



## Assessing Genetic Diversity in Squash Pumpkin (*Cucurbita moschata*) through Computational Analysis of Plastid Genes

Uduak L. Edem\*, Aniefiok N. Osuagwu, Ndem E. Edu, Lasbrey I. Emeagi, Reagan B. Agbor, Nne-Ese N. Nsungu, Ukam U. Uno, Ekerette E. Ekerette

Department of Genetics and Biotechnology, Faculty of Biological Sciences, University of Calabar P.M.B 1115, Calabar, Nigeria

### ARTICLE INFO

#### Article history:

Received 05 July 2024

Revised 31 October 2024

Accepted 11 November 2024

Published online 01 December 2024

**Copyright:** © 2024 Edem *et al.* This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

Grasping the genetic variation of economically significant crops, along with their Wild counterparts that serve as genetic assets, is essential for advancing the development of cultivars and strains capable of adapting to shifting climate conditions. Achieving high crop yields in agriculture is often challenging due to multiple influencing factors, such as cultivar quality, availability of nutrients and water, pathogen infection levels, natural disasters, and soil quality, all of which play a role in plant growth and development. This research evaluated the genetic diversity of 20 accessions of squash pumpkin (*Cucurbita moschata*) by analysing its' plastid genes using bioinformatics mechanism. Twenty *Cucurbita moschata* accessions were sourced from the National Center for Biotechnology Information (NCBI) database. Phylogenetic association, Guanine-cytosine (GC) composition, secondary and three-dimensional protein structure, and domain structure were evaluated. The phylogenetic assessment identified 12 clades in total, with two main clades containing four accessions; XM\_023105445.1, XM\_023105444.1, XM\_023075850.1, and XM\_023075859.1 each showing 100% bootstrap support. Accession XM\_021009681.1 showed the peak Guanine-cytosine composition, whereas XM\_023076698.1 showed the least performance. The analysis of domain structure indicated that accession XM\_021009681.1 contained the highest number of domains, while XM\_023067309.1, XM\_02105444.1, and XM\_0230758951.1, each had only one domain, implying that they could have a more specific role. Variations in secondary and three-dimensional protein structures were observed across the accessions, suggesting potential structural differences that may affect protein robustness or their ability to interact with other molecules.

**Keywords:** *Cucurbita moschata*, Squash pumpkin, Genetic variation, Phylogenetic analysis, Domain structure, Guanine-cytosine composition.

### Introduction

Food insecurity has become an increasingly critical issue due to the growing global population. Despite improvements in global food production over recent decades, more than 800 million individuals continue to experience chronic hunger.<sup>1</sup> Genetic variation contributes to various aspects of food security, such as ensuring food availability, improving access to it, and enhancing its utilisation. Food is undoubtedly the most fundamental of all human necessities. Despite significant efforts to enhance the quality and increase food availability globally, scarcity of food resources persists, mostly in Africa and Asia, especially in the southern zones. Nigeria has faced challenges due to malnutrition, leading to the loss of many lives.<sup>1</sup> The African Food Security Briefs (AFSB) reported that roughly one-third of individuals in Africa's sub-Saharan regions experience malnutrition.<sup>3</sup> Without an adequately nourished and thriving population, promoting enduring developmental strides in Nigeria's economy and Africa at large will remain a distant goal.

\*Corresponding author. Email: [edemlinus@unical.edu.ng](mailto:edemlinus@unical.edu.ng)  
Tel: +2347035493359

**Citation:** Edem UL, Osuagwu AN, Edu NE, Emeagi LI, Agbor RB, Nsungu NN, Uno UU, Ekerette EE. Assessing Genetic Diversity in Squash Pumpkin (*Cucurbita moschata*) through Computational Analysis of Plastid Genes. Trop J Nat Prod Res. 2024; 8(11): 9310 – 9314. <https://doi.org/10.26538/tjnpr/v8i11.47>

Official Journal of Natural Product Research Group, Faculty of Pharmacy, University of Benin, Benin City, Nigeria

Despite Nigeria's rank as the largest economy in Africa since 2014, the country faces an alarming poverty rate.<sup>4</sup> At least 70% of Nigerians earn below a dollar daily, with inadequate access to nutritious food affecting 79% of economically disadvantaged urban households and 71% of remote areas.<sup>3</sup> Rural areas are especially susceptible to poor nutrition, inconsistent food supply, high food prices, poor-quality foods, and, in some cases, total food scarcity. This issue is prominent in Borno, Yobe, Kano, Katsina, and Zamfara States due to several interrelated factors like poverty, limited infrastructure, and frequent insecurity issues that disrupt food supply chains.<sup>5</sup> The genetic diversity found in edible plants offers a promising approach to alleviating hunger and enhancing nutrition and health outcomes.<sup>6</sup> The precision of genetic variation allows for the identification of the most suitable crop varieties for cultivation. From a conceptual viewpoint, variation in genetic traits can be seen as a reflection of the equilibrium linking the creation and extinction of genetic variants. Our environment is rich in healthful crops that are often neglected and one such crop is squash pumpkin. Investigating its' variation in genetic traits could support the stability of resource objectives outlined in the Sustainable Development Goals (SDG) Computational analysis refers to the use of code-based algorithms and applications to process and analyse large datasets.<sup>8</sup> By employing statistical and mathematical methods, researchers can identify significant patterns and derive insights from complex datasets. These insights can then contribute to the creation of accurate forecasts, guide deliberation processes, and foster progress in disciplines biology as biology, economics, sociology, and engineering.<sup>9</sup> In *in-silico* studies, data is inserted into a software tool that subsequently executes necessary arithmetic operations and data analysis procedures to generate outcomes. Such outcomes may subsequently be put to use to draw conclusions, forecast trends, or guide decision-making. Computational analysis serves as a crucial tool across various

disciplines, enabling researchers and analysts to efficiently process vast datasets and uncover trends and connections that might not be observable through manual approaches. This approach often involves the use of genetic codes or DNA chains from a biological entity to examine genetic diversity within a specific living thing.<sup>10</sup> In plant and algal cells, plastid genetic materials are located inside the plastid organelles.<sup>11</sup> These are specialised cell structures that perform a variety of crucial tasks, such as enabling photosynthesis, storing starch, and synthesizing key compounds like proteins, Fatty acids, and Chromophores.<sup>12</sup> Plastid genes are generally divided into two primary categories: plastome genome genes and genes responsible for encoding in the nucleus but targeted to plastids. Plastome is a diminutive, circular DNA molecule responsible for encoding proteins essential for plastid activity. In contrast, genes from the nuclear genome that are plastid-targeted are initially found in the nuclear region of the cell and subsequently transported inside the plastid compartment.<sup>13</sup> Proteins produced by genetic material in plastids play vital roles in various biochemical pathways, including light-dependent synthesis, metabolic pathways of starch, pigment formation, lipids, proteins, and nucleotides.<sup>14</sup> Plastid genetic material serves an important function in ensuring proper operation and upkeep of plastids, which are essential for the progression of flora and marine life. Genetic alterations within these sequences result in a variety of modifications, such as disrupted pigment production, decreased photosynthetic efficiency, and hindered progression.<sup>13</sup> Thus, this work seeks to assess the genetic variation of 20 *Cucurbita moschata* accessions by analysing their plastid genes through bioinformatics approaches. Investigating the genetic material of plasmid through *in-silico* methods is crucial for unraveling the genetic foundations of *Cucurbita moschata*, along with devising innovative approaches to enhance crop yield and long-term viability. A preprint of this research is available at <https://doi.org/10.21203/rs.3.rs-4587542/v1>.

## Materials and Methods

### Obtaining DNA and protein sequences

DNA and protein sequences of squash pumpkin were sourced from the GenBank database. The FASTA format files for plastid gene sequences were downloaded from the NCBI, USA repository. The GenBank variety numbers for the DNA sequences are as follows: XM\_023100969.1, XM\_023100971.1, XM\_023100967.1, XM\_023100968.1, XM\_023100968.1, XM\_023100829.1, XM\_023068156.1, XM\_023070845.1, XM\_023105444.1, XM\_023105445.1, XM\_023076698.1, XM\_023077049.1, XM\_023065942.1, XM\_023067134.1, XM\_023023068071.1, XM\_023075859.1, XM\_023075860.1, XM\_023067310.1, XM\_023067314.1, and XM\_023067315.1.

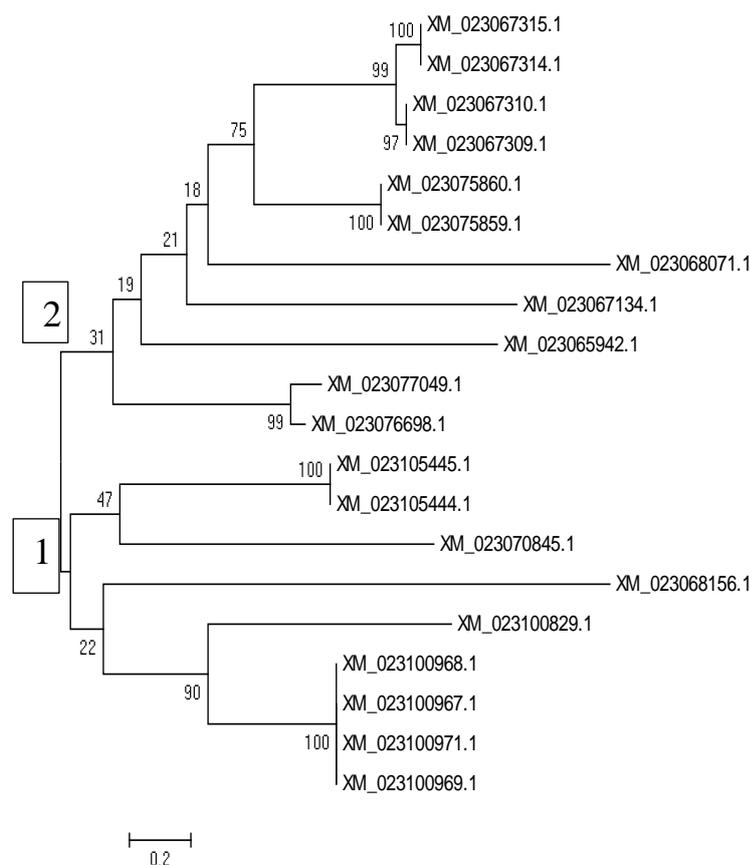
### Analyses of data using *in silico* methods

MEGA 6 software was used to analyse the phylogenetic relationships among all *C. moschata* varieties.<sup>14</sup> Genscan software was used to predict the GC content of the plastid gene, where the nucleotide sequence of plastid amino acids in FASTA format was entered into the designated area to perform the analysis.<sup>14</sup> The variety chosen for evaluation was grounded in their evolutionary classifications. The patterns within the protein chains of the genetic material in plastid from twenty *C. moschata* varieties were estimated with the Simple Modular Architecture Research Tool, accessible at <https://smart.embl-heidelberg.de/>. Phyre2 online software was used to predict the three-dimensional protein structure from the reference protein sequence sourced from the NCBI database.<sup>14</sup> The samples for protein patterns and structure analysis were selected according to their grouping in the phylogenetic tree.

## Results and Discussion

Figure I depicts the phylogenetic analysis of 20 squash pumpkin varieties. The phylogenetic analysis revealed two primary clades for 20 squash pumpkin varieties evaluated. The tree depicted the Evolution-

based relationships among the 20 *C. moschata* varieties, with the primary clade splitting into two distinct sub-groups. among these groups, XM\_023100968.1, XM\_023100968.1, XM\_023100971.1, and XM\_023100969.1 clustered together showing completely bootstrap reliability, thus suggesting a close evolutionary relationship and a likely shared common ancestor.<sup>15</sup> The phylogenetic analysis also demonstrated that XM\_02310544.1 and XM\_023105445.1 formed a distinct clade, supported by 100% bootstrap value. The next prominent clade was further split into two subgroups, with XM\_023075859.1 and XM\_023075860.1 placed in the same group, achieving complete bootstrap reliability. This suggests that both samples are intimately connected and probably have a common origin.<sup>15</sup> In general, the evolutionary analysis demonstrates multiple strongly supported clades among the *C. moschata* accessions, indicating genetic variation within



**Figure 1:** Phylogenetic association within 20 accessions of squash pumpkin based on plastid genome.

the 20 accessions. This observation is consistent with the conclusions of Sanjur et al<sup>16</sup> that the accessions could be categorised within distinct clades classified according to their genetic or evolutionary ties. The closely associated accessions within the sub-groups could further be explored to examine the evolutionary similarities and divergence, that could have an effect on breeding, visible traits, promising applications, and management strategies.<sup>17,18</sup>

The GC content of genetic material is a key characteristic that offers valuable knowledge of its' framework and functionality.<sup>17</sup> The GC values for chosen accessions are displayed in Table 1. Guanine and cytosine values are known to vary among different species and in distinct segments of the same genome. Typically, higher Guanine-cytosine content is linked to greater DNA stability, as the Guanine-cytosine base pairs form more stable hydrogen bonds than the A-T base pairs. Elevated GC content can have a significant impact on gene expression, protein function, and structure, as well as on the evolution of genomes.<sup>18</sup>

**Table 1:** Guanine-cytosine (GC) composition in squash pumpkin.

Accessions	Guanine-cytosine content (%)
XM_021009681	44.86
XM_023068156.1	43.01
XM_023105444.1	41.93
XM_023076698.1	40.94
XM_023075859.1	44.68
XM_023067309.1	43.41
XM_023067310.1	43.47

On the other hand, reduced Guanine-cytosine composition is often related to an increase in the frequency of mutations, genomic uncertainty, and heightened sensitivity to Genetic alterations.<sup>18</sup> Our

findings revealed that the highest Guanine and cytosine composition with a value of 86% was seen in XM\_02100968.1, with the lowest being 40.94%. The variations in GC content, particularly in XM\_023076698, suggest potential differences in the genomic structure and functional properties among the selected *C. moschata* accessions. For instance, XM\_021009681, exhibiting the highest GC content, could exhibit greater genomic stability and possibly more highly conserved genetic sequence compared to XM\_023076698.1 with low Guanine-Cytosine composition. A deeper exploration of these variations could offer better insights into the genetic variation and possible adaptability of these selected accessions to diverse climatic conditions<sup>19</sup>

Table 2 presents the findings for the secondary protein framework of the squash pumpkin. The accession XM\_023076698.1 exhibited the highest alpha helix content at 38.95%, while the lowest was observed in XM\_023105444.1 at 7.92%. The peak proportion of the extended strand, 40.27%, was recorded in XM\_023105444.1, whereas XM\_023067309.1 had the lowest, 17.46%. XM\_02100968 with regards to random coil, displayed the peak proportion at 56.65%, while XM\_023076698.1 exhibited the lowest at 39.06%. The results indicate variations in the secondary protein framework across the distinct accessions of squash pumpkin. These structural variations may pose effective consequences for the proteins, potentially influencing their variability or their connections with other compounds.<sup>20</sup>

**Table 2:** Secondary protein structures in *Cucurbita moschata* accessions.

Accessions	Alpha helix (%)	Extended strand (%)	Random coil (%)
XM_02100968.1	27.95	17.49	54.56
XM_023068156.1	19.60	23.74	56.65
XM_023105444.1	7.92	40.27	51.81
XM_023076698.1	38.95	21.99	39.06
XM_023075859.1	27.87	28.34	43.79
XM_023067309.1	33.33	17.46	49.21
XM_023067310.1	26.92	18.06	55.00

**Table 3:** Domain structure of *Cucurbita moschata* accessions

Accessions	Domain	Start	End
XM_02100968.1	Low complexity	39	54
	Low complexity	83	98
	Low complexity	404	416
	Low complexity	447	466
XM_023068156.1	Low complexity	43	56
	Low complexity	84	103
	Low complexity	187	195
XM_02105444.1	Low complexity	755	768
	Transmembrane region	45	62

	Transmembrane region	77	79
	Transmembrane region	142	164
XM_023076698.1	Transmembrane region	179	198
	Transmembrane region	286	308
	Transmembrane region	435	457
	Transmembrane region	477	494
	Low complexity	557	565
XM_023075859.1	Low complexity	5	21
XM_023067309.1	Transmembrane region	206	228
	Low complexity	4	21
XM_023067310.1	Low complexity	67	80
	Transmembrane region	206	228

Figure 2 depicts the diversity in the three-dimensional protein structures of chosen *C. moschata* accessions. Each accession displayed a distinct folding pattern, with differing proportions of  $\alpha$ -helix,  $\beta$ -sheet, extended regions, and disordered coil. The tertiary protein structures of seven of these accessions display distinct variations in their folding patterns, particularly in the proportions  $\alpha$ -helix,  $\beta$ -sheet, and disordered coil.

The domain structure predicted for the chosen *C. moschata* chains in Table 3 revealed distinct domains of two types; low-complexity and transmembrane regions. The results revealed alterations in the classes and quantity of domains across various sequences. Accession XM\_023076698.1 displayed the highest number of domains, featuring seven transmembrane domains and one low-complexity domain. In comparison, the XM\_02100968.1 accession has four low-complexity domains. This arrangement implies that these sequences could be involved in roles related to transmembrane transport or signaling pathways.<sup>21</sup> XM\_023067309.1, XM\_023075859.1, and XM\_02105444.1 each possess one domain, implying they might have several specialized specific roles. These findings revealed differences in the domain structure of squash pumpkin sequences, potentially indicating functional variations between them.

Yellow ( $\beta$ -sheet); White (extended strands); Blue (random coils); Pink ( $\alpha$ -helix)

The existence of transmembrane and low-complexity domains points to possible involvement in biological pathways including signal transmission and protein-protein connections.<sup>22</sup> Future investigation is essential to substantiate these hypotheses and to elucidate the precise roles of these domains in squash pumpkin.

## Conclusion

The outcomes of this research offer insights into the genetic variation of *Cucurbita moschata*, highlighting the importance of bioinformatics in understanding the genetic makeup of important crops. Results highlighted phylogenetic relationships, GC content variations, domain architecture differences, and structural variations in secondary and tertiary protein structures. These findings enhance knowledge of the genetic foundations of traits in *C. moschata* and may assist in the selection and breeding of varieties with enhanced traits for food security. Further research into the functional significance of observed variations will deepen our understanding of squash pumpkin genetics and aid in the development of improved cultivars to meet the challenges of climate change.

## Conflict of Interest

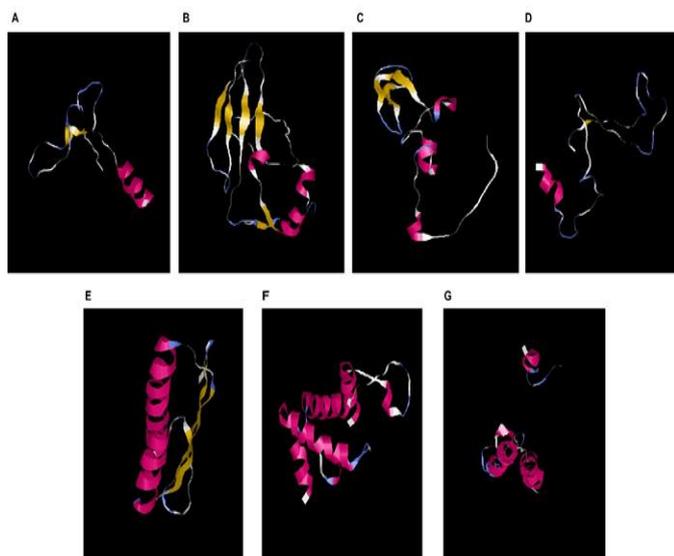
The authors declare no conflicts of interest.

## Authors' Declaration

The authors confirm that the research presented in this article is original and accept full responsibility for any claims related to its content.

## References

1. Wudil AH, Usman M, Rosak-Szyrocka J, Pilař L, Boye M. Reversing years for global food security: A review of the food security situation in Sub-Saharan Africa (SSA). *Int J Environ Res Public Health*. 2022; 19(22):14836. Doi: <https://doi.org/10.3390/ijerph192214836>
2. Ekerette EE, Etukudo OM, Efiokwu JN, Etta HE, Henry II, Ekpo PB, Edu EN, Agbor RB, Edem UL, Ikpeme EV.



**Figure 2:** Tertiary protein structures of various *Cucurbita moschata* accessions: (A) XM\_02100968.1, (B) XM\_023068156.1, (C) XM\_023105444.1, (D) XM\_023076698.1, (E) XM\_02307859.1, (F) XM\_023067309.1, (G) XM\_023067310.1.

- Evaluation of genetic variation in *Oreochromis* tilapia species from South-South Nigeria using mitochondrial DNA hypervariable region. Trop J Nat Prod Res. 2024; 8(9):8527-8536. Doi: <https://doi.org/10.26538/tjnpr/v8i9.41>.
3. Akerele D, Momoh S, Aromolaran AB, Oguntona CRB, Shittu AM. Food insecurity and coping strategies in South-West Nigeria. Food Sec. 2013; 5:407-414. Doi: <https://doi.org/10.1007/s12571-013-0264-x>.
  4. Omorogiwa O, Zivkovic J, Ademoh F. The role of agriculture in the economic development of Nigeria. Eur Sci J. 2014; 10:113-147.
  5. Akinyele IO. Ensuring food and nutrition security in rural Nigeria: An assessment of the challenges, information needs, and analytical capacity (NSSP Working Paper No. 7). International Food Policy Research Institute (IFPRI); 2009.
  6. Edem UL, Osuagwu AN, Udensi OU. Assessment of genetic diversity in *Sphenostylis stenocarpa* (Hosht Ex. A Rich Harms) using morphological markers. Asian J Biol. 2023; 19(4):74-85.
  7. Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. Genet Res Int. 2015; 2015:431487. Doi: <https://doi.org/10.1155/2015/431487>.
  8. Rohman MS, Sulistomo HW, Kusumastuty I, Nugroho DA, Nugraheni NI, Lukitasari M, Alfata FH. In silico discovery of green tea and green coffee bioactive compounds against IGF-1R, PPAR- $\alpha$ , and TLR4 as a therapeutic candidate for metabolic disorder. Trop J Nat Prod Res. 2024; 8(3):6594–6603. Doi: <https://doi.org/10.26538/tjnpr/v8i3.18>.
  9. Kasemsap K. Bioinformatics: Applications and implications. In: Lytras M, Papadopoulou P, editors. Applying Big Data Analytics in Bioinformatics and Medicine. Idea Group Inc. Global; 2018. p. 26-47. Doi: <https://doi.org/10.4018/978-1-5225-2607-0.ch002>.
  10. Oyebamiji AK, Akintelu AS, Mutiu OA, Adeosun IJ, Kaka MO, M T, Semire B. In-silico study on anti-cancer activity of selected alkaloids from *Catharanthus roseus*. Trop J Nat Prod Res. 2021; 5(7):1315–1322. Doi: <https://doi.org/10.26538/tjnpr/v5i7.25>.
  11. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011; 76(3-5):273-297. Doi: <https://doi.org/10.1007/s11103-011-9762-4>.
  12. Choi H, Yi T, Ha SH. Diversity of Plastid Types and Their Interconversions. Front Plant Sci. 2021; 12:692024. Doi: <https://doi.org/10.3389/fpls.2021.692024>
  13. Neuhaus HE, Emes MJ. Non-photosynthetic metabolism in plastids. Annu Rev Plant Biol. 2000; 51(1):111-140.
  14. Edem UL, Osuagwu AN. Evaluation of secondary and tertiary protein structure variation in *rbcL* gene of selected legumes using a computational approach. J Proteomics Bioinformatics. 2023; 16(2):1000642.
  15. Edem UL, Emeagi LI, Osuagwu AN, Edu NE, Ekerette EE, Ogbeche AA, Udensi OU. Assessment of genetic diversity analysis in African yam bean (*Sphenostylis stenocarpa* Hochst Ex.A.Rich Harms) using qualitative and quantitative attributes. J Adv Biol Biotechnol. 2024; 27(10):122-131. Doi: <https://doi.org/10.9734/jabb/2024/v27i101436>.
  16. Sanjur OI, Piperno DR, Andres TC, Wessel-Beaver L. Phylogenetic relationships among domesticated and wild species of Cucurbita (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. Proc Natl Acad Sci U S A. 2002; 99(1):535-540. Doi: <https://doi.org/10.1073/pnas.012577299>.
  17. Bowers JE, Tang H, Burke JM, Paterson AH. GC content of plant genes is linked to past gene duplications. PLoS One. 2022; 17(1):261-748. Doi: <https://doi.org/10.1371/journal.pone.0261748>.
  18. Bernardi G. Isochores and the evolutionary genomics of vertebrates. Gene. 2000; 241(1):3-17. Doi: [https://doi.org/10.1016/S0378-1119\(99\)00485-0](https://doi.org/10.1016/S0378-1119(99)00485-0).
  19. Dufresne F, García-González F, Bell G. Genome size, environment, and genetic diversity in fish. Biol Lett. 2005; 1(4):411-414. Doi: <https://doi.org/10.1098/rsbl.2005.0366>.
  20. Zhou Y, Zhou H. Prediction of protein secondary structure. Proteins Struct Funct Bioinform. 2004; 56(4):517-530. Doi: <https://doi.org/10.1002/prot.2003>.
  21. Apic G, Gough J, Teichmann SA. Domain combinations in archeal, eubacterial, and eukaryotic proteomes. J Mol Biol. 2001; 310(2):311-325. <https://doi.org/10.1006/jmbi.2001.4774>.
  22. Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, Tompa P, Fuxreiter M. Protein Phase Separation: A New Phase in Cell Biology. Trends Cell Biol. 2018; 28(6):420-435. <https://doi.org/10.1016/j.tcb.2018.02.004>.