plants

MDPI

Article

The organization features of the mitochondrial genome of sunflower (*Helianthus annuus* L) with ANN2-type of male-sterile cytoplasm

Maksim S. Makarenko[1,6,*], Alexander V. Usatov[1], Tatiana V. Tatarinova[2,6-8], Kirill V. Azarin[1], Maria D. Logacheva[3,6], Vera A. Gavrilova[4], Kornienko I.V.[1,9], Renate Horn[5]

[1] Southern Federal University, Rostov-on-Don, Russia
[2] Department of Biology, University of La Verne, La Verne, CA, USA
[3] Skolkovo Institute of Science and Technology, Moscow, Russia
[4] The N.I. Vavilov All-Russian Institute of Plant Genetic Resources, Saint Petersburg, Russia
[5] University of Rostock, Institute of Biological Sciences, Plant Genetics, Rostock, Germany
[6] Institute for Information Transmission Problems, Moscow, Russia
[7] Institute for General Genetics, Moscow, Russia
[8] Siberian Federal University, Krasnoyarsk, Russia
[9] Southern Scientific Center of the Russian Academy of Sciences, Rostov-on-Don, Russia
* Correspondence: mcmakarenko@yandex.ru

Received: date; Accepted: date; Published: date

Abstract: The current study provides insights into the flexibility of the mitochondrial genome in sunflower (*Helianthus annuus*) as well as the causes of ANN2 type cytoplasmic male sterility (CMS). *De novo* assembly of the mitochondrial genome of the male-sterile HA89(ANN2) sunflower line was performed using Illumina short reads. Analysis of CMS ANN2 mitochondrial DNA sequence revealed the following reorganization events: twelve rearrangements, seven insertions, and nine deletions. From comparison between coding sequences of the male-sterile line with the male-fertile line seven new transcriptionally active open reading frames (ORF) were established: *orf324, orf327, orf345, orf558, orf891, orf933, orf1197*, and deletion of *orf777*. Three of these ORF are presented by chimeric genes involving *atp6* (*orf1197*), *cox2* (*orf558*) and *nad6* (*orf891*). In addition, *orf558*, *orf891*, *orf1197* as well as *orf933* encode proteins containing membrane domain(s), making them the most likely candidate genes for ANN2 CMS type development. Although the CMS phenotype may be caused by simultaneous action of several candidate genes, we assume that *orf1197* plays the major role in male sterility development in ANN2. Comparison of the mtDNA organization in sunflower lines with different CMS sources also allowed identification of reorganization hot-spots in the mitochondrial genome of sunflower.

Keywords: sunflower, cytoplasmic male sterility (CMS), mitochondrial genome reorganizations, NGS sequencing

1. Introduction

Low substitution rate in genes along with considerable variability in size and structure are distinct features of plant mitochondrial genomes [1, 2]. Reorganization events in mitochondrial DNA (mtDNA) are primarily caused by disruption of a fragile equilibrium of intramolecular recombinations, maintained by nuclear-mitochondrial genomes interactions [3, 4]. The runaway recombination of mtDNA can lead to changes in gene content and expression patterns of mitochondria [5, 6]. The flowering plant mitogenomes carry genes for rRNAs, tRNA, subunits of the respiratory chain complexes, as well as genes for the ribosomal proteins (*rps* and *rpl*). Maturase-related protein gene (*matR*) and genes responsible for the biogenesis of cytochrome C (*ccmB*, *ccmC*, *ccmFc,* and *ccmFn*) are also part of the plant mitochondrial gene set [7]. Alterations in the transcription activity of mitochondrial genes can have profound adverse effects on the functionality of mitochondria and thus on different plant traits. Among phenotypic traits caused by

mitochondria impairment, special attention is devoted to cytoplasmic male sterility (CMS) [8]. Cytoplasmic male sterility (inability to produce or shed functional pollen) was described in more than 140 higher plant species [9]. As a result of mtDNA rearrangements, new open reading frames are created, possibly resulted in male sterility [10]. In turn, dominant nuclear-encoded restorer-of-fertility genes (Rf genes) can restore normal development of pollen. The CMS/Rf systems are important both for studying pollen development in plants and for commercial applications [11, 12]. Existence of the CMS trait in plants eliminates the need for laborious, manual emasculations for the directional crossing of plants, thus promoting utilization of hybrid breeding [13].

Comparing mtDNA configurations in sunflower is especially interesting, as more than 70 CMS sources have been described within a Food and Agriculture Organization (F.A.O.) program in sunflower [14], even though modern sunflower hybrid breeding predominantly relies on a single male-sterile cytoplasm, the so-called CMS PET1 [15]. CMS PET1 originates from an interspecific cross of *H. petiolaris* with *H. annuus* [16]. Molecular characterization of the CMS mechanism helps introducing new CMS sources into breeding programs. So far, the mtDNA of only a few CMS sources have been sufficiently characterized to be used in sunflower hybrid production. CMS can arise spontaneously in wild populations or in breeding lines after wide crosses, interspecific exchange of nuclear and/or cytoplasmic genomes, and mutagenesis [11]. It has been demonstrated that some CMS sources obtained from different inter- or intraspecific crosses showed the same mechanism of sterility formation as the PET1 CMS type [17]. Even though these CMS sources had different origins, they have the same mitochondrial genome organization indicating that some configurations may be preferentially maintained in sunflower [17].

Less is known about the spontaneously occurring CMS sources in *Helianthus annuus* [14]. ANN2 was derived from the wild sunflower population N517 in Texas [18]. In this population 40% of the plants were male-sterile [19]. However, ANN2 developed from this population by maintaining with HA89 or RHA265 lines, has 100% male-sterile progeny plants, which indicates a stable mitochondrial DNA configuration and absence of heteroplasmy. ANN2 and other spontaneously occurring CMS sources like ANN1, ANN3, and ANN4 maintained by RHA265 were hard to restore [20]. None of the tested maintainer and restorer lines of CMS PET1 were able to restore pollen production in CMS ANN1 and CMS ANN3. Only 12.5% and 15.8 % of all investigated lines showed restorer capacity towards CMS ANN2 and CMS ANN4, respectively, indicating very different CMS mechanisms compared to CMS PET1. Three restorer lines, Rf ANN2-PI 413178, Rf ANN2-P21 and Rf ANN2-RMAX1, carry a restorer-of-fertility gene for ANN2 [21, 22]. Besides, a suppressor gene S1 overpowering the restorer gene action has been described recently [23], thus making the CMS-Rf interactions in ANN2 source even more convoluted.

Previous mtDNA investigation of some spontaneously occurring CMS sources was based on Southern blot hybridizations with mitochondrial genes (*atp6*, *atp9*, *cob*, cox1, *cox2*, *cox3*, 18S/5S/nad5, *orfH522*, *orfH708, orfH873*) [24]. It revealed unique banding patterns for CMS cytoplasms, as well as provided clusterization based on CMS source origin and mechanism of action. Moreover, this study revealed unique banding patterns for 4 out of 10 probes for these cytoplasms [24]. Banding patterns allowed clusterization of CMS sources according to their origin and mechanism of action. For example, CMS ANN4 and ANN5 are similar to each other, and form a distinct group from CMS ANN1, ANN2 and ANN3 [24]. It was also shown that ANN1/ANN2/ANN3 mtDNA-type significantly differs from both the male-fertile sunflower cytoplasm and the PET1 CMS source [24].

Using the complete mitochondrial genome sequence of HA89(ANN2) CMS line of sunflower, we describe the first assembly of the CMS source ANN2, that occurred spontaneously in *Helianthus annuus*. The current study also provides insights into the flexibility of the sunflower mitochondrial genome by comparing different isonuclear male-sterile lines HA89(ANN2, MAX1, PET1, PET2) and the male-fertile line HA89, allowing identification of hot spots for rearrangements in the sunflower mtDNA. For the CMS mechanism in ANN2, new open reading frames were identified, which were transcriptionally active. ANN2 is remarkable as it might not only be an interesting CMS source for oilseed hybrid breeding but also for horticultural purposes, as it is difficult to restore. This is a highly desirable trait, since pollen production is usually not required nor required for ornamental sunflowers, except if the pollen color enhances the contrast with the florets.

2. Results

2.1 Rearrangements in the mitochondrial genome of the male-sterile line HA89(ANN2)

We assembled the complete mitochondrial genome of HA89 sunflower line with ANN2 cytoplasmic sterility type (MN175741.1). The master chromosome of HA89 (ANN2) consists of 306,018 bp (Figure 1), and it is 5,071 bp longer than the mtDNA genome of the male-fertile isonuclear line HA89 (MN171345.1).

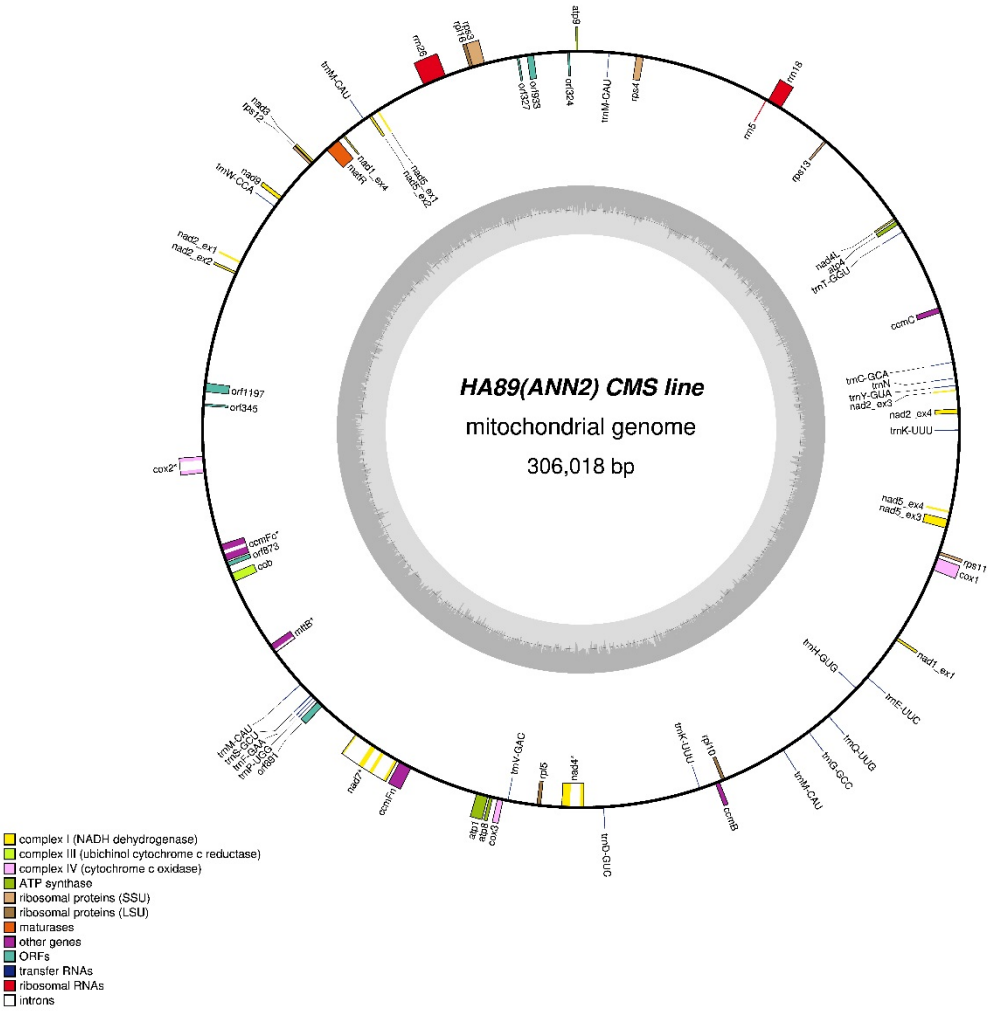

Figure 1. *Graphical mitochondrial genome map of HA89(ANN2) CMS line of sunflower.*

The HA89 (ANN2) mitochondrial genome has a wide range of rearrangements as compared to the male-fertile HA89 mitogenome. The summary of whole mitogenome alignment of sterile and fertile lines is presented in Table 1.

Table 1. *Alignment of the mitochondrial genomes of the male-fertile HA89 and the male-sterile HA89(ANN2) line*.

№ of alignment region	The alignment region length, bp	Positions in mtDNA of male-fertile HA89	Positions in mtDNA of male-sterile HA89(ANN2)	Orientation	% similarity	Localized genes

1	29196	1-29196	1-29204	Plus/Plus	99	*nad2_ex3, nad2_ex4, trnY, trnN, trnC, ccmC, trnT, atp4, nad4L*
2	557	33772-34328	78343-78899	Plus/Plus	99	-
3	1245	34329-35573	148163-149411	Plus/Minus	98	-
4	77441	36739-114179	217575-295553	Plus/Plus	95	*atp8, cox3, trnV, rpl5, nad4, trnD, trnK, ccmB, rpl10, trnM, trnG, trnQ, trnH, trnE, nad1+, cox1, nad5+*
5	41702	114180-155882	35657-77315	Plus/Minus	99	*atp9, trnM, rps4, rrn26, rrn5, rrn18, rps13, nad1_ex1, nad1_ex2, nad1_ex3*
6	4150	155883-160032	300892-305041	Plus/Plus	99	-
7	8584	160320-168903	129358-137946	Plus/Plus	99	*nad2_ex1, nad2_ex2, nad6**
8	21433	168906-190275	171388-192871	Plus/Minus	98	*nad6*, trnP, trnF, trnS, trnM, mttB, cob*
9	8158	194543-202700	163232-171387	Plus/Plus	99	*ccmFc, orf873*
10	24687	202701-227387	192915-217574	Plus/Minus	99	*atp1, ccmFn, nad7*
11	41505	227396-268900	87945-129446	Plus/Plus	99	*nad1+, rps3, rpl16, trnM, matR, nad3, rps12, nad9, trnW, nad5_ex3, nad5_ex4*
12	6029	269217-275245	141704-147732	Plus/Minus	99	*atp6**
13	12520	275536-288055	150723-163231	Plus/Plus	99	*nad1_ex4, cox2, nad2+,*
14	977	299971-300947	305042-306018	Plus/Plus	99	*trnK*

* genes, that had impaired sequence as the result of rearrangement

Mitochondrial genomes of ANN2 and the male-fertile HA89 share 14 complement regions, but their localizations and orientation may differ. We illustrated the localization of complement regions in a scheme with both genomes shown in linear forms in Figure 2. Since regions # 1 and # 14 in case of the circular molecule represent the same region, we classified the other twelve regions (#2-#13) as rearrangements.

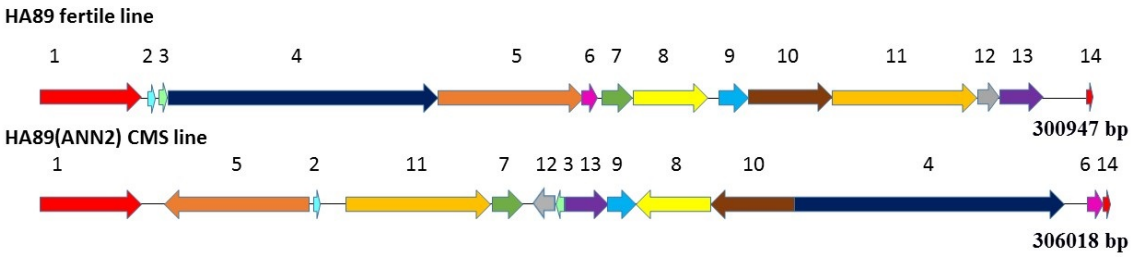

Figure 2. *The schematic illustration of homologous regions between mitochondrial genomes of HA89 male-fertile and HA89(ANN2) CMS lines. 1 – 29196 bp; 2 – 557 bp; 3 – 1245 bp; 4 – 77441 bp; 5 – 41702 bp; 6 – 4150 bp; 7 – 8584 bp; 8 – 21433 bp 9 – 8158 bp; 10 – 24687 bp; 11 – 41505 bp; 12 – 6029 bp; 13 – 12520 bp; 14 – 977 bp.*

In most cases, the rearrangements only resulted in a reversal of a genes direction or a change in genes order. However, the 8,584 bp (#7) and 21,433 bp (#8) rearrangements influenced the coding sequence of *nad6*, and the 6,029 bp rearrangement (#12) impaired *atp6*. The largest part of the *nad6* gene sequence (~ 88%) is in the 21,433 bp rearrangement while the 3' terminal part of *nad6* lies in the 8,584 bp rearrangement. As a result of the convergence of the 21,433 bp and 24,687 bp rearrangements (#8 and #10) in the mitochondrial DNA of HA89(ANN2) the new *nad6*-chimeric open reading frame - *orf891* – was created. The analysis of *orf891* transcription had ambiguous results: for instance, the transcription level was determined for both *nad6* (HA89) and *orf891* (HA89(ANN2)) while using primers to 5' identical sequence of their mRNA (Supplement 1). Nevertheless, using the same forward primer, but different reverse primers (Supplement 1) complementary to the 3' sequence of *nad6* and *orf891*, transcription was detected only for *nad6* (the fertile line), but not for *orf891* (CMS line). It is important to note that almost all the rearrangements found in mtDNA of HA89(ANN2) are accompanied by other types of genome reorganizations - deletions and insertions.

2.2 Deletions and insertions in mitochondrial genome of the male-sterile line HA89(ANN2)

In comparison with the male-fertile analog, we identified nine long (more than 100 bp) deletions in the mtDNA of HA89(ANN2): 287 bp, 290 bp, 299 bp, 316 bp, 583 bp, 1165 bp, 4204 bp, 4575 bp and 11901 bp (Table 2). Most of the deletions did not affect the protein coding sequences, except for the two deletions of 316 bp and 1165 bp. The 1165 bp deletion resulted in the total elimination of *orf777*, while the 316 bp deletion only affected a part of the *atp6* gene. Interestingly, in previous studies we also discovered the removal of *orf777* from the mitochondrial genomes of two other CMS lines - HA89(PET2) [25] and HA89(MAX1) [26].

Table 2. *Deletions (> 100 bp) localized in the mitochondrial genome of HA89(ANN2) CMS line.*

Deletion length, bp	Positions in mtDNA of the male-fertile line HA89	Deletion localization according to the male-fertile line HA89 genetic map	Deleted genes
287	160032-160319	*rps13-nad6*	-
290	275246-275535	*atp6-cox2*	-
299	56701-56999	*nad4-ccmB*	-
316	268901-269216	*nad9-atp6*	*atp6* (partial)
583	70338-70920	*rpl10-nad1*	-
1165	35574-36738	*nad4L-orf777-atp8*	*orf777*
4204	190339-194542	*cob-ccmFc*	-

4575	29197-33771	*nad4L-orf777*	-
11901	288070-299970	*cox2-nad2*	-

We have also detected seven long (more than 100 bp) insertions in the mtDNA of HA89(ANN2) CMS line: 430 bp, 1027 bp, 1310 bp, 3757 bp, 5338 bp, 6452 bp, 9045 bp (Table 3). As a result of these insertions in the mitochondrial DNA of HA89(ANN2), five new open reading frames have appeared: *orf324*, *orf327*, *orf345*, *orf558*, and *orf933*. We detected transcriptional activity of all five ORFs in plants of ANN2, but not in plants of the male-fertile line HA89.

Table 3. *Localization of insertions (> 100 bp) in the mitochondrial genome of HA89(ANN2) CMS line.*

Insertion in bp	Positions in mtDNA of HA89(ANN2)	New ORFs based on insertion	Homologies to
430	147733-148162	*orf1197**	*atp6*
1027	77316-78342	*orf324*	*orf285* (CMS PET2)
1310	149412-150722	*orf345*	
3757	137947-141703		
5338	295554-300891	*orf558*	*cox2*
6452	29205-35656		
9045	78900-87944	*orf327*, *orf933*	

* appeared as the result of several simultaneous reorganizations of mtDNA structure

A search for transmembrane domains (TD) revealed that the protein encoded by *orf558* contained single TD, and in the case of *orf933,* there are two TD. The *orf933* encoded protein did not show homology to other sunflower proteins in GeneBank, and only had limited similarity (40-60 amino acids) to hypothetical mitochondrial proteins with unknown functions in *Lactuca sativa* (PLY70338.1), *Salvia miltiorrhiza* (YP_008992338.1), *Beta vulgaris* (CBJ23356.1), etc. Forty-six amino acids of the N-terminus of the protein encoded by *orf558* matched the N-terminus of cytochrome c oxidase subunit 2 (*cox2* gene). Moreover, most of the amino acids that form the transmembrane domain in *orf558* protein are identical to those in COX2. However, sunflower cytochrome c oxidase subunit 2 has two TD and the protein encoded by *orf558* - only the single one (Figure 3). So *orf558* represents a chimeric *cox2* gene and could potentially play a role in CMS phenotype development.

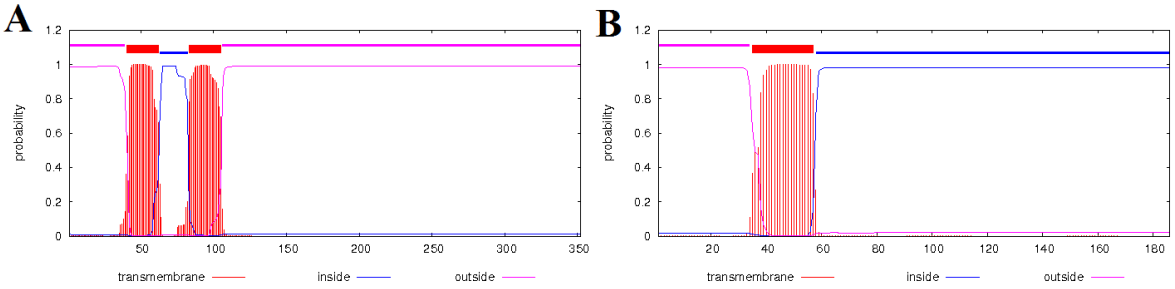

Figure 3. *The comparison of transmembrane domains of proteins encoded by cox2 (A) and orf558 (B).*

The most complex among the discovered ORFs in the mtDNA of HA89(ANN2) was *orf1197*, which has appeared from three simultaneous reorganization events involving a 316 bp deletion, 430 bp insertion, and 6029 bp rearrangement. The *orf1197* represents a chimeric *atp6* gene, with transcription activity specific for the CMS line HA89(ANN2). In sunflower, the *atp6* gene normally encodes a protein consisting of 351 aa, whereas the translation product of *orf1197* is 399 aa long. Both proteins share 251 identical amino acids in the C-terminus. Thus, the protein encoded by *orf1197* has all seven TD present in the C-terminus of a normal ATP synthase Fo subunit 6 (Figure 4).

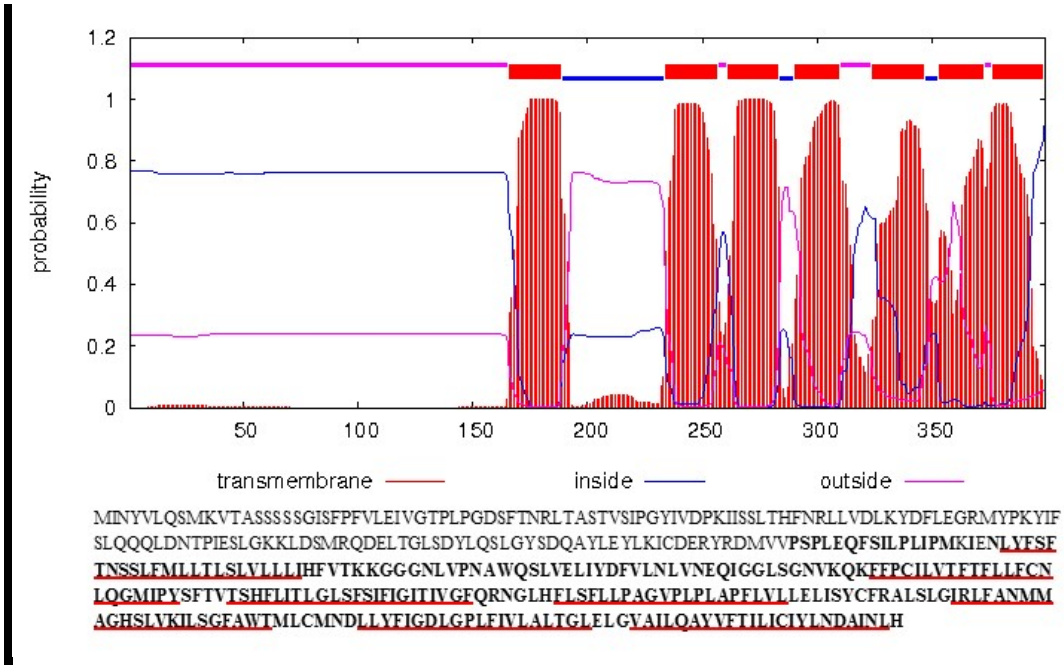

Figure 4. *Comparison of orf1197 and atp6 encoded proteins and prediction of transmembrane helices. Identical amino acids are in bold. Amino acids forming transmembrane domains are marked by red bars.*

3. Discussion

Recently, we had investigated complete mitochondrial DNA sequences for three CMS sources in sunflower: PET1, PET2, MAX1 [25, 26]. The comparison of HA89(ANN2) mitogenome with fertile mitochondrial genome assemblies [25, 27] and other HA89 sterile analogs provides insights into reorganizations of mtDNA associated with CMS phenotypes. While male-fertile lines (HA89, HA412) have only slight variations in mtDNA sequence [25], the mitogenomes of CMS sources (HA89(ANN2), HA89(PET1), (HA89(PET2), HA89(MAX1)) showed a significant difference as compared to their alloplasmic male-fertile analog. The complete mitochondrial genome of the male-fertile line HA89 adds up to 300,947 bp (MN171345.1), while HA89(PET1) has a size of 305,217 bp (MG735191.1), HA89(PET2) of 316,582 bp (MG770607.2), HA89(MAX1) of 295,586 bp (MH704580.1) and HA89(ANN2) of 306,018 bp (MN175741.1). The difference in the genomes sizes is due to several deletions and insertions. For instance, in the mtDNA of all investigated CMS sources, except HA89(PET1), a similar deletion in the *nad4L-orf777-atp8* region was observed. In the case of HA89(PET2), this is due to a 711 bp deletion, in HA89(MAX1) to a 978 bp deletion, HA89(ANN2) to a 1195 bp deletion. All these deletions resulted in removal of *orf777* from the mtDNA of the CMS lines. Another region enriched by deletions is the area between *cob-ccmFc*, here three overlapping deletions were detected: a 451 bp deletion in HA89(PET1), one of 3780 bp in HA89(PET2) and another one of 4204 bp in HA89(ANN2).

The agreement between locations of these deletions is not accidental. There are three 265 bp repeats in the sunflower mitochondrial genome, with following positions in mtDNA of male-fertile HA89 line: 36537-36801 (adjacent to *atp8*), 190074-190338 (next to *cob*), and 202902-202638 (*orf873-atp1*). These repeat regions are shown by red stars in Figure 5. Repeats represent common recombination points in the mtDNA molecules [4, 28]. Identification of small repeats involved in recombination is important because they influence maintenance and evolution of mitochondrial genomes [28]. Imbalance in the nuclear-mitochondrion relationship, that may occur in distant hybridizations, impairs the recombination of mtDNA sub-genomic molecules, therefore, leading to reorganization in the mitochondrial master chromosome. For instance, in HA89(PET1) a deletion, an insertion, and inversion were mentioned in the *cob-atp1* region, directly between two repeats (Figure 5). In HA89(PET2) there were also several rearrangements in "hot-spots", resulting in a new gene cluster *cob-atp8-cox3* formation, as well as translocation of the *ccmFc-orf873-atp1* gene cluster into the *nad4L-orf777* region, combined with deletion and a huge insertion (Figure 5). In mtDNA of both HA89(MAX1) and HA89(ANN2) lines a specific *atp1-atp8-cox3* genes order was created, while in MAX1 CMS source the *ccmFc-orf873* translocated into the *nad4L-orf777* region (with additional deletion and insertion in this region). In the case of ANN2, the *ccmFc-orf873* region is located next to the *cob* gene (Figure

5). Thus, we have established that these three 265 bp repeats represent reorganization hot-spots in the mitochondrial genome of sunflower.

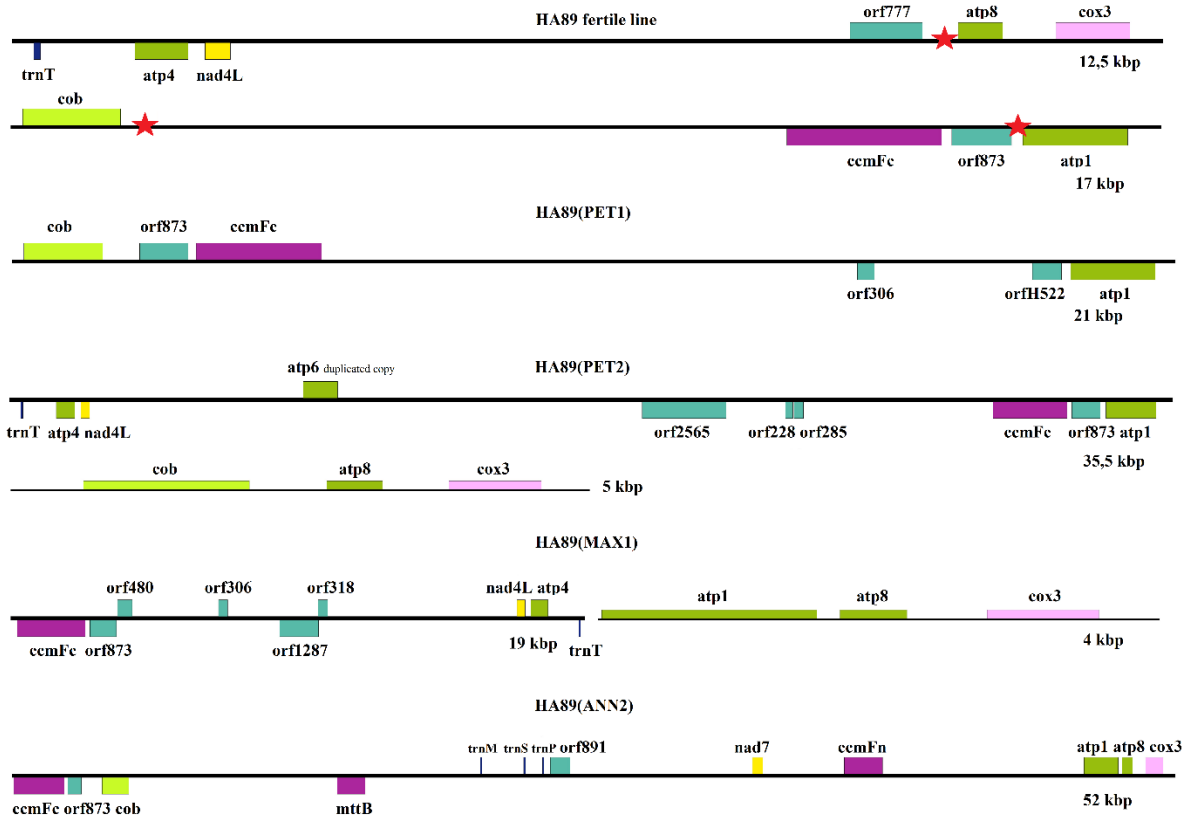

Figure 5. *The reorganizations in CMS lines that revealed nearby 265 bp repeats (red stars).*

Considering the insertions discovered in the HA89(ANN2), we also observed similarity between insertions of different CMS sources in sunflower. For instance, about 85% of the 1,027 bp insertion sequence (ANN2) is complementary to the part of the 15,885 bp insertion (PET2). The 3,757 bp insertion (ANN2) contains 1,959 nucleotides identical to the 15,885 bp insertion (PET2) and 1,215 – to the 5,272 bp insertion (MAX1). Also, 2,343 bp of 5,338 bp insertion in ANN2 are similar to another region of sunflower mtDNA proximal to the *cob* gene (position 185,987-188,330 in the mtDNA of HA89 fertile line), so this sequence is duplicated in the mitochondrial genome of HA89(ANN2) CMS line. About 10 % of 6,452 bp insertion (ANN2) is complementary to both the 5,050 bp (PET2) and the 5,272 bp (MAX1) insertions. As well as 1,158 bp of the 9,044 bp insertion (ANN2) are identical to 5,050 bp and 15,885 bp insertions (PET2) and 5,272 bp insertion (MAX1).

Although there are similarities in deletions, insertions, and rearrangements between the mitochondrial genome of HA89(ANN2) line and other CMS lines, the discovered ORFs were different. We summarized the data about all identified ORFs in Table 4. The ORFs encoding proteins with similarity to other mitochondrial proteins and especially having transmembrane domains are of particular interest, since the chimeric proteins with TD most often cause the CMS phenotype [10, 29]. In HA89(ANN2) mtDNA characterized by three new transcriptionally active ORFs, encoding proteins with TD – *orf558* (one TD), *orf933* (two TD), *orf1197* (seven TD). The *orf933* shows no homology to other sunflower genes, while *orf558* represents a chimeric *cox2* gene, and *orf1197* a chimeric *atp6* gene. It is difficult to estimate the exact contribution of *orf558*, *orf933*, *orf1197* to the development of male sterility phenotype in ANN2. However, the possibility of the involvement of more than one open reading frame might be one explanation why ANN2 is so difficult to restore. On the other hand, the presence of a suppressor gene S1 discovered by [23] might be the reason for low rates of fertility restoration of the ANN2 CMS source. According to previous studies, the chimeric *atp6* genes or new ORFs that are co-transcribed with *atp6* most often cause CMS phenotypes in flowering plants [10, 30-33]. Therefore, we suggest that *orf1197* is the major CMS candidate gene for ANN2 CMS source. Moreover, chimeric *atp6* genes were also identified in MAX1 [26] and CMS3/ANT1 [34] CMS types of sunflower. In CMS lines, chimeric *atp6* genes encode N-terminal extended proteins compared to the normal

ATP synthase subunit 6 (351 aa): ANN2 - 399 aa, MAX1- 429 aa, (AYV91168.1), CMS3/ANT1 - 437 aa (CAA57790.1). Moreover, 397 of 399 amino acids in *orf1197* protein are identical to the chimeric *ATP6* of CMS3/ANT1 line, and therefore this protein represents a shorter version of this 437 aa long protein. Such similarities support our hypothesis about the importance of *orf1197* in shaping the CMS phenotype in ANN2.

Table 4. *Summary of transcriptionally active open reading frames in the mitochondrial genome of fertile and CMS lines*

Fertile mtDNA	CMS ANN2	CMS MAX1	CMS PET2	CMS PET1
-	-	-	***orf228***	-
-	-	-	***orf285***	-
-	-	*orf306*	-	*orf306*
-	*orf324*	-	-	-
-	*orf327*	-	-	-
-	*orf345*	-	-	-
-	-	*orf480*	-	-
-	-	-	-	***orfH522***
-	***orf558***	-	-	***-***
-	-	*orf645*	*orf645*	-
orf777	-	-	-	*orf777*
-	***orf891***	-	-	-
-	***orf933***	-	-	-
-	***orf1197***	-	-	-
-	***-***	***orf1287***	-	-
-	-	-	*orf2565*	-

In bold: ORFs encoding proteins with transmembrane domains

Orf558 (as a chimeric *cox2* gene) might also cause the cytoplasmic male sterility in ANN2. In other plants species, modified *cox2* sequences seem to be involved in the male sterility phenotype. For instance, the CMS specific *pcf* gene in petunia is composed from sequences of the 5' portion of *atp9*, segments of *cox2* and large region of unknown origin - *urfS* [35]. In wild beets, the CMS-associated ORF (*orf129*) shows homology to the 5' flanking and coding sequence of *cox2* [36]. In mitochondrial DNA of the inap CMS source of *Brassica napus*, which was created by somatic hybridization with *Isatis indigotica*, there was detected a novel *cox2-2* gene, which represents recombination of the *cox2* of woad and *cox2-2* of rapeseed [37].

Another unique feature of HA89(ANN2) mitogenome is formation of *orf891*. According to ORF prediction, 3' elongation of the *nad6* gene (*orf891*) may occur. However, the cDNA analysis does not agree with the genomic data. Perhaps due to the *nad6* mRNA editing instead of 3' elongated transcript, the shorter one is formed. The *nad6* transcript length heteromorphism was also observed in *Mimulus guttatus x M. nasutus* hybrids with CMS phenotype [38]. Both fertile and sterile hybrids have a single copy of *nad6* gene, and the divergence in mRNA length was observed only for CMS plants [38]. Therefore, in our future studies, we plan to provide the mRNA analysis of hybrids with ANN2 cytoplasm, but with restored fertility. It will help to evaluate the role of each revealed new ORF in ANN2 CMS phenotype development.

4. Materials and Methods

4.1. Plant material

The CMS line HA89(ANN2) of sunflower was obtained from the genetic collection of the N. I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR, Russia). The original source of ANN2 sterile cytoplasm was obtained by Serieys in 1984 [18]. All sunflower lines were grown in regularly irrigated pots in the growth chamber KBWF 720 (Binder, Germany) at the following growth conditions: temperature - 26 °C, humidity - 70%, photoperiod - 10/14 h (dark/light).

4.2. Mitochondrial DNA extraction, NGS library preparation and sequencing

First, the organelle fraction from the leaves of 14-days-old sunflower seedlings was isolated as described by [39]. This significantly reduced the amount of nuclear DNA. Then DNA extraction was performed with PhytoSorb kit (Syntol, Russia), according to the manufacturer's protocol. Equal amounts of DNA from seven plants were mixed, and we used 1 ng of DNA pull for the NGS library preparation step. The library was made with Nextera XT DNA Library Prep Kit (Illumina, Mountain View, CA, USA), following the guidelines of Illumina. The quality of the library was evaluated using Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). The library was quantified at the Qubit fluorimeter (Invitrogen, USA) and by qPCR, then diluted up to the concentration of 8 pM. Sequencing was performed on two different Illumina sequencing platforms: HiSeq 2000 using TruSeq SBS Kit v3-HS 200-cycles and MiSeq using MiSeq Reagent Kit v2 500-cycles (Illumina, Mountain View, CA, USA). A total number of 3,063,836 100-bp paired reads and 3,305,268 250-bp paired reads were generated.

4.3. Mitochondrial genome assembly and annotation

Quality control of reads was done using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimming of adapter-derived and low quality (Q-score below 25) reads was performed with Trimmomatic software [40]. For the contigs generation, we used SPAdes Genome Assembler v.3.10.1 [41]. The whole mitochondrial genome was manually assembled using scaffolds based on high coverage (depths >100) contigs (length > 1 kbp) and available bridge contigs (length = 0.3-1 kbp). The genome assembly was validated by remapping the initially obtained reads using Bowtie 2 v.2.3.3 [42]. All observed rearrangements were verified by PCR analysis. For variant calling, we used samtools/bcftools software [43] and manually revised polymorphic sites using the IGV tool [44].

The mitochondrial genome was annotated with MITOFY [45], BLAST tool [46] and ORFfinder (https://www.ncbi.nlm.nih.gov/*orf*finder). Using GeSeq [47] we provided comparison of our annotation with the current reference annotations (NCBI accessions NC_023337.1, CM007908.1) of sunflower mitochondrial genome. Graphical genome maps were generated using the OGDRAW tool v.1.3.1 [48]. The prediction of transmembrane domains was made with TMHMM Server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/).

4.4 The mRNA detection

The RNA was extracted from leaves of seven 28-days-old sunflower plants using a guanidinium thiocyanate-phenol-chloroform based method with ExtractRNA reagent kit (Evrogen, Russia). Quality and concentration of the RNA were evaluated with the Qubit fluorimeter (Invitrogen, USA) and the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). 0.5 μg of total RNA was treated with DNAse I (Thermo Fisher Scientific, USA) and then cDNA was synthesized using the MMLV RT kit (Evrogen, Russia) with random primers. As a negative control for each DNAse treated mRNA sample, there was performed the same reverse transcription protocol but without MMLV enzyme. The quantitative PCR was performed with EvaGreen based RT-PCR kit (Syntol, Russia) on Rotor-Gene 6000 (Corbett Research, Australia). A summary of all primer sequences is given in Supplementary Table 1.

5. Conclusions

The assembly of CMS ANN2 mitochondrial genome (HA89(ANN2) line) revealed several rearrangements, insertions and deletions as well as seven new open reading frames: *orf324*, *orf327*, *orf345*, *orf558*, *orf891*, *orf933*, and *orf1197*. Transcripts were detected for all seven new open reading frames in CMS ANN2, but not in the fertile cytoplasm. Only *orf558*, *orf891*, *orf933*, *orf1197* encoded proteins containing

membrane domains, making them the most likely CMS candidate genes for the ANN2 source. *Orf1197* represents a chimeric *atp6* gene and presumably plays a major role in the CMS phenotype development associated with ANN2. However, CMS ANN2 may be caused by the simultaneous action of several candidate genes. Hot spots for rearrangements were identified, and we propose that they influence maintenance and evolution of the mitochondrial genome in sunflower.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: The primers sets used for transcription activity analysis.

Author Contributions: conceptualization, Azarin K. and Gavrilova V.; **methodology,** Makarenko M. and Kornienko I.; **software,** Tatarinova T. and Logacheva M.; **validation,** Makarenko M., Azarin K., Kornienko I. and Horn R.; **formal analysis,** Makarenko M. and Tatarinova T.; **investigation,** Makarenko M., Azarin K. and Logacheva M.; **resources,** Gavrilova V.; **data curation,** Makarenko M. and Logacheva M.; **writing—original draft preparation,** Makarenko M., Tatarinova T. and Horn R.; **writing—review and editing,** Tatarinova T. and Horn R.; **visualization,** Horn R.; supervision, Usatov A. and Horn R.; **project administration,** Usatov A. and Gavrilova V.; **funding acquisition,** Usatov A. and Logacheva M.

Funding: The study was supported by the Ministry of Education and Science of Russia project no. 6.929.2017/4.6. The NGS sequencing was provided with the support of budgetary subsidy to IITP RAS (Laboratory of Plant Genomics 0053-2019-0005). Analytical work was carried out on the equipment of centers for collective use of Southern Federal University "High Technology."

Acknowledgments: We are grateful to the cooperators of Northern Crop Science Laboratory (Fargo, USA) for providing seeds of HA89 alloplasmic lines of sunflower to the genetic collection of the N. I. Vavilov Institute of Plant Genetic Resources.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Yang, J.; Liu G., Zhao N., Chen S., Liu D., Ma W., Hu Z., Zhan M. Comparative mitochondrial genome analysis reveals the evolutionary rearrangement mechanism in *Brassica*. *Plant Biol*. **2015,** *18*, 527-536. DOI: 10.1111/plb.12414.
2. Wang, S.; Song Q., Li S., Hu Z., Dong G., Song C., Huang H., Liu Y. Assembly of a Complete Mitogenome of *Chrysanthemum nankingense* Using Oxford Nanopore Long Reads and the Diversity and Evolution of Asteraceae Mitogenomes. *Genes* **2018**, *9*, 547. DOI: 10.3390/genes9110547.
3. Liu, H.; Cui P., Zhan K., Lin Q., Zhuo G., Guo X., Ding F., Yang W., Liu D., Hu S., Yu J., Zhang A. Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. *BMC Genomics* **2011**, *12*, 163. DOI: 10.1186/1471-2164-12-163.
4. Storchova, H.; Stone J.D., Sloan D.B., Abeyawardana O.A.J., Muller K., Walterova J., Pazoutova M. Homologous recombination changes the context of *Cytochrome b* transcription in the mitochondrial genome of *Silene vulgaris* KRA. *BMC Genomics* **2018**, *19*, 874. DOI: 10.1186/s12864-018-5254-0.
5. Marechal, A.; Brisson N. Recombination and the maintenance of plant organelle genome stability. *New Phytologist* **2010**, *186*, 299–317. DOI: 10.1111/j.1469-8137.2010.03195.x.
6. Gualberto, J.M.; Mileshina D., Wallet C., Niazi A.K., Weber-Lotfi F., Dietrich A. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* **2013**, *100*, 107-120. DOI: 10.1016/j.biochi.2013.09.016.
7. Zervas, A.; Petersen, G., Seberg, O. Mitochondrial genome evolution in parasitic plants. *BMC Evol. Biol*. **2019**, *19*, 87. DOI: 10.1186/s12862-019-1401-8.
8. Ding, B.; Hao M., Mei D., Zaman Q.U., Sang S., Wang H., Wang W., Fu L., Cheng H., Hu Q. Transcriptome and Hormone Comparison of Three Cytoplasmic Male-sterile Systems in *Brassica napus*. *Int. J. Mol. Sci*. **2018**, *19*, 4022. DOI: 10.3390/ijms19124022.
9. Laser, K.D.; Lersten, N.R. Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. *Bot. Rev*. **1972**, *33*, 337–346.
10. Horn, R.; Gupta K.J., Colombo N. Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* **2014**, *19*, 198-205. DOI: 10.1016/j.mito.2014.04.004.
11. Chen, L.; Liu Y.-G. Male sterility and fertility restoration in crops. *Annu. Rev. Plant Biol*. **2014**, *65*, 579-606. DOI: 10.1146/annurev-arplant-050213-040119.

12. Wu, Z.; Hu K., Yan M., Song L., Wen J., Ma C., Shen J., Fu T., Yi B., Tu J. Mitochondrial genome and transcriptome analysis of five alloplasmic male-sterile lines in *Brassica juncea*. *BMC Genomics* **2019**, *20*: 348. DOI: 10.1186/s12864-019-5721-2.
13. Chen, Z.; Zhao N., Li S., Grover C.E., Nie H., Wendel J.F., Hua J. Plant Mitochondrial Genome Evolution and Cytoplasmic Male Sterility. *Crit*. *Rev. Plant*. *Sci*. **2017**, *36*, 55–69. DOI: 10.1080/07352689.2017.1327762.
14. Serieys, H. Identification, study and utilisation in breeding programs of new CMS sources in the FAO subnetwork. In Proceedings of the 2005 Sunflower Subnetwork Progress Report, Novi Sad, Serbia and Montenegro, FAO: Rome, Italy: 47–53, 17–20 July 2005.
15. Vear, F. Changes in sunflower breeding over the last fifty years. *OCL*. **2016**, *23*, D202. DOI: 10.1051/ocl/2016006.
16. Leclercq, P. Une sterilite male chez le tournesol. *Ann*. *Amelior. Plant*. **1969**, *19*, 99–106.
17. Horn, R.; Friedt W. CMS sources in sunflower: different origin but same mechanism? *Theor. Appl*. *Genet*. **1999**, *98*, 195-201. DOI: 10.1007/s001220051058.
18. Serieys, H. Report on the activities of the F.A.O. working group: "Identification, study and utilization in breeding programs of new CMS sources", for the period 1991-1993. *HELIA* **1994**, *17*, 93-102.
19. Serieys, H.; Vincourt, P. Caracterisation de nouvelles sources de sterility male chez le tournesol. Les Colloques de lNRA Paris. **1987**,*45*, 53-64. INRA, Paris.
20. Horn, R.; Friedt W. Fertility restoration of new CMS sources in sunflower (*Helianthus annum* L.). *Plant Breeding* **2006**, *116*, 317-322. DOI: 10.1111/j.1439-0523.1997.tb01005.x.
21. Jan, C.C. Cytoplasmic male sterility in two wild *Helianthus annuus* L. accessions and their fertility restoration. *Crop Sci*. **2000**, *40*, 1535-1538. DOI: 10.2135/cropsci2000.4061535x.
22. Jan, C.C. Silencing of fertility restoration genes in sunflower. *HELIA* **2003**, *26*, 1-6. DOI: 10.2298/hel0339001j.
23. Liu, Z.; Long Y., Xu S.S., Seiler G., Jan C.C. Unique fertility restoration suppressor genes for male-sterile CMS ANN2 and CMS ANN3 cytoplasms in sunflower (*Helianthus annuus* L.) *Mol*. *Breeding* **2019**, *39*, 22. DOI: 10.1007/s11032-018-0922-y.
24. Horn, R. Molecular diversity of male sterility inducing and male-fertile cytoplasms in the genus *Helianthus*. *Theor. Appl*. *Genet*. **2002**, *104*, 562-570. DOI: 10.1007/s00122-001-0771-6.
25. Makarenko, M.; Kornienko I., Azarin K., Usatov A., Logacheva M., Markin N., Gavrilova V. Mitochondrial genomes organization in alloplasmic lines of sunflower (*Helianthus annuus*) with various types of cytoplasmic male sterility. *PeerJ* **2018**, *6*, e5266. DOI: 10.7717/peerj.5266.
26. Makarenko, M.S.; Usatov A.V., Tatarinova T.V., Azarin K.V., Logacheva M.D., Gavrilova V.A., Horn R. Characterization of the mitochondrial genome of the MAX1 type of cytoplasmic male-sterile sunflower. *BMC Plant Biol*. **2019**, *19*, 51. DOI: 10.1186/s12870-019-1637-x.
27. Grassa, C.J.; Ebert D.P., Kane N.C., Rieseberg L.H. Complete Mitochondrial Genome Sequence of Sunflower (*Helianthus annuus* L.). *Genome Announc*. **2016**, *4*, e00981-16. DOI: 10.1128/genomeA.00981-16.
28. Wynn, E.L.; Christensen A.C. Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. *G3 (Bethesda)* **2019**, *9*, 549-559. DOI: 10.1534/g3.118.200948.
29. Mower, J.P.; Case A.L., Floro E.R., Willis J.H. Evidence against Equimolarity of Large Repeat Arrangements and a Predominant Master Circle Structure of the Mitochondrial Genome from a Monkeyflower (*Mimulus guttatus*) Lineage with Cryptic CMS. *Genome Biol*. *Evol*. **2012**, *4*, 670-86. DOI: 10.1093/gbe/evs042.
30. Yamamoto, M.P.; Kubo T., Mikami T. The 5'-leader sequence of sugar beet mitochondrial atp6 encodes a novel polypeptide that is characteristic of Owen cytoplasmic male sterility. *Mol*. *Gen*. *Genomics*. **2005**, *273*, 342-349. DOI: 10.1007/s00438-005-1140-y.
31. Kim, D.H.; Kim B.D. The organization of mitochondrial atp6 gene region in male-fertile and CMS lines of pepper (*Capsicum annuum* L.). *Curr. Genet*. **2006**, *49*, 59–67. DOI: 10.1007/s00294-005-0032-3.
32. Jing, B.; Heng S., Tong D., Wan Z., Fu T., Tu J., Ma C., Yi B., Wen J., Shen J. A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. *J*. *Exp*. *Bot*. **2012**, *63*, 1285–1295. DOI:10.1093/jxb/err355.
33. Tan, G.F.; Wang F., Zhang X.Y., Xiong A.I. Different lengths, copies and expression levels of the mitochondrial *atp6* gene in male-sterile and fertile lines of carrot (*Daucus carota* L.). *Mitochondrial DNA Part A* **2017**, *29, 446-454*. DOI: 10.1080/24701394.2017.1303492.
34. Spassova, M.; Moneger F., Leaver C.J., Petrov P., Atanassov A., Nijkamp H.J., Hille J. Characterisation and expression of the mitochondrial genome of a new type of cytoplasmic male-sterile sunflower. *Plant Mol*. *Biol*. **1994**, *26*, 1819-1831. DOI: 10.1007/BF00019495.
35. Young, E.G; Hanson M.R. A fused mitochondrial gene associated with cytoplasmic male sterility is developmentally regulated. *Cell* **1987**, *50*, 41-49. DOI: 10.1016/0092-8674(87)90660-x.

36. Yamamoto, M.P.; Shinada H., Onodera Y., Komaki C., Mikami T., Kubo T. A male sterility-associated mitochondrial protein in wild beets causes pollen distribution in transgenic plants. *Plant J*. **2008**, *54*, 1027-1036. DOI: 10.1111/j.1365-313X.2008.03473.x.
37. Kang, L.; Li P., Wang A., Ge X., Li Z. A Novel Cytoplasmic Male Sterility in *Brassica napus* (inap CMS) with Carpelloid Stamens via Protoplast Fusion with Chinese Woad. *Front*. *Plant Sci*. **2017**, *8*, 529. DOI: 10.3389/fpls.2017.00529.
38. Case, A.L.; Willis J.H. Hybrid male sterility in *Mimulus* (*Phrymaceae*) is associated with a geographically restricted mitochondrial rearrangement. *Evolution* **2008**, *62*, 1026-1039. DOI: 10.1111/j.1558-5646.2008.00360.x.
39. Makarenko, M.S.; Usatov A.V., Markin N.V., Azarin K.V., Gorbachenko O.F., Usatov N.A. Comparative Genomics of Domesticated and Wild Sunflower: Complete Chloroplast and Mitochondrial Genomes *OnLine J*. *Biol*. *Sci*. **2016**, *16*, 71-75. DOI: 10.3844/ojbsci.2016.71.75.
40. Bolger, A.M.; Lohse M., Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114-2120. DOI: 10.1093/bioinformatics/btu170.
41. Nurk, S. et al. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. *In Research in Computational Molecular Biology. RECOMB 2013. Lecture Notes in Computer Science*; Editor 1, Deng M., Editor 2, Jiang R., Editor 3, Sun F., Editor 4, Zhang X.; Springer: Berlin, Heidelberg, **2013**, 7821, 158-170. DOI: 10.1007/978-3-642-37195-0_13.
42. Langmead, B.; Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nat*. *Methods* **2012**, *9*, 357-359. DOI: 10.1038/nmeth.1923.
43. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987-2993. DOI: 10.1093/bioinformatics/btr509.
44. Thorvaldsdóttir, H.; Robinson J.T., Mesirov J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. **2013**, *14*, 178-192. DOI: 10.1093/bib/bbs017.
45. Alverson, A.J.; Wei X., Rice D.W., Stern D.B., Barry K., Palmer J.D. Insights into the Evolution of Mitochondrial Genome Size from Complete Sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol*. *Biol*. *Evol*. **2010**, *27*, 1436-1448. DOI: 10.1093/molbev/msq029.
46. Altschul, S.F.; Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. *J*. *Mol*. *Biol*. **1990**, *215*, 403-410. DOI: 10.1016/S0022-2836(05)80360-2.
47. Tillich, M.; Lehwark P., Pellizzer T., Ulbricht-Jones E.S., Fischer A., Bock R., Greiner S. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. **2017**, *45*, W6-W11. DOI: 10.1093/nar/gkx391.
48. Greiner, S.; Lehwark P., Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res*. **2019**, 47, W59-W64. DOI: 10.1093/nar/gkz238