UDC 631.522:58.083.5

BIOINFORMATICAL ANALYSIS OF MAIZE DWARF MOZAIC VIRUS GENOME FORMAIZE MARKER-ASSISTED SELECTION

T. M. Satarova, *doctor of biological sciences*
Agricultural Steppe zone Institute of the National Academy of Agrarian Sciences of Ukraine
A. I. Vinnikov, *doctor of biological sciences;*
S. V. Khoprichkova
Dniepropetrovsk state university in honour to O. Gonchar

The analysis of genome of maize dwarf mosaic virus (MDMV) is conducted according to computer databases Genome, Nucleotide, Gene, and Protein of bioinformatic server NCBI. It is determined that genome of MDMV includes 9515 nucleotides, but only 8553 of them take place in the encoding of amino acid sequences. MDMV is characterized with economy of genome organization, particularly 2 of 10 genes of this virus are overlapped. In maize the resistance to dwarf mosaic virus infection is controlled by dominant allele of gene Mdm1, which is situated in chromosome 6, in bin 6.1 and is localized between nucleotide coordinates 9491573 and 14931786. Markers of single nucleotide polymorphism (SNP-markers) in maize genome closest to locus mdm1 are determined. The obtained results should be applied to production of maize hybrids resistant to dwarf mosaic virus with methods of MAS-selection.

Keywords: *maize dwarf mosaic virus, maize, nucleotide, amino acid, SNP-marker.*

Viral diseases of maize are possessed of great injuriousness through plant depression. Among them dwarf mosaic disease is especially dangerous. Its causative agent is *Maize dwarf mosaic potyvirus* (MDMV) which belongs to kingdom *Vira*, family *Potyviridae*, genus *Potyvirus*.

The first features of dwarf mosaic disease appear on maize seedlings as small chlorotic spots along leaf ribs. In progress the leaf mosaic broadens, it can be small-or wide-spread, light- or dark-green. In strongly injured plants entire chlorosis or tissue reddening can be observed. If after mosaic appearance temperature decreases and keeps three days below 17 °C, the mosaic transforms into the necrosis [1]. Injured plants retard in their growth, the deformation of tassels and ear aplasia occur, plants lose their yieldness in 4,1–42,1 % [2]. Maize dwarf mosaic virus is transferred by different species of aphids (*Aphididae*) non-persistentively. One of the natural reservoirs for MDMV is *Sorghum galepense,* from where with aphids it comes across to maize [3]. Optimal conditions for this infect agent are high temperature, drought and a lot of weeds [4].The area of expansion of this viral disease covers the southern zone of Europe including Northern Caucasus, Moldova, Georgia, so as Kazakhstan [5]. In Ukraine MDMV is also disseminated, its injuriousness here is considered middle. Due to absence of massive explosions of this disease in Ukraine it is little-studied. Although the dangerous effect of MDMV can become enhanced through importation of this quarantine infection with maize seeds from abroad and susceptibility of native breeding material. The genome of Bulgarian isolate of MDMV had been sequenced first [6], later regional races of MDMV were studied [7 et al.]. Consequently, the production of maize inbreds and hybrids tolerant to MDMV is actual. The successfully attempts of maize genetic modification through RNA interference with inverted repeats of MDMV coat protein gene were undertaken [8]. As an effective approach for maize improvement marker-assisted selection (MAS) is also considered [9]. As the first step of genome-wide assisted breeding the aim of given investigation was the bioinformatical analysis of structure and functional characteristics of MDMV genome, correlative genomic peculiarities of maize and the determination of potential molecular markers of resistance to MDMV.

Materials of the investigation were genome of *Maize dwarf mosaic potyvirus* (Bulgarian isolate) and maize genome (inbred B73). Methods of the investigation included bioinformatical analysis of nucleotide and amino acid sequences according to resources of server NCBI (National

Centre for Biotechnological Information) [10], particularly, computer databases Genome, Nucleotide, Gene and Protein [11–14].

According to [11] *Maize dwarf mosaic potyvirus* is (+) single-stranded RNA virus. It has no DNA stage in its ontogenesis. To data of sequencing the Bulgarian isolate of MDMV in project PRJNA15355 its genome size arranges in 9,515 Kb. The contents of guanine and cytosine in RNA-genome of MDMV are 40,6 %, so the contents of adenine and uracil are 59,4 %. There are ten genes in MDMV genome which together encode the polyprotein (includes 9 proteins) and the PIPO-protein. Table 1 represents coordinates and lengths of coding sequences in MDMV genome for 10 proteins.

1. Coordinates and lengths of coding protein nucleotide sequences of MDMV

Gene number	Coordinates of coding nucleotide sequence	Length of coding nucleotide sequence, nucleotides
Polyprotein		
1	2…697	696
3	2078…3118	1041
4	3119…3319	201
5	3320…5233	1914
6	5234…5392	159
7	5960…6685	726
8	6686…8248	1563
9	8249…9121	873
Protein PIPO		
10	2542…2781	240
Total	2…5392; 5960…9121	8553

Table 1 shows that in MDMV genome 8553 of total 9515 nucleotides take part in amino acid encoding. Remaining 962 nucleotides take no participation in this process. Nucleotide sequence for the protein PIPO is situated within gene 3 of the polyprotein, which codes protease 3. In such a way in genome of MDMV 2 genes of total 10 ones are overlapped. The positions of non-coding islands in MDMV genome are introduced in table 2.

2. Coordinates and lengths of non-coding regions in MDMV genome

Region number	Coordinates of non-coding nucleotide sequences	Length of non-coding nucleotide sequence, nucleotides
1	1	1
2	5393…5959	567
3	9122…9515	394
Total	1, 5393…5959, 9122…9515	962

There are two groups of non-coding nucleotides in MDMV genome. The first one presents nucleotides on the 3'-end (1 nucleotide) and the 5'- end (394 nucleotides). Non-coding nucleotides near the 5'-end contain the sequence of 16 adenine remainders as a possible defence of the 5'-end against affection of exonuclease which is able to destroy RNA of MDMV after introgression into host cells. The second region consists of 567 non-coding nucleotides which are situated in the middle part of MDMV genome. It is interesting that the amount of non-coding nucleotides in this district is divisible to 3. It assumes that this particular region also can also be considered as a coding region for a protein not revealed until now.

The total number of amino acids in the polyprotein and the PIPO-protein counts 3041. Table 3 contains the general description of MDMV proteins concerning the contents of different classes of amino acids in accordance with coding genome sequences.

3. The ratio of classes of amino acids coded by MDMV genome

Classes of amino acids	A part in proteins of MDMV	
	amount	%
Monoaminomonocarboxylic amino acids, without serine and threonine	1031	33,90
Oxymonoaminomonocarboxylic amino acids	382	12,56
Monoaminodicarboxylic amino acids	360	11,84
Amides of monoaminodicarboxylic amino acids	290	9,54
Diaminomonocarboxylic amino acids	368	12,10
Sulphur-containing amino acids	122	4,01
Aromatic amino acids	280	9,21
Heterocyclic amino acids	208	6,84
Total	3041	100

Total protein of MDMV which is presented by the polyprotein and the PIPO-protein above all contains monoaminomonocarboxylic amino acids (glycine, alanine, valine, isoleucine, leucine). The significant fraction of oxymonoaminocarboxylic (serine, threonine) and diaminomonocarboxylic amino acids (lysine, arginine) is revealed. The least part is represented by sulphur-containing amino acids cysteine and methionine. Heterocyclic amino acids also are rare. As for contents of individual amino acids in MDMV proteins the highest amount falls on leucine (8,85 %) and lysine (7,37 %), followed by alanine (6,77 %), valine (6,68 %) and glutamic acid (6,68 %). The least part falls to the share of cysteine (1,51 %) and tryptophan (1,25 %).

4. Coordinates of locus mdm1 on maize genome map IBM2 2008 Neighbors 6

Locus name	Chromosome number	Coordinates
umc1606	6	9491573…9492295
*mdm*1	6	within 9492295…14929916
IDP1701	6	14929916…14931786

According to MaizeDatabase [15] maize genome contains in short arm of chromosome 6 the gene *Mdm*1 which ensures the tolerance to virus of maize dwarf mosaic disease. Dominant allele of this gene encodes resistance, but recessive one – susceptibility to the given virus. The level of resistance due to gene *Mdm*1 may be considerably modified with environmental and ontogenetical factors. Today the position of locus of *Mdm*1 gene is not precisely localized, but the results of genome sequencing of maize inbred B73 allows to identify coordinates of its flanking loci – IDP1701 and umc1606.

The maize genome map IBM2 2008 Neighbors 6 [15] fixes the position of locus *mdm*1 between nucleotides 9492295 and 14929916. The size of locus *mdm*1 puts together 5437621 nucleotides. Table 4 demonstrates coordinates of flanking loci and the island presumably occupied with locus *mdm*1.

The information in table 3 and the total maize genome characteristics from MaizeGD [11] allows concluding that locus *mdm*1 is situated in bin 6. 01 which is delimited with the core marker umc85a.

Today few panels of maize DNA markers of single nucleotide polymorphism of (SNP-markers) are known, among them BDI-III-chip [16] and 1536-chip [17] which are widely used in marker-assisted breeding. Bioinformatical analysis shows that markers in maize genome closely situated to locus *mdm*1 and potentially connected with resistance to maize dwarf mosaic disease are SNP-markers N 232-239, 378 i 379 on BDI-III-chip and markers PZA01509.1 and PHM8909.12 on 1536-chip. The settled potential SNP-markers are recommended for application in maize breeding on resistance to dwarf mosaic virus.

Bibliography

1. *Павленко Л. А.* Вирусная мозаика кукурузы и меры по ее предупреждению: автореф. дис. на соискание ученой степени канд. с.-х. наук: спец. 06.01.09 «Растениеводство» / *Л. А. Павленко*; Укр. НИИ защиты растений. – К., 1983. − 22 с.
2. *Мельничук М. Д.* Фітовірусологія: [навч. посібник] / *Мельничук М. Д.* − К.: Поліграф Консалтинг, 2005. − 200 с.
3. *Землина А. Г.* Тли − переносчики вируса мозаики кукурузы. Вирусы и вирусные болезни растений / *Землина А. Г.* – К.: Наук. думка, 1974. − 203 с.
4. *Назаров И. Х.* Влияние ВКМ на различные сорта кукурузы при искусственном заражении / *И. Х. Назаров, Д. Каримова, А. Х. Вахабов* // Узбекский биол. журнал. − 1991. − № 5. − С. 14–17.
5. *Панарин И. В.* Защита злаковых культур от вирусных болезней / *Панарин И. В.* – М.: Россельхозиздат, 1985. – 77 с.
6. *Kong P.* Complete nucleotide sequence and analysis of putative polyprotein of maize dwarf mosaic virus genomic RNA (Bulgarian isolate) / *P. Kong, H. H. Steinbiss* // Arch. Virol. – 1998. – Vol. 143. – P. 1791–1799.
7. Complete sequence and development of a full-length infectious clone of an Ohio isolate of Maize dwarf mosaic virus (MDMV) / *L. R. Stewart, R. Bouchard, M. G. Redinbaugh, T. Meulia* // Virus Res. – 2012. – Vol. 165, N 2. – P. 219–224.
8. *Zhang Z. Y.* Improvement of resistance to maize dwarf mosaic virus mediated by transgenic RNA interference / *Z. Y. Zang, L. Yang, S. F. Zhou* [et al.] // J. Biotechnol. – 2011. – Vol. 153, 3–4. – P. 181–187.
9. *Marker-assisted selection. Current status and future perspective in crops, livestock, forestry and fish* / *Ed. E. P. Guimaräes* [et al.]. – FAO: Rome, 2007. – 471 p.
10. *National centre for biotechnology information* [Електронний ресурс]: U. S. National Library of Medicine. – Режим доступу: www.ncbi.nih.gov. – Назва з екрану.
11. *Genome* [Електронний ресурс]: National centre for biotechnology information. – Режим доступу: www.ncbi.nih.gov/genome. – Назва з екрану.
12. *Nucleotide* [Електронний ресурс]: National centre for biotechnology information. – Режим доступу : www.ncbi.nih.gov/nucleotide. – Назва з екрану.
13. *Gene* [Електронний ресурс]: National centre for biotechnology information. – Режим доступу: www.ncbi.nih.gov/gene. – Назва з екрану.
14. *Protein* [Електронний ресурс]: National centre for biotechnology information. – Режим доступу : www.ncbi.nih.gov/protein. – Назва з екрану.
15. *Maize genetics database* [Електронний ресурс]. – Режим доступу: www.maizeGDB.org. – Назва з екрану.
16. *Biodiagnostics, Inc.* [Електронний ресурс]. – Режим доступу: www.biodiagnostics.net. – Назва з екрану.
17. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms / *Y. Lu, J. Yan, C. T. Guimaräes* [et al.] // Theor. Appl. Genet. – 2009. – Vol. 120. – P. 93–115.

Анотація

Сатарова Т. М., Вінніков А. І., Хопрічкова С. В. Біоінформаційний аналіз геному вірусу карликової мозаїки кукурудзи для маркер-асоційованої селекції. *Аналіз геному вірусу карликової мозаїки кукурудзи (ВКМК) проведено за даними комп'ютерних баз даних Genome, Nucleotide, Gene, Protein біоінформаційного сервера NCBI. Визначено, що геном ВКМК містить 9515 нуклеотидів, причому в кодуванні амінокислотних послідовностей приймають участь лише 8553 нуклеотиди. ВКМК характеризується економічністю геномної організації, зокрема, 2 з 10 генів цього вірусу є такими, що перекриваються. У кукурудзи стійкість до ураження вірусом карликової мозаїки визначається домінантним алелем гена Mdm1, який знаходиться на хромосомі 6, в біні 6.1, і локалізується між координатами нуклеотидів 9491573 та 14931786. В геномі кукурудзи визначено маркери однонуклеотидного поліморфізму (SNP-маркери), найближчі до локусу mdm1. Отримані результати можуть бути використані при створенні гібридів кукурудзи, стійких до вірусу карликової мозаїки, методами MAS-селекції.*

Ключові слова: *вірус карликовий мозаїки кукурудзи, кукурудза, нуклеотид, амінокислота, SNP-маркер.*