\sum_{k,m} \tilde{s}_{k,m}^{1,3} \tilde{s}_{k,m}^{2,3}}{\sum_{k,m} (\tilde{s}_{k,m}^{1,3})^2} \right\rangle, \tag{9}$$

depending on the number of the substitutions N_r . Averaging was carried over the groups of 10 binary maps of extreme states $\{\tilde{s}_{k,m}^{2,3}\}$, synthesized for the sequences with a given number N_r of random substitutions at the threshold values $s_{th}^3 = -0.95$ and $s_{th}^3 = -0.99$.

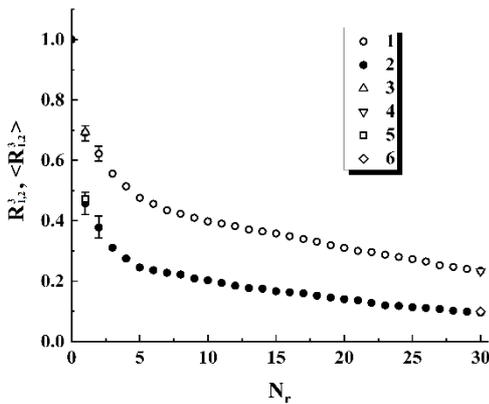


Fig. 4 Average correlation coefficients of the binary distributions $\tilde{s}_{k,m}^3$ (Eq. (9)) to the “HuB20” strain in the random substitutions of a given number N_r of unit nucleotides in triplets (1, 2) compared with the values of correlation coefficients to the “HuB20” – “Ulyanovsk” (3, 5) and “HuB20” – “Zaire” (4, 6) pairs. (1, 3, 4) discrimination level is -0.95 ; and (2, 5, 6) discrimination level is -0.99 .

The obtained values $\langle R_{1,2}^3 \rangle = f(N_r)$ are shown in Fig. 4; selectively shown confidence intervals for $N_r = 1$ and 2 correspond to the significance level of 0.9. Note a sharp drop in $\langle R_{1,2}^3 \rangle$ when replacing single elements in the original sequences and the subsequent slow decrease when N_r further increases. Assuming a uniform distribution of the displayed polarization extreme states at a given threshold s_{th}^3 with the density $P_{th} = N_{th}/4N^2$ (i.e., the density is equal to the ratio of the number of extreme states to the number of pixels in the frequency plane), we can show that if $\aleph \rightarrow N^2$

(random nucleotide substitution in all triplets), then $\langle R_{1,2}^3 \rangle \rightarrow P_{th}$. Fig. 4 also shows $R_{H,U}^3$ (sequences for the “HuB20” and “Ulyanovsk” strains) and $R_{H,Z}^3$ (sequences for the “HuB20” and “Zaire” strains). Note to the absence of the overlapping confidence intervals in the figure for a small number of triplets with nucleotide substitutions; this indicates high sensitivity of the correlation coefficient of the binary distributions of extreme local polarization states to minor the changes in the structure of nucleotide sequences of the analyzed target genes.

An increase in discrimination threshold leads to a decrease in the correlation coefficient. Fig. 5 shows the dependences of correlation coefficients for the pairs of the nucleotide sequences of the ASF virus strains “HuB20” – “Ulyanovsk” ($N_r = 1$) and “HuB20” – “Zaire” ($N_r = 30$) (Table 1) on the value of the discrimination level detuning from the minimum possible value $s_{th}^3 = -1$.

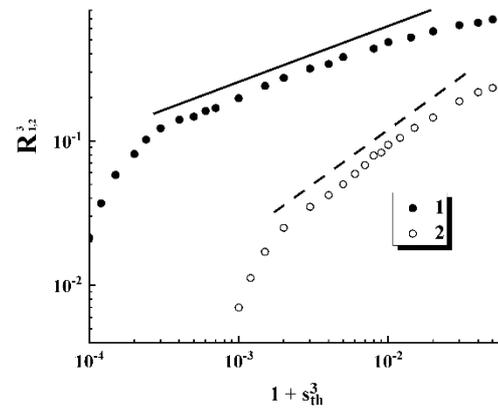


Fig. 5 Correlation coefficients for the pairs of nucleotide sequences of the ASF virus strains “HuB20” – “Ulyanovsk” (1) and “HuB20” – “Zaire” (2) depending on detuning of the discrimination threshold s_{th}^3 from the minimum possible value (-1) . The solid and dashed lines are applied as guides for the eye and show trends in the behavior of the dependences that are close to the power-law trend.

Note that for sufficiently wide intervals of the changes in the detuning parameter $1 + s_{th}^3$ (on the order of 1.5–2 decades), the dependences presented in Fig. 5 admit with an acceptable accuracy the power approximations of the form

$$R_{1,2}^3 \propto (1 + s_{th}^3)^\alpha, \tag{10}$$

similar to the behavior of complex systems near the critical points (for example, in the region of the phase transition).

At the same time, in contrast to the classical critical phenomena, the established close-to-power decrease in correlation coefficients as they approach the minimum

possible value s_{ih}^3 is not universal in the broad sense (the α values for the pairs of nucleotide sequences of the ASF virus strains “HuB20” – “Ulyanovsk” ($\alpha \approx 0.39$) and “HuB20” – “Zaire” ($\alpha \approx 0.75$) differ significantly in the case under consideration, i.e. depend on the number of triplets substituted. The issue of such “quasi-critical” behavior of the synthesized binary distributions of polarization extreme states depending on the structure of synthesized phase screens is beyond the scope of this work and is an object of further research.

5 Conclusion

Thus, the considered method of polarization imaging of the structure of nucleotide sequences in the targeted genes of the model ASF virus strains using binary mapping of extreme local polarization states of the diffracted reading beam demonstrates high sensitivity to small changes in the structure of the sequences. The synthesized map of extreme states at an appropriately high discrimination threshold is a unique identifying object that corresponds solely to the analyzed nucleotide sequence.

Note that the proposed method, in contrast to the correlation analysis of the synthesized GB speckles [4, 5], allows rather simple instrumental implementation for research and demonstration purposes. The corresponding polarimetric system can consist of a single-mode HeNe laser with a beam expander-collimator and a spatial filter used to improve the quality of the linearly polarized beam, a multi-element liquid crystal spatial light modulator with computer control (for example, manufactured by Holoeye (Germany) or a similar unit), the Fourier-transform lens, a set of polarization filters, and a CMOS camera. In particular, when using a He-Ne-laser, the transmissive spatial light modulator Holoeye LC 2012 (the pixel size is 36 μm), and a Fourier transform lens with a focal length of about 100 mm, the characteristic dimensions of the diffraction pattern in the focal plane are covered by the photosensitive zones of modern CMOS sensors. The number of CMOS sensor pixels within the area of interest will be of the order of 3×10^4 for a CMOS pixel size of 10 μm ; this is more than enough to implement the discussed technique. Following from the expected dimensions of the used zone of the spatial light modulator (no more than 3×3 mm) and the

requirement for readout beam uniformity within this zone, it can be assumed that the beam expansion factor can be no more than 20–25. Note that in the considered model of the formation of diffraction patterns in the focal plane of the Fourier-transforming lens (see Eq. (1)), a uniform distribution of the readout field amplitude over the surface of the synthesized phase screen ($\varphi_{i,t}^{x,y}$) is assumed. This uniformity can be ensured at the stage of formation of the readout beam by using a beam-expanding telescopic system with a pinhole diaphragm as the beam cleaner. With a sufficiently large magnification factor of the telescopic system (as noted above, about 20–25 or more), in the paraxial region of the phase screen used for phase modulation, a close-to-uniform amplitude distribution will be formed. On the other hand, the problem of spatial inhomogeneity of the readout beam can be taken into account within the framework of the model (1) by additionally introducing a matrix of weighting amplitude coefficients. This matrix can be obtained using preliminary calibration of the considered polarimetric system. Note that the instrumental implementation is beyond the scope of this study and will be carried out in the future.

Also, it is necessary to note, that the considered approach based on the four-gradation matrix coding of initial nucleotide sequences, superposition of two Fourier transforms with a certain phase shift, and nonlinear transformations of the obtained superposition with a subsequent binarization and synthesis of the maps of extreme states can be considered as the basis for computer algorithms applied in identification of the objects with a complex quasi-random structure.

The proposed method is sufficiently flexible and allows for various modifications of the boundary field phase modulation algorithms that extend the functional capabilities of mapping and identifying the structure of nucleotide sequences in the target genes of infectious agents.

Disclosures

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the Russian Science Foundation (Project № 22-21-00194).

References

1. S. Kumar, G. S. Kumar, S. S. Maitra, P. Malý, S. Bharadwaj, P. Sharma, and V. D. Dwivedi, “[Viral informatics: bioinformatics-based solution for managing viral infections](#),” Briefings in Bioinformatics 23(5), bbac326 (2022).
2. S. Goodwin, J. D. McPherson, and W. R. McCombie, “[Coming of age: ten years of next-generation sequencing technologies](#),” Nature Reviews Genetics 17, 333–351 (2016).
3. N. N. Pozdnicenko, M. S. Stupin, A. S. Gumenyuk, M. S. Doroshenko, and O. P. Shafeeva, “[Algorithms and software for obtaining dissimilar order and high order of symbolic sequences](#),” Journal of Physics: Conference Series 1050(1), 012063 (2018).

4. S. S. Ulyanov, O. V. Ulianova, S. S. Zaytsev, Y. V. Saltykov, and V. A. Feodorova, “[Statistics on gene-based laser speckles with a small number of scatterers: implications for the detection of polymorphism in the Chlamydia trachomatis omp1 gene](#),” *Laser Physics Letters* 15(4), 045601 (2017).
5. S. S. Ulyanov, S. S. Zaytsev, O. V. Ulianova, Y. V. Saltykov, and V. A. Feodorova, “[Using of methods of speckle optics for Chlamydia trachomatis typing](#),” *Proceedings of SPIE* 10336, 103360D (2017).
6. J. W. Goodman, *Statistical Optics*, 2nd ed., John Wiley & Sons, Hoboken, New Jersey, USA (2015). ISBN: 978-1-119-00945-0.
7. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed., Publishers, Englewood, Colorado, USA (2005). ISBN: 978-0974707723.
8. A. Mazloun, A. van Schalkwyk, A. Shotin, A. Igolkin, I. Shevchenko, K. Gruzdev, and N. Vlasova, “[Comparative Analysis of Full Genome Sequences of African Swine Fever Virus Isolates Taken from Wild Boars in Russia in 2019](#),” *Pathogens* 10(5), 521 (2021).
9. D. Beltran-Alcrudo, M. A. C. C. Gallardo, S. A. Kramer, M. L. Penrith, A. Kamata, and L. Wiersma, [African swine fever: detection and diagnosis](#), Food and Agriculture Organization of the United Nations (FAO), Rome, Italy (2017). ISBN: 978-92-5-109752-6.
10. R. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed., McGraw Hill, New York, USA (1986). ISBN 978-0070070165.