

UDC 577.2:633.15

© 2016

T. Satarova, doctor of biological sciences
Institute of grain crops of NAAS

STRUCTURAL ARCHITECTURE OF GENOME OF CORN

The purpose. Analysis of modern state of researches of genome of corn, in particular structural architecture of kernel DNA and infraspecific polymorphism of the genome caused by emersion transposon, indels, microsatellites, one-nucleotide substitutions. **Conclusions.** The high level of scrutiny of structural architecture of genome of corn enables to identify molecular markers for assessment of polymorphism of selection specimens, to search functional markers of valuable economic attributes and to use them in programs of marker-associate selection.

Key words: genome, genes, transposons, genome polymorphism, microsatellites, SNP, corn.

High level of the development of the selection in Ukraine allows national breeding institutions to create highly heterosis maize hybrids of all classes of maturity, which comply with international requirements to yield, grain quality, adaptability to stressful abiotic and biotic factors. The strategic directions of the development of maize breeding in Ukraine are located in the plane of the use of modern biotechnological methods, in particular, the marker-associated selection for polymorphism assessment, search and application of functional markers [20,32,40]. The rapid improvement of molecular genetics methods for evaluation of genomic polymorphism, the transition from the selection on a phenotype to the selection on a genotype, the increasing prevalence of the screening according to molecular markers, the introduction of the genetic certification and copyright protection for commercial maize genotypes create the need for a detailed awareness in structural and other characteristics of maize genome. The presented overview contains the generalization of the current state of maize genome investigations, in particular, structural organization of nuclear DNA.

According to P. S. Schnable et al. [34], the length of the genome of maize line B73 is $2.3 \cdot 10^9$ bp, however intraspecific genetic polymorphism leads to a certain variability in lengths of different lines and hybrids genomes [38]. The content of GC-pairs to [34] is 46.8 %. Modern maize genome size has been formed during several rounds of genome duplication of an ancestral form, the main of which took place 70 million years ago, and the additional one – 5-12 million years ago. A significant contribution to the lengthening of maize genome was made by transposone elements; in particular, about 3 million years ago the genome of this crop has acquired its final size through the activation of retrotransposons [31].

Maize genome contains a significant amount of repetitive elements of different origin, functional activity, duration of the existence in genome. Their total amount for maize is estimated by G. Haberer et al. [17] in 66% of the genome length, by P. S. Schnable et al. [34] – in 85% for line B73, by J. Messing et al. [24] – only in 58-63%. The main fraction of the repetitive elements in maize genome is consisted by such mobile genetic elements as transposons. A significant portion of the genome is represented by the numerous families of two classes of transposons – retrotransposons and DNA-transposons, irregularly spaced along the nuclear DNA.

Retrotransposons of maize. LTR-retrotransposons are the main type of maize retrotransposons; they contain long terminal repeats on their ends. LTR-retrotransposons according to various estimates take 60-72% length of the nuclear maize genome. Concerning their location they tend to nest associations, that is, they are often embedded one in another. 80% maize LTR-retrotransposons belong to 5 major families – *Opie-Ji*, *Cinful-Zeon*, *Huck*, *Prem1* and *Grande*, which vary in nucleotide sequences of transposon specific domains, of 5'- and 3'-ends, in the specificity of transposition process, ontogenetical, tissue- and organ peculiarities [5,24,25]. The number of LTR-retrotransposons copies per genome for

some families comes to 200 thousand, but there are families that contain even less than 100 elements. Just such, few copied transposons "nest" within other retrotransposons that are located generally in intergenic segments. For maize *nonLTR-retrotransposons* are also discovered, they are similar to LTR-retrotransposons, but have no long terminal repeats. They are represented by family *Cin4*, the members of which vary in length from 1.1 to 6.6 thousand bp [19].

DNA-transposons of maize. DNA-transposons are detected in maize genome considerably less often than retrotransposons [13]. Only about 2% repeats are represented by DNA-transposons. According to [34] DNA-transposons are predominantly located inside the introns, but to [12] the majority of DNA-transposons are embedded in nonmethylated areas near the active genes.

In the genome of maize almost all groups of DNA-transposons are presented. Moreover, just in maize a significant portion of the groups was first identified and marked. *Ac-like elements* in maize are the first transposons which have been identified and described in living beings. The honour of this discovery belongs to Barbara McClintock [23]. From the group of *Ac-like elements* in maize genome autonomous DNA-transposons *Ac* (number of copies is 0-2, the size is 4565 bp), *Bg* (number of copies is not identified, the size is 4869 bp), nonautonomous DNA-transposons *rDT* (number of copies is up to 20, the size is 704 bp), *Ds* (number of copies is greater than 50, the size is up to 30,000 bp) are attended [19,22]. To the group of *En/Spm-like elements* autonomous DNA-transposon *En/Spm* (number of copies is 0-2, the size is 8287 bp), its defective, nonautonomous copy *I/dSpm* (number of copies is greater than 50, the size is 900-8300 bp), undefined about the autonomy and the number of copies DNA-transposon *MRII* of 9000 bp belong [5]. DNA-transposons of *Mu-type* (*Mutator*) and *MULE* (*Mu-like elements*) are also discovered and investigated in the research of maize line with extremely high (100-fold)

capacity to mutating described by D. Robertson [30]. *Helitrons* are a type of DNA-transposons, which are widespread in maize genome; they are situated in the regions of DNA enriched in genes [35]. During the transposition helitrons are capable to capture the adjacent nucleotide sequences, gene fragments, among which specific introns and exons can be appeared. This leads to the appearance of new transcripts, the number of which in maize is estimated at the level of 25% of the total number of genes in the genome [3]. M. Morgante et al. [26] reported that in maize more than 4,000 gene fragments had been borrowed by helitrons. Nonautonomous helitron inserts containing gene fragments are responsible for the mutations leading to the loss of functions of genes *schrunk 2* and *barren stalk 1* [16].

Analysis of the recombinant inbred lines of maize revealed almost 100-fold variation of the recombination frequency along the genome, from 0.08-0.10 to 10.43-11.50 morgantes/Mbp [27]. Meiosis recombination in maize takes place mainly in single copied components of the genome where genes are located, but in the locations of retrotransposons the recombination takes place very rarely or never takes place [5].

Thus, in the process of evolution, artificial hybridization and selection different types of transposons through the mechanism of transposition, capture and transfer of DNA-regions with different functional activity played a significant role in the formation of the structural organization of maize genome, in the occurrence of genomic polymorphism, gene functions and the regulation of their expression.

The polymorphism of maize genome. For breeding practice the intraspecific comparison of genomes, namely the characterization of the genomic (genetic) polymorphism, the determination of its patterns among the gene pool of breeding material and the relationship with the phenotypic diversity are very important. Yu. M. Syvolap et al. [38] determine the genetic polymorphism as the existence in a population of two or more

fragmentary variants of genes (alleles) or nongenic regions of DNA. The selection process leads to the reduction in nucleotide diversity, increased proportion of rare alleles, a wide variation in number of copies of repetitive elements, in single nucleotide polymorphism, indels etc. [18].

The polymorphism of maize genome via transposons. The greatest variation in sizes of plant genomes is caused by differences in the amount of repetitive DNA. 74% differences in the nucleotide sequences between maize and *Sorghum bicolor* are considered to be the result of retrotransposons accumulation during the divergence of these two species [39]. S. Brunner et al. [8] conducted a comparative study of the genomes of maize lines B73 and Mo17, which demonstrated the contribution of different types of transposons in the increase of maize genomic polymorphism. It was found that nuclear DNA of these two lines has a common (collinear) and noncommon nucleotide sequences. The latter ones are present (absent) in only one of the lines. The majority of noncommon sequences are represented by LTR-retrotransposons and other mobile elements. The density of genes in noncommon segments is much lower than in common. In most cases noncommon genes are grouped into clusters and are oriented in the same direction within the clusters. It is believed that the complementation of haplotypes that carry different noncommon genes contributes into the phenomenon of heterosis. Noncommon sequences can participate in flanking actively expressed genes, be activated by different stresses and affect the expression of neighbouring genes through the formation of single chimerical or antisense transcripts [8] and significantly reduce the degree of recombination of retrotransposon fraction [43].

The polymorphism of maize genome via CNVs, PAVs. Nucleotide sequences which have different number of copies for different individuals are denoted as CNVs (copy number variations), while present in one line and absent in another one, noncommon, are denoted as PAVs (presence

and absence variations). The various types of transposons, as already mentioned, can be considered as the causes of CNVs occurrence, as well as the usual duplications, deletions and insertions. A statement of the polymorphism through a number of repetitive segments in a genome does not always permit to give a reason of the appearance and the origin of variation in the number of copies, but can be considered as a comparative structural feature. Genome-wide analysis of lines Mo17 and B73 undertaken as the development of S. Brunner's studies [8] found about 400 tandem duplications, in the other 14 maize lines thousands of CNVs were discovered [4], their presence in 14 teosinte populations was shown as well [37]. This allowed identifying 479 genes with increased and 3410 ones with a reduced number of copies per maize genome. It is established that maize genome regions which rarely recombine, in particular, telomeres, contain more CNVs [4,37]. The comparison of sequences for Mo17 and B73 lines revealed also 1783 PAVs associated with 1270 genes. Analysis of PAVs proves their connection with ancestral evolution and domestication [34]. PAVs are also found in mitochondrial DNA of maize [2,33].

Polymorphism of the maize genome via microsatellites. Using the program MISA (**M**icro**s**atellite identification tool), which allows scanning of microsatellites (SSRs), J. Qu, J. Liu [29] conducted a detailed comparative analysis of the genome of reference line B73, as well as 345 other maize lines. According to their results the genome of maize line B73 contains 179681 microsatellite, which are founded into all 10 chromosomes, in average, 1 microsatellite per every 11,46 Kb. The largest proportion of SSRs (77.25%) falls on intergenic genome sections, while the smallest one (1.86%) drops into the coding sequences in genes. Microsatellite density decreases in the direction of the 5'UTR → 3'UTR → promoter → intron → intergenic sections → coding sequences. The overall average length of SSRs of B73 line is 20,34 bp, and the most common motif is (C/G)_n (24,44%). In general, polymorphism of microsatellites is affected by

the influence of the artificial selection during the breeding process. From the point of view of SSR-polymorphism the genetic variation in maize inbred lines is very high compared to other plant species [9].

Polymorphism of maize genome via single nucleotide changes. The above-discussed polymorphism of maize genome via transposons, duplications, deletions and microsatellites manifests itself as in qualitative changes of DNA regions, so in a variation of their length. In turn, the polymorphism via single nucleotide changes leads only to alterations in the nucleotide sequence, but does not affect the length of DNA. The number of single nucleotide changes in maize is estimated as 1 per 44-75 bp, in average one substitution per 31 bp in noncoding and 124 bp in coding parts of the genome [11,15]. For selected in the Steppe zone of Ukraine maize lines V. V. Borisova et al. [6] have shown that in the studied 384 marker SNP-sites, specially selected as high polymorphic and biallelic, purine nucleotides are dominated, whereas pyrimidine nucleotides are presented in a much less number of marker sites. In the gene pool of maize lines of Ukrainian selection transitions of purine nucleotides A↔G (63,2%) were the most common types of nucleotide changes, while transitions of pyrimidine nucleotides (C↔T) were not found at all. Transversions as a type of nucleotide changes occurred much less frequently than transitions. Among transversions the transversions of cytosine were recorded more often than transversions of thymine. For thymine its replacement by guanine (T↔G) in the national breeding material was not founded at all. A similar distribution of the types of single nucleotide changes is detected also for maize lines in collections of CIMMYT, China and Brazil [21]. Various maize subspecies and types of germplasm vary considerably in the intensity and final volume of nucleotide changes in DNA [14,31,42], that is the basis for the use of SNP-analysis in the breeding process.

Maize genes. Genome of maize B73 line contains 39 454 thousand genes that encode 63 335 thousand protein types [34]. However N. N. Alexander et al. [1] on the basis of cDNA analysis have determined that the number of protein-coding genes in maize is about 50 thousand. The density of genes varies widely, from 0.5 to 10.7 genes per 100 thousand bp [17,36]. Structure of maize genes has a number of features generally specific for plant genes. Really, plant genes are relatively compact, mainly due to small intron size; promoters and other regulatory elements in plant genes are very close to the structural part of a gene [5,41].

By G. Haberer et al. [17] maize genes are very variable in size. In the studied breeding material genes in average were of 4 thousand bp in length and contained 5 exons each, while the longest gene was 59 thousand bp long and had 31 exons. The average intron size in maize is 607 bp [1], according to other data it is less than 150 bp [7], but genes without introns which contain a single long exon are also discovered [10]. The genes of maize, as well of representatives of the whole family *Poaceae*, on the nucleotide composition of codons are divided into two classes. The first class incorporates genes, transcripts of which contain less guanine and cytosine in the third position of a codon, that is generally more typical for dicots. The second class includes genes, transcripts of which have a high guanine and cytosine content in the third position of a codon, as is more specific for herbaceous plants. Precisely genes in the second group much more often lack introns, which allowed assuming an important role of genetic information horizontal transfer in their evolution [10]. In 51% transcripts a stop codon is UGA, in 30% of them it is UAG, but in 19% of transcripts a stop codon is UAA [17]. The promoters of maize genes are rich in sites of binding of transcription factors. By the level of the expression they are divided into 5 groups. The greatest number of TATA-boxes among nucleotide sequences of promoters was found in the

group of the strongest promoters and in genes with a high content of guanine or cytosine in the third position of a codon [1].

Conclusions.

Structural organization of maize genome has a considerable intraspecific variability of quantitative and qualitative traits. This creates the conditions for the use of the results of molecular and genetic analysis in the characterization of maize intraspecific variability and the attraction of the received data for a targeted selection of the best genotypes. The main areas of the application of genomic data in maize breeding is the screening and the application of molecular markers for evaluation of genomic polymorphism in maize breeding samples, the retrieval and the application of functional markers of economic valuable signs in programs of marker-associated selection.

Bibliography

1. Alexander N.N. Insights into corn genes derived from large-scale cDNA sequencing / N.N.Alexander, V.V.Brover, S.Freidin et al./Plant Mol. Biol.–2009.–Vol.69, N 1–2.– P.179–194.
2. Allen J.O. Comparisons among two fertile and three male–sterile mitochondrial genomes of maize / J.O.Allen, C.M.Fauron, P. Minx et al. // Genetics. 2007, vol. 177, no 2, pp. 1173–1192.
3. Barbaqlia, A.M., Klusman, K.M., Higgins J. et al. Gene capture by Helitron transposons reshuffles the transcriptome of maize, Genetics. – 2012. Vol. 190, N 3. – P. 965–975.
4. Belo A. Allelic genome structural variations in maize detected by array comparative genome hybridization / A.Belo, M.K.Beatty, D.S.Hondred et al. // Theor. Appl. Genet. – 2010. – Vol. 120, N 2. – P. 355–367.
5. Bennetzen J.L. Handbook of maize : genetics and genomics / J.L. Bennetzen, S.C.Hake. –Springer, 2009. – 800 p.
6. Borisova V.V. Frequencies of SNP-alleles in maize lines of Ukrainian selection (in Ukrainian) / V.V.Borisova, V.Yu.Cherchel, B.V.Dzubetskij, T.M.Satarova // Tavrijskij naukovij visnik. – 2013. – Issue. 84. – P. 21-25.

7. Bruggmann R. Uneven chromosome contraction and expansion in the maize genome / R.Bruggmann, A. K.Bharti, H.Gundlach et al. // *Genome Res.* – 2006. – Vol. 16. – P.1241–1251.
8. Brunner S. Evolution of DNA sequence nonhomologies among maize inbreds / S.Brunner, K.Fengler, M.Morgante et al. // *Plant Cell.* – 2005. – Vol. 17. – P.343–360.
9. Buckler E.S. Molecular and functional diversity of maize / E.S.Buckler, B.S.Gaut, M.D. McMullen // *Cur. Opin. Plant Biol.* – 2006. – Vol. 9, N 2. – P.172–176.
10. Bull C.T. Interactions between Myxobacteria, plant pathogenic fungi, and biocontrol agents / C.T. Bull, K.G. Shetty, K.V. Subbarao // *Plant Dis.*– 2002. – Vol. 86. – P. 889–896.
11. Ching A. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines / A.Ching, K.S.Caldwell, Z.M. MaurineDolan et al. // *BMC Genet.* – 2002. – Vol. 3. – DOI: 10.1186/1471–2156–3–19.
12. Fedoroff N.V. Isolation of the transposable maize controlling elements Ac and Ds / N.V. Fedoroff, S. Wessler, M. Shure // *Cell.* – 1983. – Vol. 35, N 1. – P. 235–242.
13. Fedoroff N.V. Transposons and genome evolution in plants / N.V. Fedoroff // *Proc. Natl. Acad. Sci. USA.* – 2000. – Vol. 97, N 13. – P. 7002–7007.
14. Frascaroli E. Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs / E. Frascaroli, T. A. Schrag, A.E.Melchinger // *Theor. Appl. Genet.*– 2013–Vol.126,N 1.–P.133–141.
15. Ganal M. W. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome / M. W. Ganal, G. Durstewitz, A. Polley et al. // *PLoS ONE.* – 2011. – Vol. 6, N 12. – DOI: 10.1371/journal.pone.0028334.
16. Gupta S. A novel class of helitron–related transposable elements in maize contain portions of multiple pseudogenes / S. Gupta, A. Gallavotti, G.A. Stryker et al. // *Plant Mol. Biol.* – 2005. – Vol. 57. – P. 115–127.
17. Haberer G. Structure and architecture of the maize genome / G. Haberer, S. Young, A.K. Bharti et al. // *Plant Physiol.* – 2005. – Vol. 139, N 4. – P. 1612–1624.
18. Jiao Y. Genome–wide genetic changes during modern breeding of maize / Y. Jiao, H. Zhao, Ren L. et al. // *Nat. Genet.* – 2012. – Vol. 44, N 7. – P. 812–815.
19. Kunach V.A. Mobile genetic elements and plant genome plasticity (in Ukrainian) / V.A.Kunach. – Kyiv : Logos, 2013. – 288 p.

20. Kytayova S.S. The utilization of marker systems in the selection of maize hybrids with high heterosis (in Ukrainian) / S.S.Kytayova, V.V.Kyrichenko, L.M.Chernobaj // *Selekcija i nasinnictvo*. – 2014. – Issue 105. – P. 23-31.
21. Lu Y. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms, / Y. Lu, J.Yan, C. T. Guimarães et al. // *Theor. Appl. Genet.* – 2009. – Vol. 120. – P. 93–115.
22. Lutova L.A. Genetics of plant development (in Russian) / L.A.Lutova, N.A.Provorov, O.N.Tikhodeev et al. – St.Petersburg. – Nauka, 2000. – 536 p.
23. McClintock B. Controlling elements and the gene / B. McClintock // *Cold Spr. Harb. Symp. Quant. Biol.* –1956. – Vol. 21. – P.197–216.
24. Messing J. Sequence composition and genome organization of maize / J.Messing, A.K.Bharti, W.M.Karlowski et al.//*Proc.Natl.Acad.Sci. USA.*–2004.–Vol.101.– P. 14349–14354.
25. Meyers B.C. Abundance, distribution and transcriptional activity of repetitive elements in the maize genome, / B.C. Meyers, S.V. Tingey, M. Morgante // *Genome Res.* – 2001. – Vol. 11, N 10. – P. 1660–1676.
26. Morgante M. Gene duplication and exon shuffling by helitron like transposons generate intraspecies diversity in maize, / M. Morgante, S. Brunner, G. Rea et al. // *Nat. Genet.* – 2005. –Vol. 37, N 9. – P. 997–1002.
27. Pan Q. Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize,/ Q. Pan, F. Ali, X. Yang et al. // *PLoS ONE*. – 2012. –Vol. 7, N 12. – DOI: 101371/ journal.pone.0052777.
28. Paterson A.H. Ancient polyploidization predating divergence of the cereals and its consequences for comparative genomics, / A.H. Paterson, J.E. Bowers, B.A.Chapman // *Proc. Natl. Acad. Sci. USA.* –2004. – Vol. 101. – P. 9903–9908.
29. Qu J. A genome-wide analysis of simple sequence repeats in maize and the development of polymorphism markers from next-generation sequence data / J. Qu, J. Liu // *BMC Res. Notes*. –2013. – Vol. 6. – DOI: 10.1186/1756–0500–6–403.
30. Robertson D.S. Differential activity of the maize mutator Mu at different loci and in different cell lineages / D.S. Robertson // *Mol. Gen. Gent.* – 1985. – Vol. 200. – P. 9–13.
31. Romay M. C. Comprehensive genotyping of the USA national maize inbred seed bank / M. C. Romay, M. J. Millard, J. C. Glaubitz et al. // *Genome Biol.* – 2013. – Vol. 14, N 6. – DOI: 10.1186/qb–2013–14–6–r55.

32. Satarova T.N. Maize : biotechnological and breeding aspects of haploidy : [monograph] (in Russian) / T.N.Satarova, V.Yu.Cherschel, A.V.Cherenkov. – Dnepropetrovsk : Novaja ideologia, 2013. – 552 p.
33. Saxena R.K. Structural variations in plant genomes / R.K. Saxena, D. Edwards, R.V.Varshney // Brief. Funct. Genomics. –2014. – Vol. 13, N 4. – P. 296–307.
34. Schnable P.S. The B73 maize genome complexity, diversity, and dynamics / P.S. Schnable, D. Ware, R.S. Fulton et al. // Science. – 2009. – Vol. 326. – P. 1112–1115.
35. Slotkin R.K. Transposable elements and the epigenetic regulation of the genome / R.K. Slotkin, R. Martienssen //Nature reviews. Genetics. – 2007. – Vol. 8, N 4. – P. 272–285.
36. Song R. Sequence, regulation, and evolution of the maize 22–kD alpha zein gene family / R. Song, V. Llaca, E. Linton, J. Messing //Genome Res. – 2001. – Vol. 11. – P. 1817–1825.
37. Swanson–Wagner R.A. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor / R.A. Swanson–Wagner, S. R. Eichten, S. Kumari et al. // Genome Res. – 2010. – Vol. 20, N 12. – P. 1689–1699.
38. Syvolap Yu.M. Variability and specificity of the genomes of agricultural plants : [monograph] (in Russian) / Yu.M.Syvolap, N.E.Koschuchova, R.N.Calendar. – Odessa : Astroprint, 2011. – 336 p.
39. Tikhonov A.P. Colinearity and its exceptions in orthologous adh regions of maize and sorghum, / A.P. Tikhonov, P.J. SanMiguel, Y. Makajima et al. //Natl. Acad. Sci. USA. –1999. –Vol. 96, – P.7409–7414.
40. Volkova N.E. Molecular and genetical investigations of maize nuclear genome : [monograph] (in Ukrainian) / N.E.Volkova. – Odessa : Astroprint, 2015. – 120 p.
41. Wendel J. F. Intron size and genome size in plants / J. F. Wendel, R. C. Cronn, I. Alvarez et al. // Mol. Biol. Evol. – 2002. – Vol. 19. – P. 2346–2352.
42. Yan J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers / J. Yan, T. Shah, M. L. Warburton et al. // PLoS ONE. – 2009. –Vol. 4, issue 12. – P. 1–13.
43. Yao H. Molecular characterization of meiotic recombination across the 140–kb multigenic 1– sh2 interval of maize / H.Yao, Q. Zhou, J. Li et al. // Proc. Natl. Acad. Sci. USA. – 2002. –Vol. 99, N 9. – P. 6157–6162.