МЕТОДИКА

https://doi.org/10.15407/frg2024.04.353
УДК 577.212.3+004.9

# DNA WALK DIAGRAM IN TRIANDER AND jsTRIANDER APPLICATIONS

**V.P. DUPLIJ[1], N.V. DUPLII[2], S.A. DUPLIJ[3]**

*[1]Institute of Cell Biology and Genetic Engineering, National Academy of Sciences of Ukraine*
*148 Akademika Zabolotnogo St., Kyiv, 03143, Ukraine*
*e-mail: duplijv@icbge.org.ua*
*[2]National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»*
*37 Prospect Beresteiskyi, Kyiv, 03056, Ukraine*
*[3]Center for Information Technology (CIT), University of Münster*
*7-13 Röntgenstrasse, Münster, 48149, Germany*

This work aimed to create software capable of presenting nucleotide sequences in a form convenient for recognition and comparison by humans. For this, the method of DNA walk by vectors of different lengths in the directions North-South-West-East was chosen. Each nucleotide has its direction. It is shown that diagrams of the triander type, represented by three branches, each corresponding to the position of a nucleotide in a codon, are similar for the same genes of different biological species and may differ for different genes. Comparing the diagrams allows one to notice even minor differences between gene sequences for different species of the same genus. The sources and binaries for the Windows operating system of Triander software were placed at https://icbge.org.ua/eng/Triander. The Web application jsTriander is located at https://triander.icbge.org.ua and can be used both online and offline.

***Key words*:** DNA walk diagram, sequence analysis, triander, software.

The avalanche-like increase in the amount of data on the genetic diversity of life due to the progress in cheapening and accelerating the sequencing process requires new methods of manipulating such information. Open genetic databases such as GenBank provide a set for primary analysis of nucleotide sequences [1]. The ability to determine the DNA sequence in combination with statistical methods provides an essential tool for obtaining hidden information about the dynamics of the evolution process, especially after the complete genomes of organisms became available [2]. Along with the need to create software for computer analysis of accumulated genetic texts, there was also a need to present them conveniently for humans.

The thumbnails of documents in modern file managers are generated to exploit the human capacity to find information based on image recognition. However, we still have no such handy instrument for files with genetic information, and all that remains for us is to use their long names.

Citation: Duplij V.P., Duplii N.V., Duplij S.A. DNA walk diagram in Triander and jsTriander applications. *Fisiol. rast. genet.*, 2024, **56**, No. 4, pp. 353—361. https://doi.org/10.15407/frg2024.04.353

Currently, there are many systems for visualizing genetic information. Most, in one way or another, do not show the sequence itself, but only reflect the relative location of genes, regulatory units, coding and non-coding regions. In these cases, one has to completely rely on the application algorithms, and lose sight of the sequence itself.

This work aimed to create applications for transforming nucleotide sequences into graphical objects acceptable for human recognition and possible visual analysis.

## Materials and methods

The desktop variant of the program for interactive visualization of nucleotide sequences (Triander) was created in the free software development environment Lazarus version 1.2.6 [3], using the Free Pascal compiler version 2.6.4 [4].

JavaScript programming language [https://developer.mozilla.org/en-US/docs/Web/JavaScript] without any additional library or framework was used for development of jsTriander Web application. The HyperText Markup Language (HTML) [5] and CSS [6] style sheet language were involved in the building of user interface in this project.

The Triander program was tested under the 32-bit version of Windows XP and 64-bit versions of Windows 7, 8.1, 10, 11. The jsTriander was tested in desktop and mobile versions of Mozilla Firefox (121.0) and Google Chrome (120.0) browsers.

The source code of the Triander application and its compiled binary code for Windows are freely available at https://icbge.org.ua/eng/Triander. The source code of jsTriander is available by full page downloading from https://triander.icbge.org.ua and can be used both online and offline.

## Results and discussion

It is widely known that the complexity of visual analysis of genetic texts written in a four-letter alphabet increases significantly with increasing text length. The human eye is designed in such a way that it either reads each letter separately, without noticing nucleotide patterns, or rejects sequences of letters that do not form familiar words [7]. The arrangement of the bases on the staff as music notes makes it easier to recognize individual patterns after some practice [7], but breaking patterns during transfer creates many more problems than with regular reading. In addition, it is quite difficult to cover a long sequence entirely.

Unlike the above-mentioned visualization methods, we offer a natural representation of nucleotide sequences in the form of curves, where multidirectional vectors are assigned to bases. In what follows, we will name such objects as nucleotide vectors and the curves formed by them as nucleotide curves.

The «H-curves» introduced in [8] provide an unambiguous representation of the sequence in three-dimensional space and, perhaps, would be ideal when analyzing them on three-dimensional display devices. Nevertheless, due to the extremely narrow distribution of such devices, it is necessary to work with two-dimensional projections of curves, which leads to partial loss of visual information. A two-dimensional version of this

method [9] has become widespread and appeared useful for detecting replication initiation sites in genomes [10]. However, such a curve often passes through the same places on the diagram due to the lack of a mandatory vertical offset. In this case, information losses can be significant; in certain areas, parts of the curve merge into spots (Fig. 1, *a*).

This issue is mostly eliminated in a visualization system, where nucleotides are represented by vectors that differ simultaneously in direction and length [11] (Fig. 1, *b*, *c*, *d*). This approach uses its «internal abstract characteristic — determinative degree» as the length of the nucleotide vector [12]. The determinative degree is a numerical characteristic of a nucleotide associated with its ability to determine an amino acid depending on its position in the codon, as well as with the so-called «evolutionary pressure». In addition, the number of hydrogen bonds is taken into account.

It is essential to construct precisely three nucleotide curves corresponding to each position in the codon. This means making three walks for each nucleotide position in the triplet [13]. When the determinative degree is taken into account, a diagram is called a triander [12]. The above-mentioned work shows that a hypothetical number of nucleotides per codon different from three or randomly generated nucleotide sequences does not lead to the appearance of visual structures such as trianders at all. Triander and jsTriander are the first applications for interactively constructing trianders and DNA walks by unequal vectors.

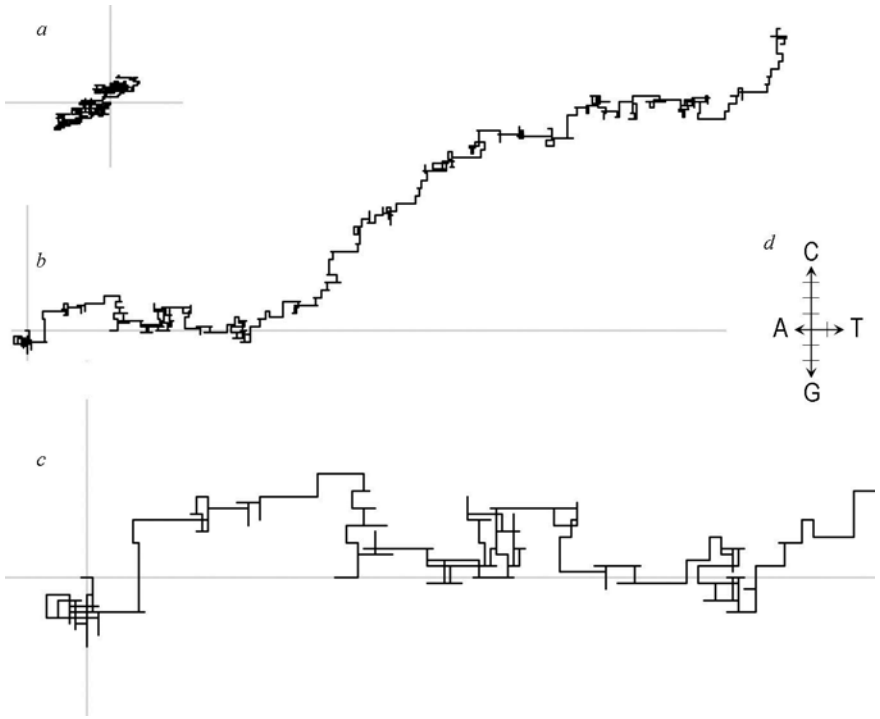


**Fig. 1.** DNA walk diagrams of *mat*K chloroplast gene (AY939873) of *Plagiomnium ellipticum* (Brid.) T.J.Kop. (*a*, *b* and *c* — beginning), and nucleotide vectors (*d*, directions and proportional length). Equal (*a*) and unequal (*b*, *c*) vector lengths, as well as regular (*a*, *b*) and enlarged (*c*) vector lengths were used

Among the possible combinations of directions of nucleotide vectors i. e. vectors that represent nucleotide on the diagram, we chose the following: C — North, G — South, T — East, A — West (Fig. 1, *d*). Since the determinative degree, and therefore the length of the nucleotide vector, are: 4 for C, 3 for G, 2 for T, and 1 for A, then the diagram of randomly selected nucleotides in our case extends in the northeast direction. This direction, like the four main ones, may be indicated on the chart.

Typically, DNA walk diagrams are constructed with one-pixel squares, so in addition to the loss of information due to the passage of the curve along the identical coordinates, losses from the merging of adjacent sections of the curve are added. The problem becomes more acute when the analysis of genomes and chromosomes shifts to the study of individual genes and regulatory sequences. We have implemented the ability to set the length of a unit vector greater than the width of the nucleotide curve (Fig. 1, *c*). This made the diagrams more readable on the one hand, and possible to correctly scale them in this way on the other.

In addition, charts may be offset in horizontal or vertical directions and set to display only a specific part of the sequence. The sequence can be depicted either as a triander (Fig. 2, *a*), the branches of which are exhibited as curves of different thicknesses or colors, or without dividing into branches, called «monander» in the program by analogy (Fig. 1, *b*, *c*). It is also possible to represent a sequence by vectors of equal length. It should be noted that the speed of plotting diagrams is sufficient to observe the animation holding down the buttons of increasing the length or shifting the position in the sequence.

The Triander (Fig. 2, *b*) and jsTriander (Fig. 2, *a*) applications visualize nucleotide sequences from FASTA, GenBank, and regular text files. The opened file is available for viewing and editing. DNA walk diagrams are constructed in the widely used SVG vector graphics format [14] and can be saved into a file.

The most popular method of graphical representation of DNA is a two-dimensional diagram constructed by walking the sequence with vectors of equal length. To build it, one can use separate programs [15] or built-in features of larger projects [16]. Such diagrams are good at exhibiting the structure of large sequences, such as chromosomes or microorganism genomes. However, the loss of visual information due to overlapping parts of the nucleotide curve prevents efficient analysis at the nucleotide level.

Fig. 3 shows that *mat*K gene trianders of three species belonging to two genera have a similar structure, but are different. It is evident that trianders of this gene for species of the same genus are almost identical, but have noticeable differences (highlighted by circles). On the other hand, the *rbc*L (Fig. 4) gene trianders are significantly different from the *mat*K gene trianders (Fig. 2 and 3) but similar to each other.

Today, only the presented software is equipped for constructing trianders and DNA walk diagrams with nucleotide vectors of unequal length. This helps both get a general idea of the sequence and distinguish individual patterns. Since the chosen method of DNA walk diagrams allows one to visually distinguish the nucleotide sequences of different genes, it would be advisable to create software that will generate thumbnails for more pro-

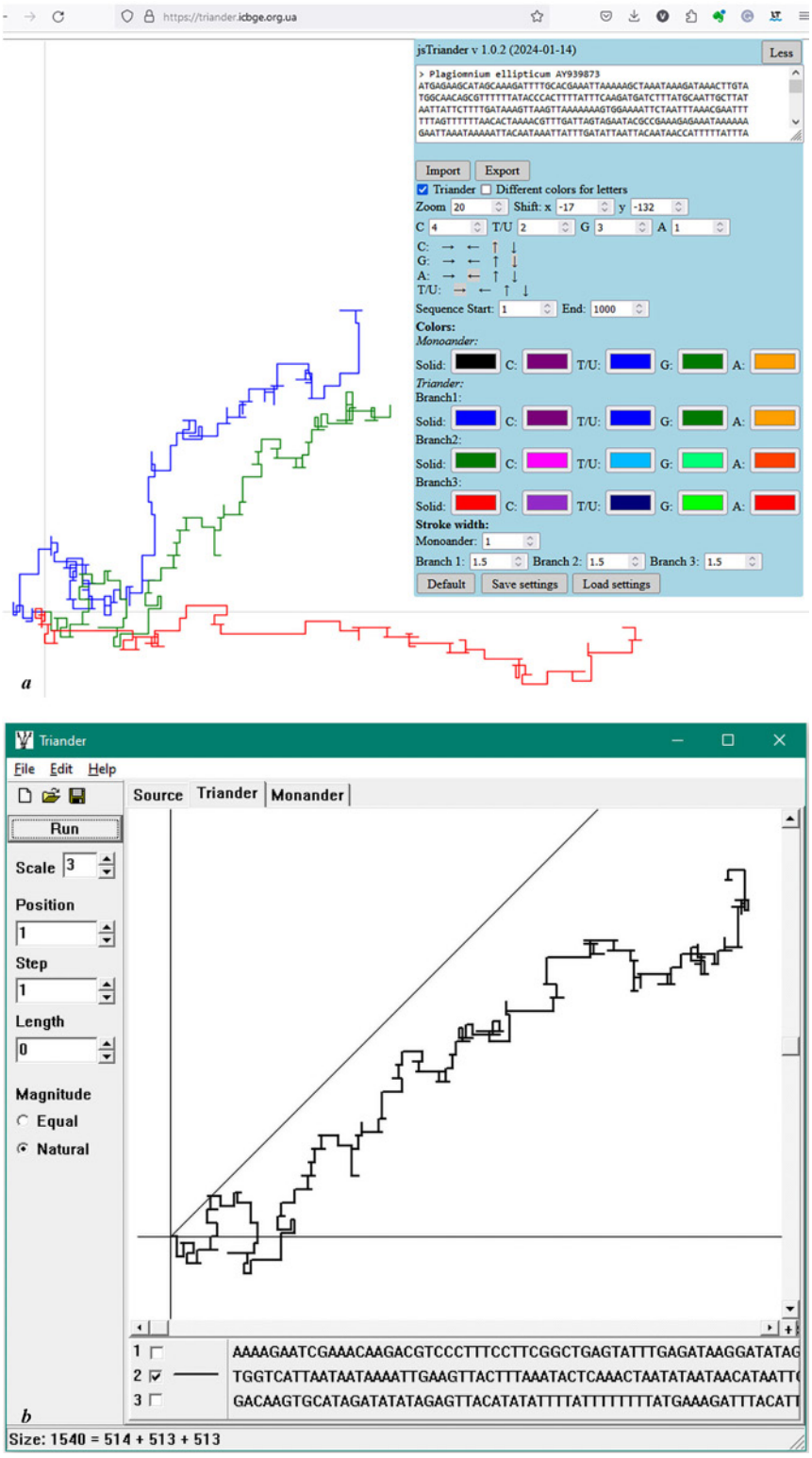


**Fig. 2.** The triander DNA walk diagrams of *mat*K chloroplast gene of *Plagiomnium ellipticum* (Brid.) T.J.Kop. in jsTriander (*a*) and Triander (*b*, 1st and 2nd branches are hidden) applications

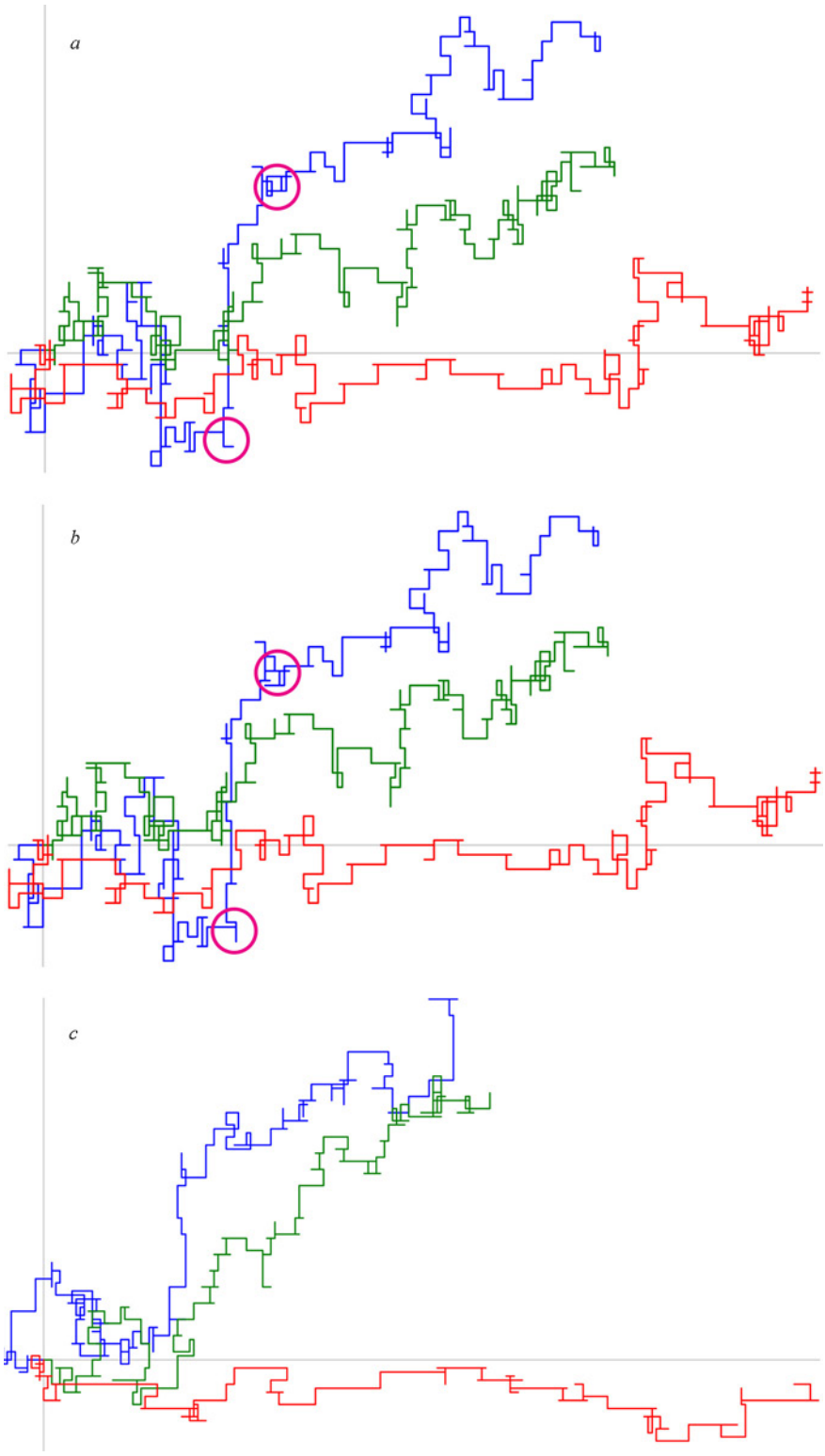


**Fig. 3.** The triander DNA walk diagrams of *mat*K chloroplast gene of *a* — *Sphagnum subsecundum* Nees (AY342155), *b* — *Sphagnum capillifolium* (Ehrh.) Hedw. (DQ185026), and *c* — *Plagiomnium medium* (Bruch & Schimp.) T.J.Kop (AY522574)

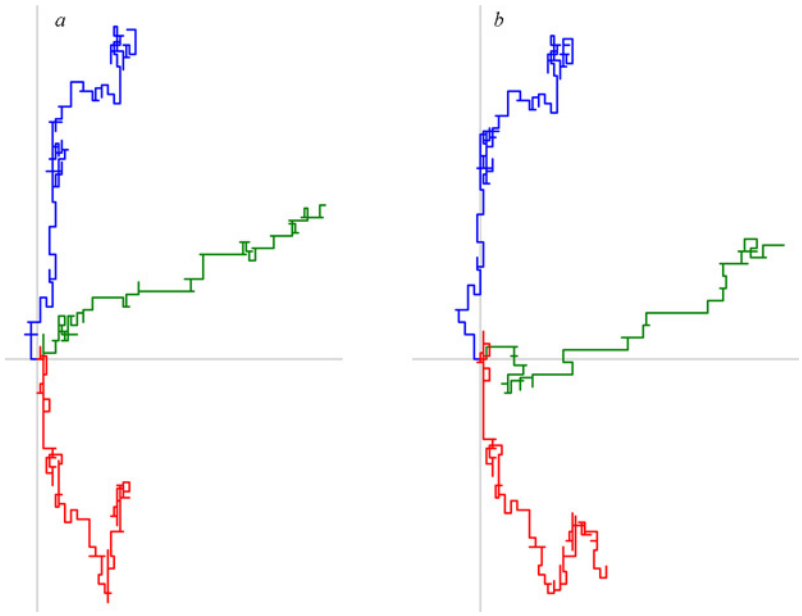


**Fig. 4.** The triander DNA walk diagrams of *rbc*L chloroplast gene of *a* — *Sphagnum subsecundum* Nees (FJ572348) and *b* — *Plagiomnium ellipticum* (Brid.) T.J.Kop. (HM625851)

ductive work in file managers. In addition, it is planned to create software for visual comparison of two or more sequences on the same chart.

Human abilities to distinguish and classify objects from a stream of visual information are well known. Only now, large language models (LLM), which today are called artificial intelligence (AI), have been able to match, and, in some cases, even surpass them. It was reported [17] that the use of DNA walking diagrams accelerates the process of learning LLM by 500 times compared to learning on text versions of the same sequences and allows it to be done on ordinary desktop computers.

Another example, where DNA walk diagrams are built for not analysis by a human, but by a computer program, can be found in cancer genetics. Fractal analysis of DNA walks allows obtaining important information about the potential consequences of mutation of one or other gene [18].

The Triander and jsTriander applications allow one to build several variants of DNA nucleotide walk diagrams. The use of an internal abstract characteristic of a base, called the determinative degree, as the length of a nucleotide vector allows both general visual analysis at the level of chromosomes and genomes and the identification of individual nucleotide patterns.

*REFERENCES*

1. Coordinators, N.R. (2013). Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 41, No. D1, pp. D8-D20. https://doi.org/10.1093/nar/gks1189
2. Nakamura, Y., Gojobori, T. & Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Research, 28, No. 1, pp. 292-292. https://doi.org/10.1093/nar/28.1.292
3. Lazarus Homepage. URL: https://www.lazarus-ide.org/ (last accessed: 09.05.2024).

4. Free Pascal — Advanced open source Pascal compiler for Pascal and Object Pascal — Home Page. URL: https://www.freepascal.org/ (last accessed: 09.05.2024).
5. HTML Standard. URL: https://html.spec.whatwg.org/ (last accessed: 09.05.2024).
6. CSS Snapshot 2023. URL: https://www.w3.org/TR/CSS/ (last accessed: 09.05.2024).
7. Cowin, J.E., Jellis, C.H. & Rickwood, D. (1986). A new method of representing DNA sequences which combines ease of visual analysis with machine readability. Nucleic Acids Res., 14, No. 1, p. 509. https://doi.org/10.1093/nar/14.1.509
8. Hamori, E. & Ruskin, J. (1983). H Curves, A Novel Method of Representation of Nucleotide Series Especially Suited for Long DNA Sequences. J. Biol. Chem., 258, No. 2, pp. 1318-1327. https://doi.org/10.1016/S0021-9258(18)33196-X
9. Gates, M.A. (1985). Simpler DNA sequence representations. Nature, 316, p. 219. https://doi.org/10.1038/316219a0
10. Lobry, J.R. (1999). Genomic landscapes. Microbiology Today, 26, pp. 164-165.
11. Duplij, D. & Duplij, S. (2005). DNA sequence representation by trianders and determinative degree of nucleotides. J. Zhejiang Univ. Sci. B, 6, No. 8, pp. 743-755.
12. Duplij, D. & Duplij, S. (2000). Symmetry analysis of genetic code and determinative degree. Biophysical Bull. Kharkov Univ., 488, pp. 60-70.
13. Cebrat, S. & Dudek, M. (1998). The effect of DNA phase structure on DNA walks. The Eur. Phys. J. B, 3, No. 2, pp. 271-276.
14. Scalable Vector Graphics (SVG) 2. URL: https://www.w3.org/TR/SVG/ (last accessed: 09.05.2024).
15. GenPatterns Home Page. URL: https://math.nist.gov/~FHunt/GenPatterns/ (last accessed: 09.05.2024).
16. Arakawa, K., Tamaki, S., Kono, N., Kido, N., Ikegami, K., Ogawa, R. & Tomita, M. (2009). Genome Projector: zoomable genome map with multiple views. BMC bioinformatics, 10, No. 1, p. 31. https://doi.org/10.1186/1471-2105-10-31
17. Akbari Rokn Abadi, S., Mohammadi, A. & Koohi, S. (2022). WalkIm: Compact image-based encoding for high-performance classification of biological sequences using simple tuning-free CNNs. PLOS ONE, 17, No. 4, e0267106. https://doi.org/10.1371/journal.pone.0267106
18. Hewelt, B., Li, H., Jolly, M.K., Kulkarni, P., Mambetsariev, I. & Salgia, R. (2019). The DNA walk and its demonstration of deterministic chaos—relevance to genomic alterations in lung cancer. Bioinformatics, 35, No. 16, pp. 2738-2748. https://doi.org/10.1093/bioinformatics/bty1021

Received 09.05.2024

## ДІАГРАМА ОБХОДУ ДНК В ЗАСТОСУНКАХ TRIANDER ТА jsTRIANDER

*В.П. Дуплій*[1], *Н.В. Дуплій*[2], *С.А. Дуплій*[3]

[1]Інститут клітинної біології та генетичної інженерії Національної академії наук України
03143 Київ, вул. Академіка Заболотного, 148
e-mail: duplijv@icbge.org.ua
[2]Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, Берестейський проспект, 37, Київ
[3]Центр Інформаційних технологій Мюнстерського університету
48149, Мюнстер, Ронтгенштрассе 7-13, Німеччина

Метою цієї роботи було створення програмного забезпечення, здатного представити нуклеотидні послідовності у зручному для розпізнавання та порівняння вигляді. Для цього було обрано метод обходу ДНК векторами різної довжини з напрямками північ-південь-захід-схід. Кожному нуклеотиду відповідає свій напрямок. Показано, що діаграми типу триандр, представленого трьома гілками, кожна з яких відповідає позиції нуклеотида в кодоні, є схожими для однакових генів різних біологічних видів і можуть відрізнятися для різних генів. Порівняння діаграм дає змогу помітити навіть

мінорні відмінності між послідовностями гена для різних видів одного роду. Вихідний код та скомпільована для операційної системи Windows програма Triander розміщені за адресою https://icbge.org.ua/eng/Triander. Веб-застосунок jsTriander розташовано за адресою https://triander.icbge.org.ua і може використовуватися як онлайн, так і офлайн.

*Ключові слова:* діаграма обходу ДНК, аналіз послідовностей, триандр, програмне забезпечення.

**ORCID**

**VOLODYMYR DUPLIJ** — https://orcid.org/0000-0002-7479-7257

**NATALIA DUPLII** — https://orcid.org/0009-0009-1838-2776

**STEVEN DUPLIJ** — https://orcid.org/0000-0003-1184-6952