The mitochondrial plasmids as a new type of mobile genetic elements in higher plants

Gorbenko I.V.*[1], Petrushin I.S.[1,2], Konstantinov Yu.M.[1,2]
[1]Siberian Institute of Plant Physiology and Biochemistry SB RAS, Irkutsk, Russia
[2]Irkutsk State University, Irkutsk, Russia
*e-mail: gorbenko@sifibr.irk.ru

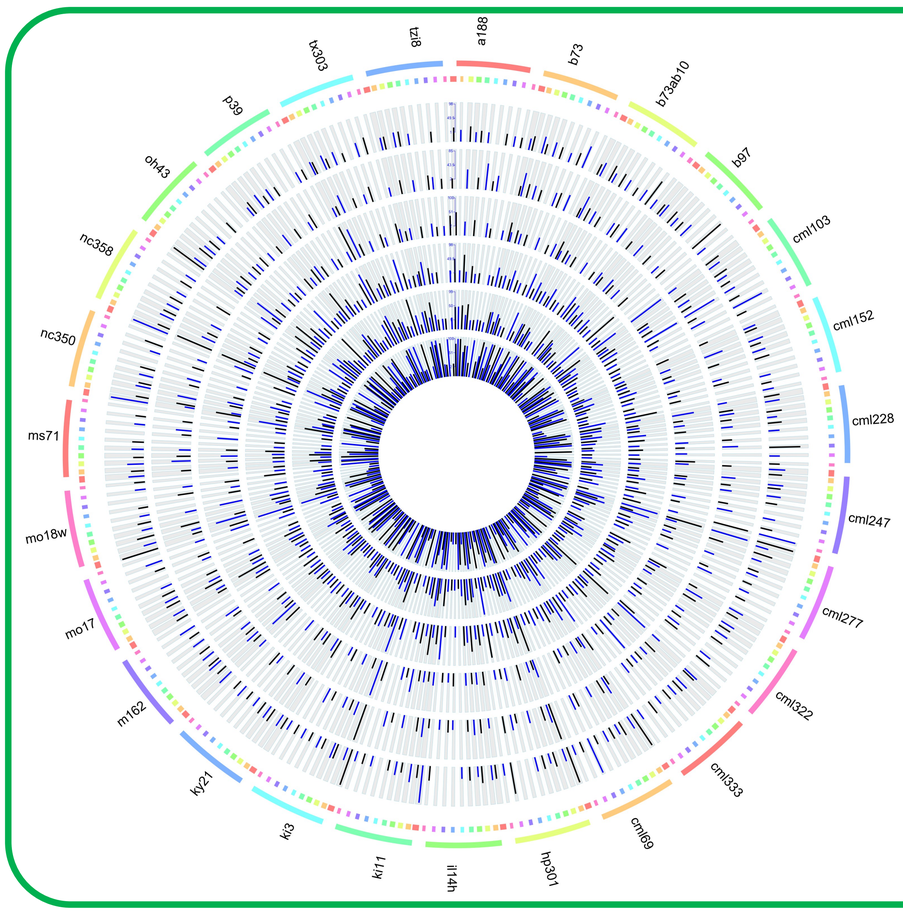


Fig.1.
Discovered mitochondrial plasmid DNA insertions into the nuclear genome of 27 inbred lines of *Zea mays*. Rings from edge to center:
1. Inbred lines
2. Genomes and chromosomes (1-10, red-to-violet)
3. J01426, S2 plasmid insertions
4. M36398, 1.4 kb plasmid insertions
5. NC_001400, 1.9 kb plasmid insertions
6. X02451, S1 plasmid insertions
7. X07041, integrated cms S-1 insertions
8. X13704, S3 plasmid insertions

Motivation:
Mitochondria of many plant species contain, in addition to the main high molecular weight genome, one or several types plasmids with a size ranging from 0.7 to more than 15 kb. These mitochondrial plasmids can exist as autonomous extrachromosomal elements, but their sequences can also be physically integrated into the mitochondrial genome. They are not known to encode an integrase function but instead undergo recombination involving repeated sequences that are shared between plasmids and the mitogenome (Warren et al. 2016; Brown and Zhang 1995; Leaver and Gray 1982). Recently, Warren et al (2016) showed that linear plasmids influence the rate of sequence evolution in plant mitochondrial genomes. However, to date understanding of the intraspecies distribution and possible evolutionary role of plasmids *per se* and plasmid-derived sequences in the whole genetic system of *Zea mays* involving nuclear, mitochondrial and chloroplast genomes remains limited. In the present study we address the following questions: 1) The search of plasmid-derived sequences for 5 mitochondrial plasmids in the nuclear genomes of 27 inbred maize lines. 2) The context analysis of structural and functional features of plasmids and plasmid-derived sequences in nuclear genomes of maize lines.3) Based on (i) the S3 plasmid full-sized insertions existence in the nuclear genome of all investigated maize lines and (ii) CRISPR-like loci existence in this plasmid and plasmid-derived sequences of nuclear genome we suggest hypothesis on the existence of adaptive immunity in maize similar to CRISPR-Cas immunity in prokaryotes.

Results:
Genome-wide search of plasmid-derived sequences from 5 mitochondrial plasmids (1.4 kb, 1.9 kb, 2.3 kb, 5.4. kb, 6.4 kb) in the nuclear genome of 27 inbred lines of *Zea mays* was done. It was shown that all nuclear chromosomes of investigated lines contain numerous insertions of all plasmids. The insertions have different sizes from 3% until 100 % of the whole plasmid length. These results could be presumably explained by the events of intracellular plasmid DNA transfer from mitochondria to nucleus which took place during this plant species evolution. Therefore the identification of full-length insertions of mitochondrial plasmids in the nuclear genome of maize allows to consider them as one of the previously unknown type of mobile genetic elements. The context analysis of plasmid sequences showed that both linear and circular maize plasmid may contain unusual structural elements. In circular 1.9. kb plasmid and linear 2.3 kb plasmid, we were able to detect CRISPR-like loci, which are usually characteristic element of the prokaryotic type immunity system. Of greatest interest is the 2.3 kb linear plasmid, in which both terminal inverted repeats, typical in their structure, contain 2 canonical CRISPR cassettes.

Conclusions:
The mitochondrial plasmid-derived sequences of different length (including full-sized insertions) were identified in the nuclear genome of 27 inbred lines of *Zea mays*. Based on these results, the mitochondrial plasmids may be qualified as new type of mobile genetic elements.

The CRISPR-like elements in the sequences of two mitochondrial plasmids (1.9 kb and 2.3 kb) were found. It is suggested that maize and presumably other plant species may contain the system of adaptive immunity similar to prokaryotic CRISPR-Cas to protect from foreign DNA.

Warren, J. M., Simmons, M. P., Wu, Z., & Sloan, D. B. (2016). Linear Plasmids and the Rate of Sequence Evolution in Plant Mitochondrial Genomes. *Genome biology and evolution*, *8*(2), 364–374. https://doi.org/10.1093/gbe/evw003

Zhang, M., & Brown, G. G. (1993). Structure of the maize mitochondrial replicon RNA b and its relationship with other autonomously replicating RNA species. *Journal of molecular biology*, *230*(3), 757–765. https://doi.org/10.1006/jmbi.1993.1198

Leaver, C. J., & Gray, M. W. (1982). Mitochondrial genome organization and expression in higher plants. *Annual Review of Plant Physiology*, *33*(1), 373-402.

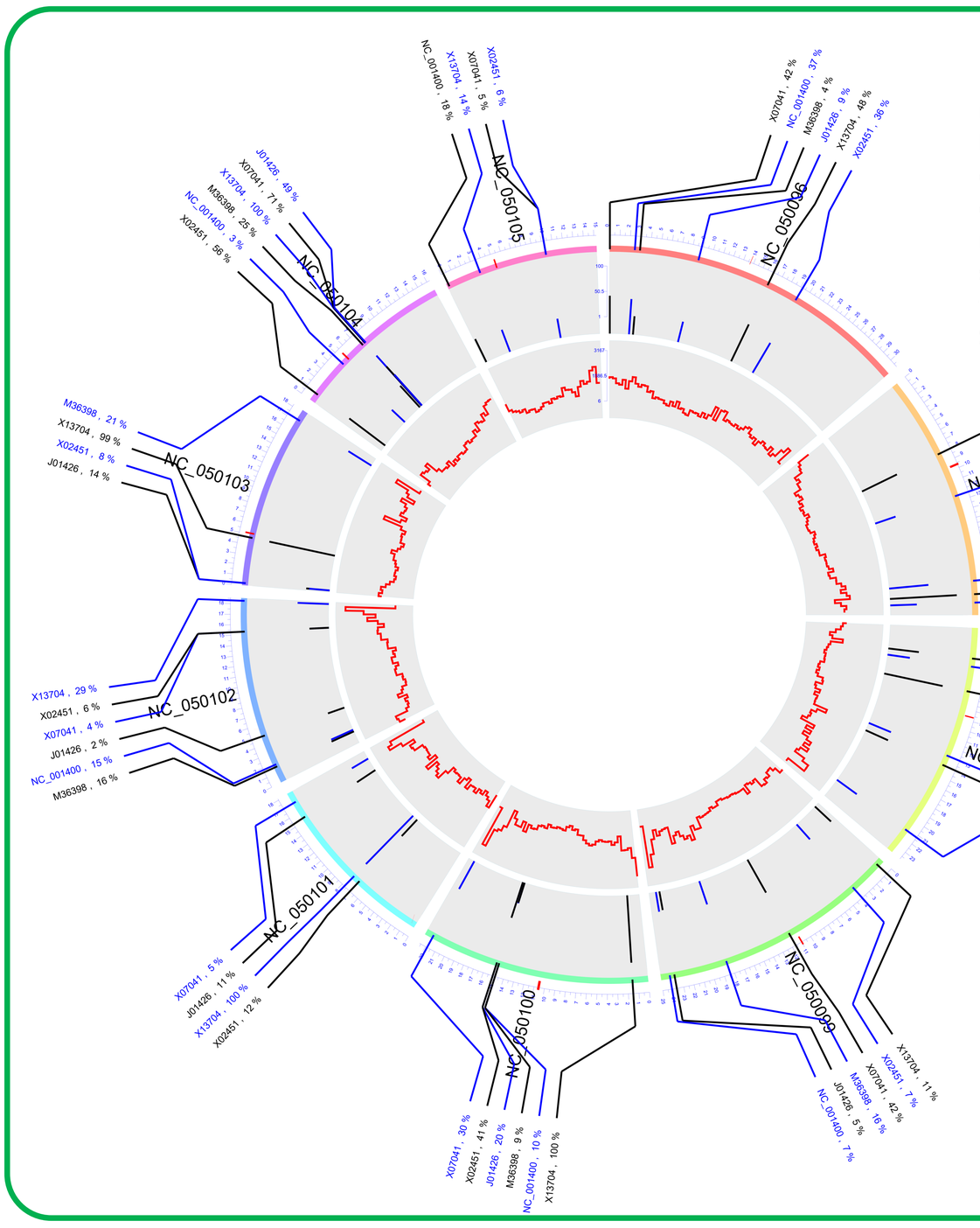


Fig.2.
Insertions of mt-plasmid DNA into the nuclear genome of B73 *Z.mays* inbred line in detail. Only the maximal homology rate variants for each plasmid per chromosome are shown. The red lines represents centromere length and position. The inner ring represent CDS feature density within chromosomes. Scale division = 10 Mb.

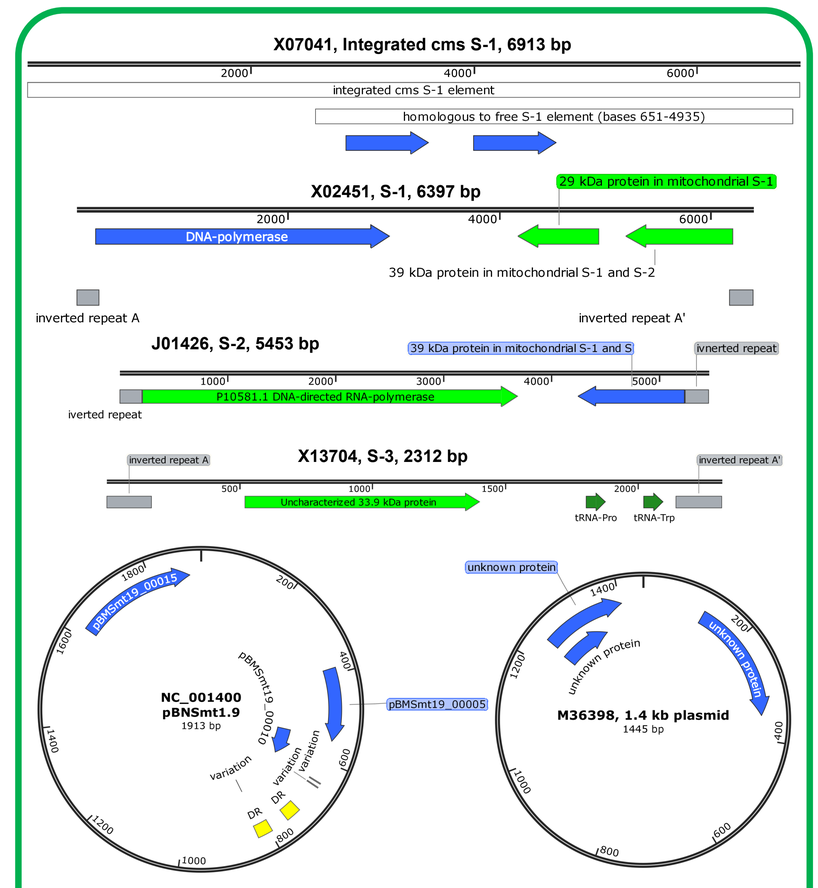


Fig.4. Graphic representation of *Zea mays* plasmids features

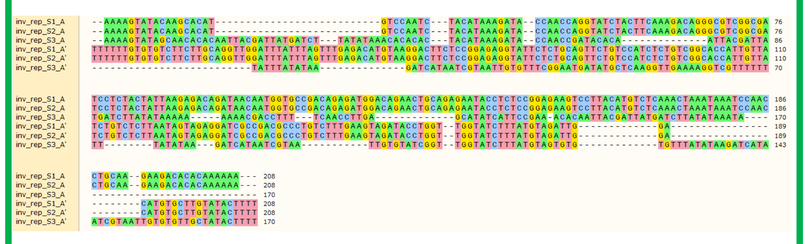


Fig.5. Alignment of *Zea mays* S1, S2 and S3 plasmids inverted repeats

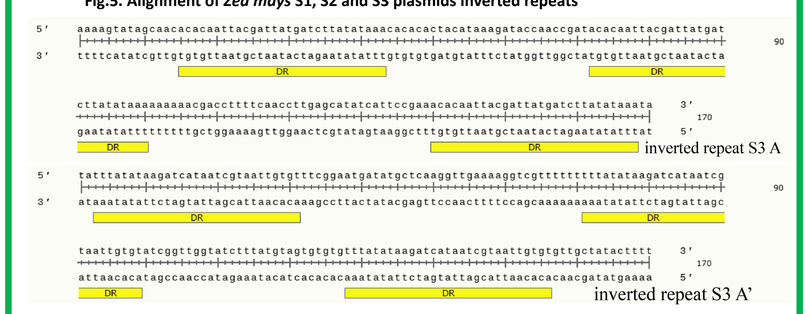
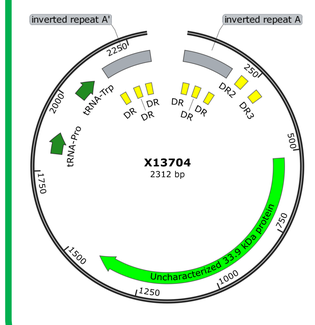


Fig.6. Crispr direct repeats in S3 plasmid inverted repeats and their RNA secondary structure



	DIRECT REPEATS:	SPACERS:	
15	**ACACAATTACGATTATGATCTTATATAAA**	CACACACTACATAAAGATACCAACCGAT	71
72	**ACACAATTACGATTATGATCTTATATAAA**	AAAAAACGACCTTTTCAACCTTGAGCATATCATTCCGAA	139
140	**ACACAATTACGATTATGATCTTATATAAA**	TACACACTACCTATTGAGCATATTCAACACATACCGAT	207
208	**ACACAATTACTATTATGATCTTATATATA**	AAAAAACTACCTTCTGAAATCCAACCACCACATACCGAA	275
276	**ACACAATTACTATTATGATCTTATATAAA**	304	
2145	**TTTATATAAGATCATAATCGTAATTGTGT**	TTCGGAATGATATGCTCAAGGTTGAAAAGGTCGTTTTTT	2212
2213	**TTTATATAAGATCATAATCGTAATTGTGT**	ATCGGTTGGTATCTTTATGTAGTGTGTG	2269
2270	**TTTATATAAGATCATAATCGTAATTGTGT**	2298	

Fig.7. Graphic representation of CRISPR DRs in S3 plasmid and DR and spacers sequences

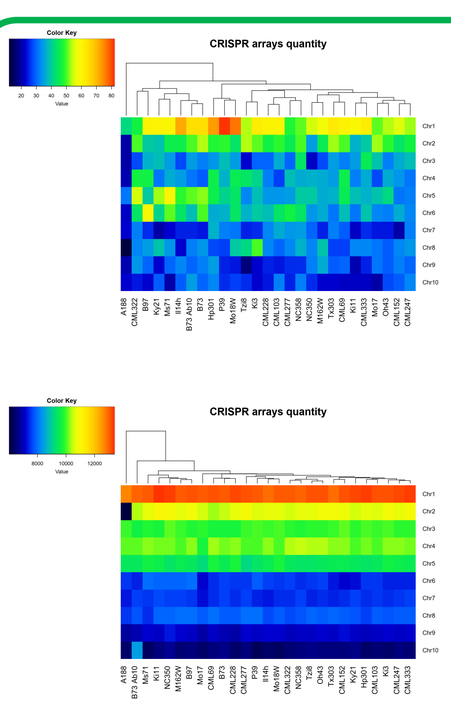


Fig.3.
Heatmaps representing relationships between maize inbred lines based on CRISPR arrays quantity — the upper one for arrays detected with CRT software (Bland et al., 2007) default settings, lower one— for arrays possessing at least one spacer.

```
Spacer:          CACACACTACATAAAGATACCAACCGAT
Query       4    ACACTACATAAAGATACCAACCGAT      28
MT460518    90336 ACACTACATAAAGATACC-ACCGAT     90359
MT448617    89151 ACACTACATAAAGATAC             89167
MK892812    28029        ATAAAGATACCAACCGA      28013

Spacer:             ATCGGTTGGTATCTTTATGTAGTGTG
Query       1       ATCGGTTGGTATCTTTATGTAGTGT      25
MT460518    90359   ATCGGT-GGTATCTTTATGTAGTGT      90336
MT448617    89167         GTATCTTTATGTAGTGT        89151
MK892812    28013   TCGGTTGGTATCTTTAT              28029

Spacer:      TTCGGAATGATATGCTCAAGGTTGAAAAGGTCGTTTTTT
Query       7       ATGATATGCTCAAGGTTGAAAAGGTCGT   34
MT460518    15497   ATGACATGGTCAAGTTTGAAAAGGTCGT   15470
LR798270    5932    ATGACATGGTCAAGTTTGAAAAGGTCGT   5905
LR796227    27700          TCAAGGTTGAAAAGGTCG      27717
MF403007

Spacer:       AAAAAACGACCTTTTCAACCTTGAGCATATCATTCCGAA
Query       6       ACGACCTTTTCAACCTTGAGCATATCAT   33
LR798270    15470   ACGACCTTTTCAAACTTGACCATGTCAT   15497
LR796227    5905    ACGACCTTTTCAAACTTGACCATGTCAT   5932
MF403007    27717   CGACCTTTTCAACCTTGA             27700

Spacer:        TACACACTACCTTATTGAGCATATTCAACACATACCGAT
Query       5       CACTACCTTATTGAGCATATTCAACACATACCGAT   39
LC433738    4646    CACTACCTTATTGACCAAAATCAACA             4621
NC_024792   143125                TATTCAACACATACCGAT    143108
KC821606    4961          ACCTTCTTGAGCATATTGAACAC         4939

Spacer:          AAAAAACTACCTTCTGAAATCCAACCACCACATACCGAA
Query       1       AAAAAACTACCTTCTGAAATCCAACCACCACATAC   35
MT002873    8220    AAAAAACTATCTTCTGAAATC                 8240
MN038176    46076                AATCCAACCACCACATAC    46059
MF684137    1111                 AATCCAACCACCACATAC    1128

Spacer:    TAAGGAGTTGCTCTCTCTACTATTTCCTACTAATGAAAACAAGTA
Query       9       TGCTCTCTCTACTATTTCCTACTAATGAAAACAAGTA 45
MN693155    3110          TCTCTACTATTTCCTACTAA          3129
KT151959    78876         TCTCT-CTAGTTCCTACCAAAGAAAACAAGTA 78846
DQ155197    314     TGCT-TCTCTACTATTTCCTAATAAT          338
```

MT460518 - Vibrio phage Direpillow8
MT448617 - Vibrio phage River4
MK892812 - Prokaryotic dsDNA virus sp. isolate...
LR798270 - uncultured Caudovirales phage
LR796227 - uncultured Caudovirales phage
MF403007 - Agrobacterium phage Atu_ph04
LC433738 - Peanut mottle virus KALRO2-23 genomic RNA
NC_024792- Bacillus phage Bobb
KC821606 - Cellulophaga phage phi12:2
MT002873 - Tenacibaculum phage JQ
MN038176 - Bacillus phage Phireball
MF684137 - HIV-1 isolate 12468 from China nonfunctional
MN693155 - Marine virus AFVG_25M94
KT151959 - Brevibacillus phage Sundance
DQ155197 - HIV-1 isolate MYDH010 from Kenya ...

Fig.8. Alignment of found CRISPR spacers with viral genomes

Acknowledgements:
We are thankful to Prof. Daniel Sloan, Dr. Igor Deineko and Prof. Ian Max Møller for the interest to the study.