Вавиловский журнал генетики и селекции. 2019;23(7):809-816
DOI 10.18699/VJ19.554

The Fifth International Scientific Conference PlantGen2019

Detection of CRISPR cassettes and *cas* genes in the *Arabidopsis thaliana* genome

Yu.M. Konstantinov[1, 2] ✉, I.S. Petrushin[2]

[1] Siberian Institute of Plant Physiology and Biochemistry, SB RAS, Irkutsk, Russia
[2] Irkutsk State University, Irkutsk, Russia
✉ e-mail: yukon@sifibr.irk.ru

The state of the art in the evolution of plant viruses allows the genetic foundations of antiviral immunity in higher (including the most important crops) plants to be categorized as one of the most pressing issues of genetics and selection. According to the endosymbiotic theory, mitochondria descended from alphaproteobacteria that had been absorbed but not degraded by the host cell. The discovery of CRISPR-Cas systems (clustered regularly interspaced short palindromic repeats (CRISPR)-associated proteins), which implement the adaptive immunity function in prokaryotes, raises the question whether such a mechanism of antiviral protection could be caught up by evolution and used by representatives of eukaryotes (in particular, plants). The purpose of this work was to analyze the complete sequences of nuclear, mitochondrial, and chloroplast genomes of *Arabidopsis thaliana* in order to search for genetic elements similar to those in CRISPR-Cas systems of bacteria and archaea. As a result, *in silico* methods helped us to detect a locus of regularly intermittent short direct repeats in the mitochondrial genome of *A. thaliana* ecotypes. The structure of this locus corresponds to the CRISPR locus of the prokaryotic adaptive antiviral immune system. The probable connection between the locus found in the mitochondrial genome of the higher plant and the function of adaptive immunity is indicated by a similarity between the spacer sequences in the CRISPR cassette found and the genome of *Cauliflower mosaic virus* affecting Arabidopsis plants. Sequences of repeats and spacers of CRISPR cassettes in Arabidopsis C24 and Ler lines are perfectly identical. However, the locations of the CRISPR locus in the mitochondrial genomes of these lines differ significantly. The CRISPR cassette in the Col-0 line was found to be completely broken as a result of four deletions and one insertion. Although *cas* genes were not detected in the mitochondrial genome of the studied Arabidopsis ecotypes, their presence was detected in the nuclear genome. Both *cas* genes and numerous CRISPR cassettes were found on all the five chromosomes in the nuclear genome of the Col-0 ecotype. The results suggest the existence of a system of adaptive immunity in plants, which is similar to the CRISPR immunity of bacteria and archaea.
Key words: *Arabidopsis thaliana*; ecotypes; mitochondrial genome; nuclear genome; CRISPR cassette; *cas* genes; homology of CRISPR spacers; plant virus genome; adaptive immunity; RNA interference.

For citation: Konstantinov Yu.M., Petrushin I.S. Detection of CRISPR cassettes and *cas* genes in the *Arabidopsis thaliana* genome. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2019;23(7): 809-816. DOI 10.18699/VJ19.554

Обнаружение CRISPR-кассет и генов *cas* в геноме *Arabidopsis thaliana*

Ю.М. Константинов[1, 2] ✉, И.С. Петрушин[2]

[1] Сибирский институт физиологии и биохимии растений Сибирского отделения Российской академии наук, Иркутск, Россия
[2] Иркутский государственный университет, Иркутск, Россия
✉ e-mail: yukon@sifibr.irk.ru

Современный уровень знаний в области эволюции растительных вирусов позволяет отнести проблему генетических основ противовирусного иммунитета высших растений (в том числе важнейших сельскохозяйственных культур) к разряду наиболее актуальных проблем генетики и селекции. В соответствии с эндосимбиотической теорией принято считать, что митохондрии произошли от альфа-протеобактерий, которые были поглощены, но не подвергнуты деструкции клеткой-хозяином. В связи с открытием у прокариот CRISPR-Cas (clustered regularly interspaced short palindromic repeats – CRISPR-associated proteins) систем, выполняющих функцию адаптивного иммунитета, возникает вопрос, мог ли подобный механизм противовирусной защиты быть подхвачен эволюцией и использован представителями эукариот, например растениями. Задачей настоящей работы был анализ полных последовательностей ядерного, митохондриального и хлоропластного геномов *Arabidopsis thaliana* с целью поиска генетических элементов, сходных с таковыми в CRISPR-Cas системах у бактерий и архей. В результате методами *in silico* в митохондриальном геноме экотипов *A. thaliana* обнаружен локус регулярно перемежающихся коротких прямых повторов, соответствующий по своей организации CRISPR-локусу адаптивного CRISPR-Cas иммунитета прокариот. На вероятную связь обнаруженного в митохондриальном геноме высшего растения локуса с функцией адаптивного иммунитета

указывает наличие у спейсерных последовательностей в составе найденной CRISPR-кассеты гомологии с геномом вируса мозаики цветной капусты, поражающего растения арабидопсиса. У линий арабидопсиса C24 и Ler последовательности повторов и спейсеров CRISPR-кассеты полностью идентичны. В то же время локализация самого CRISPR-локуса в митохондриальном геноме этих линий существенно различается. Установлено, что у линии Col-0 в результате четырех делеций и одной инсерции CRISPR-кассета полностью нарушена. Хотя гены *cas* в митохондриальном геноме исследуемых экотипов арабидопсиса не были найдены, установлено их наличие в ядерном геноме. В ядерном геноме экотипа Col-0 на всех пяти хромосомах обнаружены гены *cas* и многочисленные CRISPR-кассеты. Полученные результаты позволяют предположить существование у растений системы адаптивного иммунитета, аналогичного CRISPR-иммунитету бактерий и архей.

Ключевые слова: *Arabidopsis thaliana*; экотипы; митохондриальный геном; ядерный геном; CRISPR-кассета; гены *cas*; гомология CRISPR-спейсеров; геном растительного вируса; адаптивный иммунитет; РНК-интерференция.

Introduction

The acquisition of alphaproteobacteria (which subsequently gave rise to mitochondria) as endosymbionts by the archaeal host is now unquestionably accepted to be one of the most important events in the nascence of the eukaryotic cell (Archibald, 2015). In recent years, methods of phylogenomics provided fundamentally new data demonstrating the possibility of several evolutionary scenarios for the genesis of the eukaryotic cell, including "late" or "early" acquisition of mitochondria by the host cell (Poole, Gribaldos, 2014; Pittis, Gabaldon, 2016). The discovery of the CRISPR-Cas adaptive immunity system based on the phenomenon of RNA interference in a significant percentage of bacterial and archaeal species (Jansen et al., 2002; Mojica et al., 2005; Makarova et al., 2006; Barrangou et al., 2007; Lander, 2016) poses the question whether such a protective system may exist in eukaryotic mitochondria, organelles that have an obvious evolutionary relationship with their bacterial ancestors. In this regard, mitochondria of higher plants, which have an extremely large genome compared to the genomes of animals and yeast, are of particular interest.

The mitochondrial genome of plants is also characterized by unusual dynamism, which manifests itself as a high recombination rate caused by repetitive sequences (Gualberto, Newton, 2017). The recombination activity results in the formation of a set of subgenomic forms and high genomic variability even within the same species. Such changes in the genomic structure lead to the rapid evolution of the plant mitochondrial genome. Moreover, the mitochondrial genome of higher plants is tightly involved in horizontal gene transfer processes, where it can act as both a donor and an acceptor of the gene (Kleine et al., 2009; Zhao et al., 2018).

Another important feature of the mitochondrial genome of higher plants is the presence of species-specific sets of linear and circular plasmids in these organelles of many plant species studied in this regard. The composition of the sets within the species can vary significantly (for example, in fertile and sterile forms) (Esser et al., 1986; Thomas, 1986). The origin of mitochondrial plasmids is still unknown. Double-stranded plasmids are believed to be introduced into the cells of higher plants by a symbiotic or pathogenic pathway (Douce, Neuburger, 1989). This hypothesis is supported by the fact that mitochondrial linear plasmids are associated with a protein at their 5′-ends that resembles the structure of some viral nucleic acids (Douce, Neuburger, 1989). Moreover, the detection of genes in linear plasmids S1 and S2 of maize mitochondria encoding viral-type nucleic exchange proteins speaks for their probable viral origin (Kuzmin, Levchenko, 1987; Kuzmin et al., 1988). In recent years, a significant progress has been made in the study of mitoviruses, viruses with the simplest RNA genome that specifically infect fungal mitochondria (Shahi et al., 2019). However, there is also evidence for the existence of plant mitoviruses, which are believed to have arisen as a result of horizontal transfer events of the corresponding genes from plant-infecting fungi (Marienfeld et al., 1997; Bruenn et al., 2015; Nibert et al., 2018). Thus, if to compare bacteria and plant mitochondria, it can be said that the latter, like prokaryotes, also badly needed the protection against infectious nucleic acids of viral and/or plasmid origin during evolution.

Nevertheless, data on the existence of a similar mechanism of protection against pathogenic DNA among representatives of eukaryotes have not been obtained until recently, with the exception of the single detection of a typical CRISPR locus on the mitochondrial plasmid of the higher plant *Vicia faba* in (Mojica et al., 2000). However, that study has gone nowhere in the search for *cas* genes in the mitochondrial, chloroplast, and nuclear genomes of this plant species. In addition, no data on the existence of genetic elements of the CRISPR-Cas immunity in the nuclear plant genome has been obtained thus far.

Considering the evolutionary origin of mitochondria and the plant mitochondrial genome structure, search for genetic elements similar to those of CRISPR-Cas systems of bacteria and archaea in the mitochondrial, chloroplast, and nuclear genomes of the model plant *Arabidopsis thaliana* has been attempted by *in silico* methods. Taking into account the high dynamism of the plant mitochondrial genome, the genome-wide analysis of the mitochondrial genomes of three *A. thaliana* ecotypes (C24, Ler, and Col-0) was carried out with the purpose of searching for elements presumably associated with adaptive CRISPR-Cas immunity.

Materials and methods

The complete sequences of the nuclear (Col-0 ecotype), mitochondrial (C24, Ler, Col-0 ecotypes), and chloroplast (Ler, Col-0 ecotypes) genomes of the model plant *Arabidopsis thaliana* (L.) Heynh. were examined. DNA sequences were taken from the GenBank database (accession numbers of the nuclear genome: NC_003070, NC_003071, NC_003074,

NC_003075, NC_003076; of the mitochondrial genome: JF729200, JF729202, NC_037304; of the chloroplast genome: KX551970, NC_000932.) The sequences were first analyzed with the UGENE program (Okonechnikov et al., 2012). The same program was used to build illustrations (together with the Inkscape vector graphics package).

To seek elements of CRISPR-Cas systems in genomes, the CRISPROne online service was used (Zhang, Ye, 2017). To determine the origin of the detected CRISPR spacers, a search through the NCBI BLAST database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with the default parameters for viral taxa was carried out. Cases of coincidence with the numbers of mismatches fewer than 3 nucleotides were subsequently selected.

Sequence alignment of CRISPR loci in the mitochondrial genomes of *A. thaliana* ecotypes was carried out with the programs Matcher (paired) (Rice et al., 2000) and MUSCLE (multiple) (Edgar, 2004). The analysis of CRISPR spacer similarity to the genomes of species-specific viruses was carried out as in (Mihara et al., 2016) (according to Virus-Host DB https://www.genome.jp/virushostdb/3702).

Results and discussion

To date, the CRISPR locus, upstream leader sequence, and *cas* genes have been convincingly shown to be the critical components of CRISPR-Cas systems in bacteria and archaea as a general matter (Jansen et al., 2002; Richter et al., 2012). Resting on the known evolutionary relationship between mitochondria and bacteria, we searched for elements of the CRISPR-Cas system in the mitochondrial genome of three ecotypes of *A. thaliana* using approaches and methods of bioinformatics that are widely used in studying CRISPR-Cas systems of prokaryotes nowadays (Jansen et al., 2002; Makarova et al., 2006, 2015; Grissa et al., 2007; Zhang, Ye, 2017; Couvin et al., 2018).

The context analysis of the complete mitochondrial genome sequence of *A. thaliana* (C24 and Ler ecotypes) revealed a site whose structure is fully consistent with the organization of CRISPR cassettes of prokaryotic origin. The features of the nucleotide organization of the CRISPR-like locus in the mitochondrial genome of these ecotypes are shown in Fig. 1, *a*. As seen from the data presented, the CRISPR cassette found in the plant mitochondrial genome is formed by three 20-bp perfect direct repeats separated by two spacer sequences of 42 bp and 33 bp, respectively. By contrast, the genome-wide analysis of the Col-0 ecotype mitochondrial DNA showed that the CRISPR cassette structure is completely broken there as a result of four deletions and one insertion in the repeat unit (see Fig. 1, *b*).

In our opinion, a noteworthy result of the analysis of the ecotype-specific features of the mitochondrial CRISPR cassette is the fact that the localization of the CRISPR cassette (and its damaged variant) in the mitochondrial genomes of Arabidopsis C24, Ler, and Col-0 lines varies significantly with the complete match of the succession of repeats and spacers (Fig. 2). Such changes in the localization of the CRISPR cassette in the mitochondrial DNA of the studied Arabidopsis ecotypes are most likely to result from intense rearrangements in the mitochondrial genome due to high recombination activity, which is characteristic of the mitochondrial genomes of higher plants (Gualberto, Newton, 2017).

A special search revealed the presence of numerous CRISPR cassettes in the nuclear genome of *A. thaliana* (Fig. 3). Their sizes and arrangement on chromosomes are presented in Supplementary Material[1]. The total number of spacers present in 110 nuclear CRISPR cassettes is 330. We have not performed a detailed analysis of the similarity of the spacers of nuclear CRISPR cassettes to the genomes of plant viruses.

The results of the analysis of spacer sequences in a CRISPR cassette localized in Arabidopsis mitochondrial DNA with reference to the database of plant viruses are summarized in Table 1. The detected spacers were found to contain sections of nonrandom homology to the genomes of three strains of cauliflower mosaic virus able to infect *A. thaliana* plants. Moreover, regions of homology to mismatching genome units of different strains of this virus were identified in individual spacers (data not shown).

Search of the mitochondrial genome of the *A. thaliana* C24, Ler, and Col-0 ecotypes detected *cas* genes neither in the sequences immediately adjacent to the CRISPR locus nor in the rest of the genome. By contrast, sequences of individual *cas* genes were found in the nuclear genome (Table 2).

The *in silico* search of three Arabidopsis chromosomes (chromosomes 1, 2, and 3, respectively) made it possible to map the *cas5* gene, which is part of the effector module of type I CRISPR-Cas systems according to the existing classification of CRISPR-Cas systems (Makarova et al., 2015; Koonin et al., 2017). The *csm6* gene is located on the same chromosome 3 as the *cas5* gene. This gene encodes RNAse III-A, associated with the CRISPR-Cas system and involved (in prokaryotes) in the implementation of immunity against phages through degradation of phage transcripts (Jiang et al., 2016). The *csa5* gene, whose protein product is a universal component of type I-A CRISPR-Cas systems, was detected on chromosome 4 (Daume et al., 2014). This protein is believed to participate in the R-loop stabilization during the interference stage (Daume et al., 2014). Finally, chromosome 5 contains three regions of different lengths corresponding to the gene previously annotated in the Arabidopsis nuclear genome as *DEDDh*, which is a representative of the 3′→5′ exonuclease gene family involved in the metabolism of small noncoding RNAs (Chen et al., 2018). The attribution of this gene to the *cas* family may mean that its protein product can perform several functions *in vivo*, including plant protection from the nucleic acids of viruses and plasmids.

The reverse transcriptase (RT) genes associated with type I and III CRISPR-Cas systems were found to be represented on all the five Arabidopsis chromosomes by a significant number of copies (43 in total) (see Table 2). Presently, these enzymes are assigned a particularly important role in the functioning of type III CRISPR-Cas systems, which is to incorporate new spacers into the existing CRISPR cassette, both with the direct RT participation and with the participation of the RT Cas1 fusion protein (Silas et al., 2016; Toro et al., 2017).

Thus, for the first time ever, our search for elements of CRISPR-Cas systems in the mitochondrial and nuclear genomes of *A. thaliana* made it possible to detect the main genetic elements of prokaryotic adaptive immunity, including CRISPR loci and *cas* genes, in the genome of this plant. With

[1] Supplementary Material is available in the online version of the paper: http://www.bionet.nsc.ru/vogis/download/pict-2019-23/appx18.pdf

a

```
C24              86224  AACTCGACTGAAAGGAGAGGTTGTGAACACAAACTCGACTGAAAGGAGAG  86273
                        ||||||||||||||||||||||||||||||||||||||||||||||||||
Ler             163198  AACTCGACTGAAAGGAGAGGTTGTGAACACAAACTCGACTGAAAGGAGAG 163247

C24              86274  GTTGTGAACACAAACTCGACTGAAAGGAGAGGTCCAAGGTAATTTATTAC  86323
                        ||||||||||||||||||||||||||||||||||||||||||||||||||
Ler             163248  GTTGTGAACACAAACTCGACTGAAAGGAGAGGTCCAAGGTAATTTATTAC 163297

C24              86324  TCTTATAAAAGAGGGAACTCGACTGAAAGGAGAGG   86358
                        ||||||||||||||||||||||||||||||||||
Ler             163298  TCTTATAAAAGAGGGAACTCGACTGAAAGGAGAGG  163332
```

в

```
C24      AACTCGACTGAAAGGAGAGGTTGTGAACACAAACTCGACTGAAAGGAGAGGTTGTGAACA
Ler      AACTCGACTGAAAGGAGAGGTTGTGAACACAAACTCGACTGAAAGGAGAGGTTGTGAACA
Col-0    -ACTC----------AGAGATCAGGA----AAACTTGATAGAAAG-------CCCTAGCT
          ****          **** *   **    ***** **  *****          * *
C24      CAAACTCGACTGAA-----------------AGGAGAGGTCCAAGGTAATTTATTACTC
Ler      CAAACTCGACTGAA-----------------AGGAGAGGTCCAAGGTAATTTATTACTC
Col-0    CTAACTTGATAGAATGCCCCAGCAATGGCCATCGGAGAGGTCCAACGTAATTTATTACTC
         * **** **  ***                  ************ **************
C24      TTATAAAAGAGGGAACTCGACTGAAAGGAGAGG
Ler      TTATAAAAGAGGGAACTCGACTGAAAGGAGAGG
Col-0    TTATAAAAGAGGGAACTCGACTGAAAGGAGAGG
         *********************************
```

Fig. 1. (*a*) Pair alignment of the CRISPR loci in the mitochondrial genomes of the C24 and Ler ecotypes. Repeats are shown in boldface. (*b*) Multiple alignment of the CRISPR loci in the mitochondrial genomes of *A. thaliana* ecotypes.

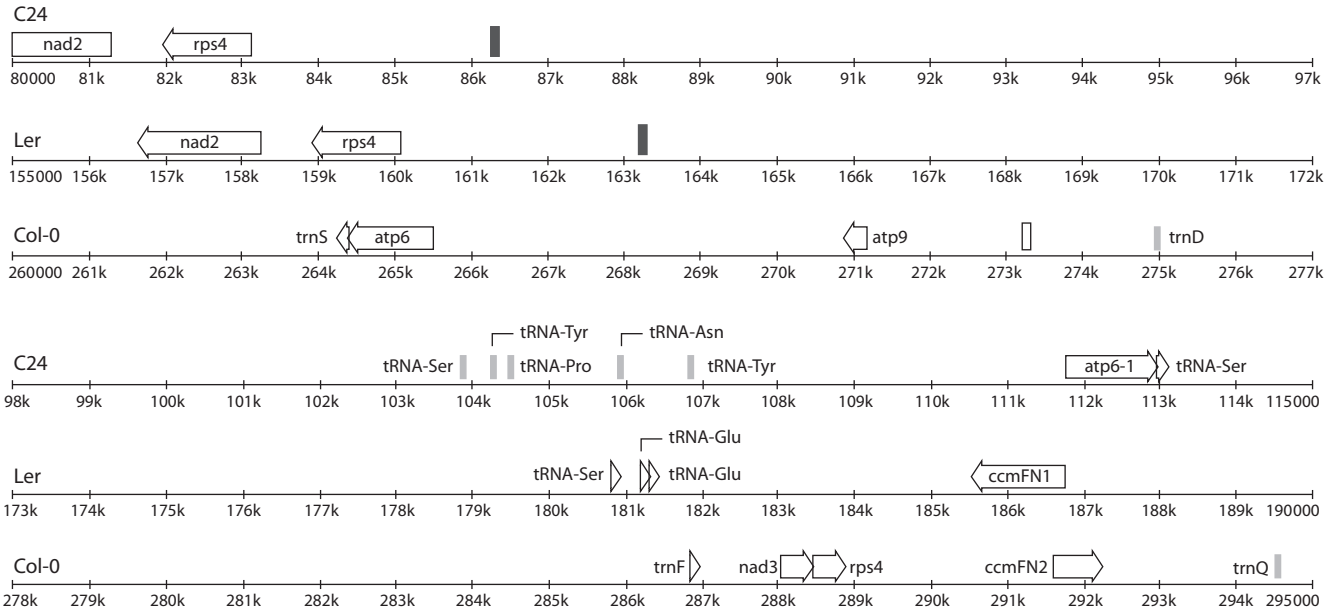

Fig. 2. Localization of the CRISPR cassette in the mitochondrial genomes of *A. thaliana* ecotypes.

The positions of the CRISPR cassette are indicated by black boxes above the C24 and Ler lines. The position of the damaged CRISPR cassette in the mitochondrial genome of the Col-0 line is indicated by open box.

the exception of the CRISPR cassette found in the mitochondrial genome, the structural elements of the system are localized in the nuclear genome. In general, in accordance with the classification proposed in (Makarova et al., 2015; Koonin et al., 2017), the still incomplete list of genes (*cas5*, *csm6*, *csa5*, cd06127, and RT) associated with CRISPR-Cas immunity of prokaryotes found in the *A. thaliana* genome allows the system found in this plant to be assigned to class 1 systems, which have a multi-subunit effector module.

However, during our study, we failed to find any structure corresponding to the characteristics of leader sequences of the prokaryotic type in the mitochondrial or nuclear genomes

of *A. thaliana* (Alkhnbashi et al., 2016). Our attempt to seek elements of CRISPR-Cas systems in the chloroplast genomes of *A. thaliana* Ler and Col-0 ecotypes was unsuccessful as well.

It is the first finding of such canonical elements of CRISPR-Cas systems of prokaryotic origin as CRISPR loci and *cas* genes in a higher eukaryote, namely, the model plant *A. thaliana*. Only a single CRISPR cassette, whose spacer sequences exhibit nonrandom homology with the cauliflower mosaic virus genome, was found in the mitochondrial genome of Arabidopsis (see Table 1). The ability of this virus to infect plants of the species under investigation is of special importance. The Arabidopsis lines studied are characterized by a significant difference between the C24 and Ler ecotypes in the genomic localization of the CRISPR locus. In the Col-0 ecotype, the CRISPR cassette structure is completely broken as a result of four deletions and one insertion in the region of direct repeats (see Fig. 1, *b*). This fact points to intense mitochondrial genome reorganization in higher plants, which manifest itself as a rapid occurrence of interline differences at the level of mitochondrial DNA.

From the evolutionary point of view, the possible existence of CRISPR-Cas immunity in plants seems quite justified since DNA-containing plant organelles – mitochondria and chloroplasts – are obviously attractive targets for viruses and plasmids of alien origin. Plant mitochondria are particularly vulnerable in this regard due to the existence of mitoviruses that attack this type of organelles (Marienfeld et al., 1997; Bruenn et al., 2015; Nibert et al., 2018) and the natural ability of plant mitochondria for DNA uptake (Koulintchenko et al., 2003). In general, with regard to the currently available data, it seems premature to form any hypotheses on the evolutionary origin of CRISPR-Cas system elements in the Arabidopsis genome. However, it should be noted that a lot of experimental data for reconstructing the scenarios of the origin and evolution of CRISPR-Cas systems in prokaryotes have already been reported (Koonin, Makarova, 2019). When trying to consider the issue of the origin of CRISPR-Cas systems in plants in the context of eukaryogenesis, the data may be useful (Koonin, 2015; Lopez-Garia et al., 2017). In this case, it should be taken into account that similar protection systems against pathogenic nucleic acids might have been present in both the alpha-proteobacterial symbiont that gave rise to mitochondria and the archaean host of the protomitochondrial endosymbiont. Notionally, in evolutionary terms the presence of such a protective system cannot be ruled out for the cyanobacterial ancestor of modern chloroplasts either.

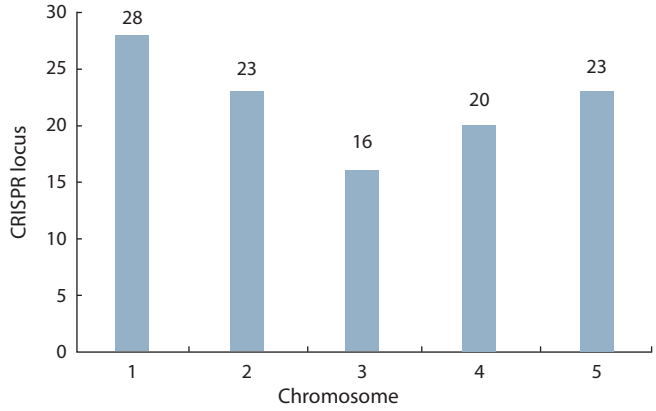

Fig. 3. Quantitative distribution of CRISPR cassettes over nuclear chromosomes of *A. thaliana.*

At the current stage of Arabidopsis CRISPR-Cas systems research, we were unable to identify a set of prokaryotic *cas* genes whose products would form the adaptation and effector modules of a class 1 CRISPR-Cas system and thus support the stages of adaptation, expression, and interference (Koonin et al., 2017). With regard to the known data on the wide variety of CRISPR-Cas systems found in bacteria and archaea (Westra et al., 2016; Koonin et al., 2017; Koonin, Makarova, 2019), it is natural to expect that the plant adaptive immunity mechanism can differ significantly from that in prokaryotes. Therefore,

Table 1. Alignment of the CRISPR spacers found in the *A. thaliana* mitochondrial genome with the cauliflower mosaic virus genome

Alignment*	Isolate	NCBI GenBank identifier
Spacer 1 TCCAAGGTAATTTATTACTCTTATAAAAGAGGG ||||.||.|||....||||||| KJ716236 7512 AAGGGAAATTAGGGTTCTTATA 7533	Cabb B-JI	KJ716236
Spacer 1 TCCAAGGTAATTTATTACTCTTATAAAAGAGGG |.||||..||..|||.|..||||||| M10376 3301 TACAAGAAAAAATATAAGGCTTATAAA 3327	D/H	M10376
Spacer 2 TTGTGAACACAAACTCGACTGAAAGGAGAGGTTGTGAACACA |.||||||| |.|||.|| AB863182 1782 ATAAACTCGA-TCAAAGAAG 1800	TUR239	AB863182

* Numerals indicate the localization of the regions of homology in the virus genome.

Table 2. Description of *cas* genes detected *in silico* in the nuclear genome of *A. thaliana*

Chromosome	Gene	Type of CRISPR-Cas system	Start–End of ORF*	Cas gene identification according to the classification (Makarova et al., 2015; Koonin et al., 2017)
1	*cas5*	I	11918718–11919893	COG1688
2	*cas5*	I	2540882–2541148	cd09693
3	*cas5*	I-B	12023929–12024258	mkCas0191
	csm6	III-A, III-D	10339967–10340668	cd09742
4	*csa5*	I-A	5072131–5072607	mkCas0163
5	*DEDDh*	I	9481429–9481962 9482093–9482377 9482399–9482647	cd06127
1	RT	I, III	11916164–11916466 11916554–11917162	pfam00078
2			10826–12517 17624–20293 23971–25653 30408–31532 3223463–3224872 3224999–3225685 3228729–3230522 5235099–5235800 5241327–5241869 5241873–5242298 5600424–5602496 5602506–5603024 5619463–5619846 5619956–5620156	
3			9435745–9436644 9436766–9437260 9444515–9444907 9445029–9445523 10341391–10343106 10343111–10343806 12021817–12022311 13167127–13168218 13174969–13175118 13175226–13175606 13467703–13469727 13469756–13470868 13476637–13477317 15489669–15490154 15490132–15490506 15493988–15496318	
4			3296005–3296457 3303483–3304247 3304260–3305717 3744233–3746887 3749914–3752901 3754635–3756176 5071003–5071551 5081577–5082176	
5			14576286–14579228 14582070–14582471 14582509–14584656	

* ORF, open reading frame.

it is obvious that only the use of a comprehensive approach, including transcriptomics and proteomics techniques along with genomics ones, will allow getting a more complete idea of the genes and their protein products that form the plant adaptive immunity system.

Conclusions

For the first time ever, such elements of the CRISPR-Cas system of prokaryotes as CRISPR cassettes and *cas* genes were detected *in silico* in the genome of the higher plant *A. thaliana*. This finding can provide a staging ground for

further detailed genomic, transcriptomic and proteomic studies of a wider set of plant species (including the most important crops) in addition to Arabidopsis in order to determine groups of genes whose expression may be associated with the activity of the natural adaptive plant immunity mechanism. The applied relevance of the expected scientific results on the molecular nature of adaptive plant immunity can hardly be overestimated.

References

Alkhnbashi O.S., Shah S.A., Garret R.A., Saundlers S.J., Costa F., Backhoven R. Characterizing leader sequences of CRISPR loci. Bioinformatics. 2016;32:i576-i585. DOI 10.1093/bioinformatics/btw454.

Archibald J.M. Endosymbiosis and eukaryotic cell evolution. Curr. Biology. 2015;25:R911-R921. DOI 10.1016/j.cub.2015.07.055.

Barrangou R., Freemaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1709-1712. DOI 10.1126/science.1138140.

Bruenn J.A., Warner B.E., Yerramsetty P. Widespread mitovirus sequences in plant genomes. PeerJ. 2015;3:e876. DOI 10.7717/peerj. 876.

Chen J., Liu L., You C., Gu J., Ruan W., Zhang L., Gan J., Cao C., Huang Y., Chen X., Ma J. Structural and biochemical insights into small RNA 3′ end trimming by *Arabidopsis* SDN1. Nat. Commun. 2018;9:3585. DOI 10.1038/s41467-018-05942-7.

Couvin D., Bernheim A., Toffano-Nioche C., Touchon M., Michalik J., Néron B., Rocha E., Vergnaud G., Gautheret D., Pourcel C. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018:46:W246-W251. DOI 10.1093/nar/gky425.

Daume M., Plagens A., Randau L. DNA binding properties of the small cascade subunit Csa5. PLoS One. 2014;9:e105716. DOI 10.1371/journal.pone.0105716.

Douce R., Neuburger M. The uniqueness of plant mitochondria. Annu. Rev. Plant Physiol. Plant Mol. Biol. 1989;40:371-414. DOI 10.1146/annurev.pp.40.060189.002103.

Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-1797. DOI 10.1093/nar/gkh340.

Esser K., Kück V., Lang-Hinrichs C., Lemke P.A., Osiewaez H.D., Stahl P., Tudzynski U. Plasmids of Eukaryotes. Fundamentals and Applications. Berlin; Heidelberg: Springer-Verlag, 1986.

Grissa I., Vergnaud G., Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:W52-W57. DOI 10.1093/nar/gkm360.

Gualberto J.M., Newton K.J. Plant mitochondrial genomes: dynamics and mechanism of mutation. Annu. Rev. Plant Biol. 2017;68:225-252. DOI 10.1146/annurev-arplant-043015-112232.

Jansen R., van Embden D.A., Gaastra W., Schouls L.M. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 2002;43:1565-1575. DOI 10.1046/j.1365-2958.2002.02839.x.

Jiang V.V., Samai P., Marraffini L.A. Dagradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. Cell. 2016;164:710-721. DOI 10.1016/j.cell.2015.12.053.

Kleine T., Maier U.W., Leister D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu. Rev. Plant Biol. 2009;60:115-138. DOI 10.1101/cshperspect.a015990.

Koonin E.V. Origin of eukaryotes from within archaea, archaeal eukaryome and bursts of gene gain: eukaryogenesis just made easier? Phil. Trans. R. Soc. B. 2015;370:201403333. DOI 10.1098/rstb.2014.0333.

Koonin E.V., Makarova K.S. Origins and evolution of CRISPR-Cas systems. Phil. Trans. R. Soc. B. 2019;374(1772):20180087. DOI 10.1098/rstb.2018.0087.

Koonin E.V., Makarova K.S., Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 2017;37:67-78. DOI 10.1016/j.mib.2017.05.008.

Koulintchenko M., Konstantinov Y., Dietrich A. Plant mitochondria actively import DNA via the permeability transition pore complex. EMBO J. 2003;22:1245-1254. DOI 10.1093/emboj/cdg128.

Kuzmin E.V., Levchenko I.V. S1 plasmid from cms-S-maize mitochondria encodes a viral type DNA-polymerase. Nucleic Acids Res. 1987;15:6758. DOI 10.1093/nar/15.16.6758.

Kuzmin E.V., Levchenko I.V., Zaitseva G.N. S2 plasmid from cms-S-maize mitochondria potentially encodes a specific RNA polymerase. Nucleic Acids Res. 1988;16:4177. DOI 10.1093/nar/16.9.4177.

Lander E.S. The heroes of CRISPR. Cell. 2016;164:18-28. DOI 10.1016/j.cell.2015.12.041.

Lopez-Garcia P., Eme L., Moreira D. Symbiosis in eukaryotic evolution. J. Theor. Biol. 2017;434:20-33. DOI 10.1016/j.jtbi.2017.02.031.

Makarova K.S., Grishin N.V., Shabalina S.A., Wolf Y.I., Koonin E.V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct. 2006;1:7. DOI 10.1186/1745-6150-1-7.

Makarova K.S., Wolf Y.I., Alkhnbashi O.S., Costa F., Shah S.A., Saunders S.J., Barrangou R., Brouns S.J.J., Charpentier E., Haft D.H., Horvath R., Moineau S., Mojica F.J.M., Terns R.M., Terns M.P., White M.F., Yakunin A.F., Garrett R.A., van der Oost J., Backofen R., Koonin E.V. An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 2015;13:1-15. DOI 10.1038/nrmicro3569.

Marienfeld J.R., Unseld M., Brandt P., Brennicke A. Viral nucleic acid sequence transfer between fungi and plants. Trends Genet. 1997;13:260-261. DOI 10.1016/S0168-9525(97)01168-2.

Mihara T., Nishimura Y., Shimizu Y., Nishiyama H., Yoshikawa G., Uehara H. Linking virus genomes with host taxonomy. Viruses. 2016;8:66. DOI 10.3390/v8030066.

Mojica F.J.M., Diez-Villasenor C., Soria E., Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol. Microbiol. 2000;36:244-246. DOI 10.1046/j.1365-2958.2000.01838.x.

Mojica F.J., Garcia-Martinez J., Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 2005;60:174-182. DOI 10.1007/s00239-004-0046-3.

Nibert M., Vong M., Fugate K.K., Debat H.J. Evidence for contemporary plant mitoviruses. Virology. 2018;518:14-24. DOI 10.1016/j.virol.2018.02.005.

Okonechnikov K., Golosova O., Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012;28;1166-1167. DOI 10.1093/bioinformatics/bts091.

Pittis A.A., Gabaldon T. Late acquisition of mitochondria by a host with chimeric prokaryotic ancestry. Nature. 2016;531(7592):101-104. DOI 10.1016/S0168-9525(00)02024-2.

Poole A.M., Gribaldo S. Eukaryotic origins: how and when was the mitochondrion acquired? Cold Spring Harb. Perspect. Biol. 2014;6:a015990. DOI 10.1101/cshperspect.a015990.

Rice P., Longden I., Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000;16(6);276-277. DOI 10.1016/S0168-9525(00)02024-2.

Richter C., Chang J.T., Fineran P. Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. Viruses. 2012;4:2291-2311. DOI 10.3390/v4102291.

Shahi S., Eusebio-Cope A., Kondo H., Hillman B.I., Suzuki N. Investigation of host range of and host defense against a mitochondrially replicating mitovirus. J. Virol. 2019;93(6):e01503-18. DOI 10.1128/JVI.01503-18.

Silas S., Mohr G., Sidote D.J., Markham L.M., Sanchez-Amat A., Bhaya D.,Lambowitz A.M., Fire A.Z. Direct CRISPR spacer acquisi-

tion from RNA by a natural reverse transcriptase-Cas1 fusion protein. Science. 2016;351(6276):aad4234. DOI 10.1126/science.aad4234.

Thomas C.M. The nucleotide sequence and transcription of minicircular mitochondrial DNA's associated with fertile and cytoplasmic male-sterile lines of sugar beet. Nucleic Acids Res. 1986;14:9353-9370. DOI 10.1093/nar/14.23.9353.

Toro N., Martinez-Abarca F., Gonzalez-Delgado A. The reverse transcriptases associated with CRISPR-Cas systems. Sci. Rep. 2017;7:7089. DOI 10.1038/s41598-017-07828-y.

Westra E.R., Dowling A.J., Broniewski J.M., van Houte S. Evolution and ecology of CRISPR. Annu. Rev. Ecol. Evol. Syst. 2016;47:307-334. DOI 10.1146/annurev-ecolsys-121415-032428.

Zhang Q., Ye Y. Not all predicted CRISPR-Cas systems are equal: isolated *cas* genes and classes of CRISPR like elements. BMC Bioinformatics. 2017;18:92. DOI 10.1186/s12859-017-1512-4.

Zhao N., Wang Y., Hua J. The roles of mitochondrion in intergenomic gene transfer in plants: a source and a pool. Int. J. Mol. Sci. 2018;19:547. DOI 10.3390/ijms19020547.

ORCID ID

Yu.M. Konstantinov orcid.org/0000-0002-0601-2788
I.S. Petrushin orcid.org/0000-0002-8788-5352

Acknowledgements. The authors are grateful to Prof. Ian Max Møller and Dr. R. Shyama Prasad Rao for their interest in the research, critical remarks, and useful discussion. The authors acknowledge Dr. Alexey Fadeev's valuable advice and useful comments on the paper structure. The authors sincerely express their deep gratitude to the anonymous reviewer, whose favorable comments and analysis of the original paper version were helpful in its substantial processing, which improved the presentation of the results obtained.

Conflict of interest. The authors declare no conflict of interest.

Received June 5, 2019. Revised August 20, 2019. Accepted August 21, 2019.