ANNOTATION

For the dissertation for the Doctor of Philosophy (PhD) degree
on the specialty "8D05105 - Biotechnology" of
Pozharskiy Alexandr Sergeyevich

on the theme: «Investigation of the gene pool and genome editing of the tomato varieties bred in Kazakhstan»

General description of the dissertation.

The work is focused n editing the genome of domestic tomato varieties selected by molecular screening for pathogen resistance markers in order to achieve resistance to powdery mildew caused by the fungus *Oidium neolycopersici*.

Relevance of the research topic.

Tomato *Solanum lycopersicum* L. is one of the most important representatives of cultivated plants of the Solanaceae family. According to the Food and Agriculture Organization (FAO) of the United Nations, tomatoes constitute an important part of the total vegetable production in Kazakhstan: in 2022, 788,760 tons were harvested from 30.2 thousand hectares. However, tomato production in the country is developing more extensively than intensively: over the last 30 years, the area under cultivation has doubled, with no improvement in specific yield per hectare of cultivated area. According to the State Revenue Committee of the Ministry of Finance of the Republic of Kazakhstan, imports account for a significant share of the local tomato market. In 2020, about 58,636 tons of tomatoes were imported, mainly from Central Asian countries, which corresponds to about 7.4% of local production. Domestic tomato growers also rely on seed imports from abroad (57.3 tons of vegetable seeds were imported in 2020, although the specific crop was not counted). Imports are associated with the risk of introducing and transmitting new pathogenic infections. Various pathogens (bacteria, fungi, viruses and oomycetes) can cause huge crop losses and, when spread, are very difficult to manage.

Among tomato varieties approved for cultivation in the country, foreign varieties prevail with a significant share of varieties from Russia and other countries of the former USSR. This dependence on imported planting material creates various risks for food security, the most dangerous of which is the possible importation of dangerous pests, weeds and pathogens. Thus, it is important for the domestic crop market to increase the use of old and newly obtained varieties of local breeding as part of integrated plant epidemiological control. To control potentially harmful phytopathogens, it is necessary not only to identify and destroy infected plants in a timely manner, but also to increase the disease resistance potential of cultivated crops through breeding and selection of varieties with genetic resistance factors. Molecular markers related to plant disease resistance play a crucial role in modern breeding programs, as their use in marker assisted selection (MAS) helps to significantly reduce the time and labor required to develop new resistant varieties. However, in Kazakhstan, the introduction of such advanced tomato breeding methods is limited by relatively low economic and scientific interests. To date, no systematic efforts have been made to lay the molecular genetic foundation of tomato breeding programs.

Powdery mildew is one of the most widespread plant diseases worldwide. More than 800 species of fungi belonging to the family Erysiphaceae are capable of infecting about 10,000 plant species, thereby threatening most economically important crops. A key role in plant response to powdery mildew pathogens is played by MLO group protein genes, which are a factor in plant susceptibility to the disease in all groups of land plants. For tomato, genes encoding SlMLO1, SlMLO5, and SlMLO8 proteins have been described as associated with susceptibility to powdery mildew caused by the fungi *Oidium neolycopersici* and *Leveillula taurica.* Incorporation of protein-disrupting mutations into these genes is seen as a promising way to achieve resistance to powdery mildew. The CRIPSR/cas9 technology has previously been successfully used to inactivate the SlMLO1 protein gene. The use of simultaneous editing of SlMLO1, SlMLO5 and SlMLO8 protein

genes in the present work is carried out for the first time, will allow to achieve stronger resistance and avoid overcoming resistance by pathogens.

The goal of the research:

Molecular genetic analysis of tomato varieties of Kazakhstani selection, selection and genome editing of promising varieties to obtain plants resistant to fungal diseases.

Research objectives:

1. To study the genetic diversity of tomato varieties of Kazakhstani selection using microsatellite markers and known markers of resistance to fungal and viral pathogens.

2. To conduct CRISPR/Cas9 editing of genes of Mlo group in selected tomato varieties to obtain plants resistant to fungal diseases.

3. Conduct phylogenetic analysis of Mlo group genes in available plant genomes to assess the limits of applicability of the results to other plant species.

The research objects: Tomato *Solanum lycopersicum* L. tomato powdery mildew *Oidium neolycopersici* L. Kiss, 2001.

The subject of the research: Resistance of tomato plants to powdery mildew and/or other fungal and bacterial diseases by inactivation of Mlo-like proteins.

Research methods: DNA, RNA and protein isolation, polymerase chain reaction, reverse transcription, gel electrophoresis, Western blotting, DNA sequencing, Golden-Gate cloning, bacterial transformation - heat shock and electroporation, production and cultivation of plant callus, plant regeneration from callus, bioinformatics analysis.

The scientific novelty of the research.

For the first time, a study of the genetic structure of open field tomato varieties of domestic selection was carried out using SSR markers, as well as screening using markers of resistance to dangerous quarantine diseases of tomato. Genome editing by CRISPR/Cas9 method of domestic tomato varieties was carried out for the first time to achieve resistance to powdery mildew caused by the fungus *Oidium neolycpersici*. The presence of this pathogen in Kazakhstan was confirmed for the first time using morphological and molecular identification. The results of simultaneous inactivation of genes *SlMlo1, SlMlo5, SlMlo8* and its effect on the overall resistance of tomato to powdery mildew were tested for the first time.

Theoretical and practical significance of the research.

Genetic diversity of tomato varieties of Kazakhstani selection in comparison with foreign varieties was characterized for the first time. Tomato plants of local varieties were used for the first time as an initial object for achieving resistance to infectious diseases by genome editing. Contributions were made to the establishment of homologous relationships between MLO group protein genes among the available genome of cover-seeded plants, and the possible effect of genomic editing of genes of this group on the achievement of resistance in different plant groups was predicted.

Genomic editing of promising tomato cultivars will allow the development of plant forms with high resistance to powdery mildew with the possibility of subsequent production of plants free of transgenic sequences. The work will lay the foundation for further research in the field of plant genome editing in IPBB and other research organizations of the Republic of Kazakhstan.

The main provisions for the defense:

1. Domestic varieties of tomato of open ground have a high degree of relatedness and are close to varieties of Russian selection, with frequencies of SSR-genotypes coincidence 60-100% and similarity assessment by STRUCTURE algorithm up to 90-95%. No Kazakh varieties carrying markers of resistance to viral diseases were identified. Seven Kazakh varieties resistant to *F. oxysporum* and two varieties resistant to *P. infestans* were identified. The varieties Meruert and Leader were selected as promising for genome editing by CRISPR/Cas9 system.

2. Mutations in the *SlMlo5* and *SlMlo8* genes do not lead to increased resistance compared to a single *SlMlo1* gene, as confirmed by the P-values of Welch's t-test, significantly exceeding the significance threshold of 0.05, taking into account the correction for foreign comparisons. Sequencing of targeted *SlMlo* gene editing sites confirms the presence of genetic mosaicism: 20 to

50% of the reads covering the deletion region demonstrate the absence of deletion. Acquired resistance to powdery mildew is maintained under mosaicism, as shown by the absence of a significant effect when comparing lines with full expression of mutations with mosaic lines by the Welch test at a significance threshold of 0.05.

3. Highly conserved sequence of *SlMlo1*, *SlMlo5*, and *SlMlo8* genes in coding regions (no variants in the *SlMlo1* gene, 2 synonymous variants in the *SlMlo5* gene, 2 synonymous and 2 missense substitutions in the *SlMlo8* gene) were revealed based on the available data on genomic variability of 166 tomato lines. In particular, the invariance of the selected gRNA recognition sites was confirmed. Thus, the designed gRNAs can be used on any cultivar and tomato line with a high chance of editing specificity due to 100% conservativity of the selected recognition sites.

4. The designed gRNAs can also potentially be used with other species of the *Solanaceae* family, but the accuracy of such mating requires further refinement taking into account the available genomic data of the respective species.

Main research results

1. Molecular genetic analysis of 68 open field tomato varieties, including 15 varieties of Kazakhstani selection using 13 SSR markers to analyze the genetic structure of the sample and 14 CAPS and SCAR markers of resistance to 5 pathogens was carried out. All local varieties showed high genetic similarity according to the used SSR markers. Among all 11 polymorphic markers, only three markers showed genotype variation within local cultivars: LEMDDNA with a set of detected alleles 211, 213, 227, 233; LELEUZIP with alleles 102, 105, 106; and TMS58 with alleles 226, 228, 230. The markers LELE25, LEATRACAb, and TM63 had only two divergent genotypes among the 15 local cultivars, and the marker LE20592 had the only divergent genotype in the cultivar Sweetgrass. This variety was the most distinctive among all the local varieties. The varieties Amber, Leader, Luchezarny, Meruert, Vostorg, and Mechta formed a group of close genotypes together with the Russian varieties Novichok, Korolek, Rassvet 365, and 33 Bogatyrs. LEPRP4 and TMS58 markers were characterized by the highest frequency of missing genotypes in this group. Other local varieties were more diverse compared to various foreign varieties. It was found that the domestic varieties of open field tomato had a high degree of relatedness and were close to the varieties of Russian selection. Analysis of pathogen resistance markers revealed the predominant presence of resistance loci to the fungus *Fusarium oxysporum* and the oomycete *Phytophtora infestans* compared to viruses. The most frequent marker was At2, associated with resistance locus I against *F. oxysporum:* half of all 64 successfully genotyped accessions were positive for resistance. Another marker for resistance to *F. oxysporum*, Z1063, linked to the I2 resistance gene, was observed in six accessions, including the local variety Meruert. Two codominant markers, Ph3-gsm and TG328, are associated with the Ph-3 locus conferring resistance to *P. infestans*. Two local cultivars, Meruert and Leader, have a resistant Ph3-gsm allele. No markers associated with resistance to virus infections were identified among the domestic varieties. The varieties Leader and Meruert possessing markers of resistance to *P. insfestans* and *F. oxisporum* were selected as promising for genome editing by CRISPR/Cas9 system.

2. Guide RNAs were designed to introduce deletions into *SlMlo1, SlMlo5,* and *SlMlo8* genes. The designed nRNAs were cloned by the GoldenGates method as part of the CRISPR/Cas9 editing cassette as part of the plasmid pCBC-DT1T2 with subsequent cloning into the binary vector pKSE401. Seventeen viable tomato plants of the cultivar Leader and 18 plants of the cultivar Meruert were obtained by agrobacterial transformation method, for which the presence of T-DNA insertion was confirmed and the expression of nRNA and Cas9 protein was confirmed. Genome editing was confirmed by sequencing in 9 plants of the cultivar Leader and 11 plants of the cultivar Meruert. These 20 plants were tested for resistance by artificial infection with *Oidium neolycopersici.* The inoculation experiment showed that there was no advantage in simultaneous editing of several genes, since the effect of mutations in the SlMlo5 and SlMlo8 genes was negligible compared to the *SlMlo1* gene. The result of genome editing accomplished through agrobacterial transformation is mosaic plants. In addition, the transgenic insertion of the CRISPR/cas9 system is retained. Thus, to obtain tomato lines suitable for use in breeding, further

cultivation is necessary to obtain pure genotypes free of mosaicism and to eliminate the transgenic insertion.

3. Analysis of SlMlo gene variation based on publicly available data on genomic variability of 166 tomato lines showed that the genes are highly conserved. The *SlMlo1* gene turned out to be one of the most stable, with 16 variations in introns and no variants in coding regions. *SlMl5* and *SlMlo8*, minor susceptibility factors to powdery mildew, have relatively high intron variability with low variability in coding regions. A phylogenetic analysis of 4886 MLO protein sequences from a wide range of plant species was performed. As a result, close homologs of *SlMLO1*, *SlMLO5*, and SlMLO8 proteins were identified in species of the genera *Solanum*, *Capsicum*, and *Nicotiana* of the *Solanaceae* family. However, their contribution to powdery mildew resistance of the respective species remains unclear. Thus, the developed guide RNAs can be used to edit *SlMlo1, SlMlo5, SlMlo8* genes of any tomato varieties and lines, since the target gene regions are highly conserved. The possibility of applying the developed guide RNAs to other species of solanaceous plants requires further clarification using new genomic and genetic data.

Conclusions

1. Kazakhstan tomato varieties have a high degree of relatedness with each other and are close to Russian varieties (from 60 to 100% coincidence of SSR-genotypes between varieties).

2. No markers of resistance to viral pathogens were found among Kazakh varieties. 8 Kazakh varieties possessed resistant genotype marker At2, variety “Meruert” also possessed resistant genotype marker Z103 (*F. oxysporum*) 2 Kazakh varieties, “Meruert” and “Leader”, possessed resistant genotype marker Ph3 (*P. infestans*).

3. Varieties 'Meruert' and 'Leader' carrying resistance markers to *F. oxysporum* and *P. infestans* were selected as candidates for genome editing.

4. Successfully assembled, cloned and transformed plants with vector constructs for *SlMlo1, SlMl5, SlMlo8* gene editing. A total of 35 plants of Leader and Meruert cultivars were obtained, of which 18 were confirmed carriers of mutations in *SlMlo* genes. The expression of gRNA and Cas9 protein in plants was confirmed by OT-PCR and western blotting, respectively

5. Sequencing of target gene regions confirmed the presence of deletions, but showed the partial nature of the mutations due to heterozygosity and possible mosaicism: on average, 50-70% of reads per line showed the presence of the deletion. Only two lines of 'Meruert' plants carrying MLO1 and MLO1+MLO8 mutations were recognized as pure mutant lines (98.89% and 97.96% of mutant reads, respectively).

6. Inoculation of plants with powdery mildew pathogen confirmed the positive effect of mutation in the *SlMlo1* gene on disease resistance (significant reduction in the average number of affected areas and their diameter compared to control plants, confirmed by Welch's t-test at the significance threshold of 0.001 with correction for multiple comparisons), but showed no effect for *SlMlo5*, *SlMlo8* genes.

7. The regions of *SlMlo1, 5, 8* genes selected for editing are highly conserved (cite figures) among a wide range of tomato genomes and thus the designed gRNAs can be used for different tomato cultivars

8. Sequence analysis of MLO proteins and genes among land plants identified the closest homologs of the corresponding genes among plants of the Solanaceae family, including 12 MLO1 homologs in the genera *Solanum*, *Capsicum*, *Nicotiana*; 3 MLO5 homologs in the genus *Solanum*; 7 MLO8 homologs in the genera *Solanum, Capsicum, Nicotiana*; however, further conclusions on the applicability of the gRNAs and the involvement of the genes in powdery mildew resistance are limited by the available data and require further consideration.

Personal contribution of the author.

The author was directly involved in all stages of the study: formation of the aim and objectives, project development, collection of biomaterial, conducting experimental work, interpretation of the results obtained, formulation of conclusions, and preparation of the results for publication.

Connection with the plan of basic scientific work. The work was carried out within the framework of the task "Obtaining high-yielding tomato lines resistant to biotic and abiotic stress by genomic editing of promising varieties" of the target funding program BR18574149 "Development of high-yielding varieties and lines of agricultural crops based on innovative biotechnologies".

Approbation of work.

The results of the work were presented at the following conferences:

6th International Scientific Conference "Genetics, Genomics, Bioinformatics and Plant Biotechnology" (PlantGen2021), June 14-18, 2021, Novosibirsk, Russia.

IV. International Agricultural, Biological & Life Science Conference Agbiol 2022 August 29-31, 2022 Edirne, Turkey.

International Forum "Modern Trends of Sustainable Development of Biological Sciences" dedicated to the 90th anniversary of Al-Farabi KazNU, 2024, March 27-28, Almaty, Kazakhstan.

The results of the dissertation work are included in the annual reports on research under the targeted funding program BR18574149 (2023-2024).

Publications.

10 scientific papers have been published on the topic of the study, including:

3 articles in scientific editions included in the 1-2 quartiles of the impact factor in the Web of Science database; 1 article in a domestic journal recommended by the Committee for Quality Assurance in Science and Higher Education, 4 abstracts of international conferences, 1 methodological manual and 1 utility model patent.