A peer-reviewed version of this preprint was published in PeerJ on 23 July 2018.

Makarenko MS, Kornienko IV, Azarin KV, Usatov AV, Logacheva MD, Markin NV, Gavrilova VA. 2018. Mitochondrial genomes organization in alloplasmic lines of sunflower (*Helianthus annuus* L.) with various types of cytoplasmic male sterility. PeerJ 6:e5266
https://doi.org/10.7717/peerj.5266

Mitochondrial genomes organization in alloplasmic lines of sunflower (*Helianthus annuus*) with various types of cytoplasmic male sterility

Maksim Makarenko Corresp., [1] , **Igor Kornienko** [1, 2] , **Kirill Azarin** [1] , **Alexander Usatov** [1] , **Maria Logacheva** [3] , **Nicolay Markin** [1] , **Vera Gavrilova** [4]

[1] Southern Federal University, Rostov-on-Don, Russia

[2] Southern scientific center of the Russian Academy of Sciences, Rostov-on-Don, Russia

[3] Moscow State University, Belozersky Institute of Physical and Chemical Biology, Moscow, Russia

[4] The N.I. Vavilov All Russian Institute of Plant Genetic Resources, Saint Petersburg, Russia

Corresponding Author: Maksim Makarenko
Email address: mcmakarenko@yandex.ru

Background. Cytoplasmic male sterility (CMS) is a common phenotype in higher plants, which often is associated with rearrangements in mitochondrial DNA (mtDNA), and is widely used to produce hybrid seeds in a variety of valuable crop species. The CMS phenomenon investigations are also promote understanding of a fundamental issue of nuclear-cytoplasmic interactions in the ontogeny of higher plants. In the present study, we analyzed the structural changes in mitochondrial genomes of three alloplasmic lines of sunflower (Helianthus annuus). The investigation was focused on CMS line PET2, as there are very few reports about its mtDNA organization.

Methods. The NGS sequencing, *de novo* assembly, and annotation of sunflower mitochondrial genomes were performed. The comparative analysis of mtDNA of HA89 fertile line and two HA89 CMS lines (PET1, PET2) occurred.

Results. The mtDNA of the HA89 fertile line was almost identical to the HA412 line (NC_023337). The comparative analysis of HA89 fertile and CMS (PET1) analog mitochondrial genomes revealed 11852 bp inversion, 4732 bp insertion, 451 bp deletion and 18 variant sites. In mtDNA of HA89 (PET2) CMS line 77 kb translocation, 711 bp and 3780 bp deletions, as well as 1558 bp, 5050 bp, 14330 bp insertions were determined. There are also revealed 83 polymorphic sites sites in the PET2 mitochondrial genome, as compared with the fertile line

Discussion. Among the revealed rearrangements the 1558 bp insertion resulted in new open reading frames formation - o*rf228* and *orf246*. The o*rf228* and *orf246* could be the main reason for the development of PET2 CMS phenotype, whereas the role of other mtDNA reorganizations in CMS formation is negligible.

Mitochondrial genomes organization in alloplasmic lines of sunflower (Helianthus annuus) with various types of cytoplasmic male sterility

Maksim S. Makarenko[1], Igor V. Kornienko[1,2], Kirill V. Azarin[1], Alexander V. Usatov[1], Maria D. Logacheva[3], Nikolay V. Markin[1], Vera A. Gavrilova[4]

[1] Southern Federal University, Rostov-on-Don, Russia

[2] Southern scientific center of the Russian Academy of Sciences, Rostov-on-Don, Russia

[3] Lomonosov Moscow State University, Belozersky Institute of Physical and Chemical Biology, Moscow, Russia

[4] The N.I. Vavilov All Russian Institute of Plant Genetic Resources, Saint Petersburg, Russia

Corresponding Author: Maksim Makarenko

Email address: mcmakarenko@yandex.ru

Abstract

Background. Cytoplasmic male sterility (CMS) is a common phenotype in higher plants, which often is associated with rearrangements in mitochondrial DNA (mtDNA), and is widely used to produce hybrid seeds in a variety of valuable crop species. The CMS phenomenon investigations are also promote understanding of a fundamental issue of nuclear-cytoplasmic interactions in the ontogeny of higher plants. In the present study, we analyzed the structural changes in mitochondrial genomes of three alloplasmic lines of sunflower (Helianthus annuus). The investigation was focused on CMS line PET2, as there are very few reports about its mtDNA organization.

Methods. The NGS sequencing, *de novo* assembly, and annotation of sunflower mitochondrial genomes were performed. The comparative analysis of mtDNA of HA89 fertile line and two HA89 CMS lines (PET1, PET2) occurred.

Results. The mtDNA of the HA89 fertile line was almost identical to the HA412 line (NC_023337). The comparative analysis of HA89 fertile and CMS (PET1) analog mitochondrial genomes revealed 11852 bp inversion, 4732 bp insertion, 451 bp deletion and 18 variant sites. In

mtDNA of HA89 (PET2) CMS line 77 kb translocation, 711 bp and 3780 bp deletions, as well as 1558 bp, 5050 bp, 14330 bp insertions were determined. There are also revealed 83 polymorphic sites sites in the PET2 mitochondrial genome, as compared with the fertile line

Discussion. Among the revealed rearrangements the 1558 bp insertion resulted in new open reading frames formation - o*rf228* and *orf246*. The o*rf228* and *orf246* could be the main reason for the development of PET2 CMS phenotype, whereas the role of other mtDNA reorganizations in CMS formation is negligible.

Introduction

In plants, the cytoplasmic male sterility (CMS) is a phenomenon of interaction between mitochondrial and nuclear genomes, in which microsporogenesis disorders occur (Touzet & Meyer, 2014). All known natural CMSs, as well as most of artificially obtained, are characterized by a special type of mitochondrial DNA (mtDNA) with numerous structural rearrangements as compared to the mtDNA of the fertile plants of the same species (Ivanov & Dymshits, 2007; Horn, Gupta & Colombo, 2014; Garayalde et al. 2015). The mitochondrial genomes of higher plants have comparatively large size with a multitude of noncoding and repetitive sequences and present the complex of subgenomes structures (Chen & Liu, 2014). Precisely these features of plant mtDNA promote a large number of recombination events, leading to the appearance of new sequences and new open reading frames, which in turn often results in CMS development (Horn, Gupta & Colombo, 2014; Touzet & Meyer, 2014).

Natural CMS forms are described in more than 150 species of flowering plants (Fujii & Toriyama, 2009). Most CMS sources in crops are obtained on the basis of interspecific hybridization. The first CMS in sunflower was discovered by P. Leclercq at the interspecific hybridization between *Helianthus petiolaris* Nutt (PET1) and *Helianthus annuus* L (Leclercq, 1969). Comparison of mitochondrial DNA organization of fertile line and the male-sterile line carrying the PET1 cytoplasm revealed the presence of an 11-kb inversion and 5-kb insertion (Siculella & Palmer, 1988; Kohler et al., 1991). Such rearrangement of the mitochondrial genome resulted in the appearance of a new open reading frame (*orfH522*) in the 3'-flanking region of the *atp1* gene encoding the alpha subunit of mitochondrial F1 ATPase. A new *orfH522* is co-transcribed with the *atp1* gene as a polycistronic mRNA (Moneger, Smart & Leaver, 1994). The use of specific antibodies to the product of *orfH522* gene has shown that *orfH522* encodes

the 16-kDa protein, represents the only difference between the in mitochondria translation products of fertile and CMS lines (Horn et al., 1996). The 16-kDa protein is synthesized in all tissues of the plant. It is embedded in the mitochondrial membranes and is believed to disturb its integrity (Horn et al., 1996). Expression of *orfH522* in tapetum cells leads to their premature apoptosis. Due to the release of cytochrome C from the mitochondria, which activates the proteolytic enzyme cascade, eventually leading to the degradation of nuclear DNA and cell death (Balk & Leaver, 2001; Sabar et al., 2003). Interesting to note, that even stable transgenic CMS tobacco lines carrying the *orfH522* gene were obtained (Nizampatnam et al., 2009). In the presence of dominant nuclear restorer gene (Rf) in sunflower genome, fertility is restored due to the anther-specific lowering of the co-transcript of *orfH522* and the *atp1* gene (Moneger, Smart & Leaver, 1994; Horn, 2003). A possible mechanism leading to a reduction in the number of the chimeric *atp1-orfH522* transcripts by the restore gene is polyadenylation of RNA matrices, which causes accelerated degradation of RNA molecules by the ribonuclease (Gagliardi & Leaver, 1999).

The CMS-Rf system is widely used for the commercial production of F1 hybrid seeds for many important crops, including maize, sorghum, sunflower etc. (Liu et al., 2011, Bohra et al., 2016). Almost all commercial sunflower hybrids are currently based on a single source of CMS discovered by Leclercq and described above. Such genetic homogeneity of cultivated hybrids makes them extremely vulnerable to new virulent strains of the pathogens and can lead to negative phenomena, for example, epiphytotics development (Levings, 1990). For instance, leaf blight epidemic which affected the maize hybrids based on the Texas type of CMS, which took on a world scale, while other types of CMS were less susceptible to this disease (Bruns, 2017). In this regard, for the creation of new CMS-Rf systems in order to prevent mtDNA unification and to reduce the genetic vulnerability of sunflower hybrids to biotic and abiotic stresses, it is urgent to search and introduce the new CMS sources into sunflower breeding. Although today more than 70 cytoplasmic male sterility types have been identified in sunflower (Garayalde et al. 2015), their insufficient study is an obstacle for utilization into commercial hybrid breeding. Also, there is no doubt that the research of the cytoplasmic male sterility phenomenon is significant for investigating the fundamental problem of nuclear-cytoplasmic interaction in the ontogeny of higher plants (Hanson & Bentolila, 2004). Previously, the comparison of mitochondrial genomes organization between 28 CMS sources of sunflower, performed with

Southern hybridization, demonstrated that some types of CMS (for example, ANN2, PET2, PEF1, etc.) have a different organization of the mtDNA from the PET1-like cytoplasms (Horn, 2002). Sequencing and comparing of whole mitochondrial genomes of various CMS sources will provide additional information about the molecular changes in their mtDNA, which in turn could help to suggest new mechanisms of the male sterility formation.

In the current study, we investigated the structural changes in mitochondrial genomes of HA89-alloplasmic lines: fertile line and two analog lines with different types of cytoplasmic male sterility - PET1, PET2. The results obtained for the PET2 CMS type are the basis for further functional research.

Materials and methods

Plant material

The study was carried out on sunflower fertile line HA89 and isonuclear CMS lines - PET1 and PET2, which were obtained from the genetic collection of the N. I. Vavilov Institute of Plant Genetic Resources (VIR, Russia). So all the lines had the same nuclear genome (HA89), but they differed by chloroplast and mitochondrial genomes, which were inherited from the wild ancestors. The CMS sources were initially obtained by the interspecific hybridization of domesticated sunflower (*Helianthus annuus*) with *H. petiolaris* Nutt (Leclercq, 1969; Whelan, 1980).

Mitochondrial DNA extraction, genome library construction and NGS sequencing

First of all, from leaves of 14-day sunflower seedlings, we extracted organelle fraction with a reduced amount of nuclear DNA as it has been described earlier (Makarenko et al., 2016). For each line, we used the same quantity of leaf tissue from 5 plants. The DNA isolation from such fraction was performed with PhytoSorb kit (Syntol, Russia), according to the manufacturer's instruction. The NGS libraries preparations were made using 1 ng of DNA and Nextera XT DNA Library Prep Kit (Illumina, USA). All the preparation steps were done pursuant to manual. For the qualitative control of libraries, the Bioanalyzer 2100 (Agilent, USA) was used. The libraries quantitation was performed with the Qubit fluorimeter (Invitrogen, USA) and qPCR. For NGS sequencing the libraries were diluted up to the concentration of 8 pM. Libraries were sequenced on different sequencing platforms. Fertile line and PET1 NGS libraries were sequenced with NextSeq 500 sequencer using High Output v2 kit (Illumina, USA). A total

number of 13,240,057 150-bp paired reads were generated for fertile line and 14,758,067 reads - for PET1 line. PET2 library was sequenced with HiSeq2000 and MiSeq platforms using TruSeq SBS Kit v3-HS and MiSeq Reagent Kit v2 500-cycles (Illumina, USA). A total number of 4,471,774 125-bp and 4,931,318 250-bp paired reads were generated for PET2 line.

Analysis of sequence data. Bioinformatics tools

The quality of reads was determined with Fast QC. Trimming of adapter-derived and low quality (Q-score below 25) reads was performed with Trimmomatic software (Bolger et al., 2014). Using Bowtie2 tool v 2.3.3 (Langmead & Salzberg, 2012) sequencing reads were aligned to reference sequence from NCBI databank (NC_023337.1). The Bowtie 2 alignments were done only for concordant paired reads (--no-mixed, --no-discordant options). Variant calling was made with samtools/bcftools software (Li, 2011) and manually revised using IGV tool (Thorvaldsdottir, Robinson & Mesirov, 2013). *De novo* assembly was performed with SPAdes Genome Assembler v 3.10.1 (Nurk et al., 2013) with different K values equal to 75, 85, 95, 127 and read coverage cutoff value equal to 30.0 (--cov-cutoff option). The potential ORFs were identified using ORFfinder. The graphical genome map was generated using OGDRAW tool (Lohse, Drechsel & Bock, 2007). Transmembrane domains were predicted using the TMHMM Server v.2.0 (available online: http://www.cbs.dtu.dk/services/TMHMM-2.0/).

Validation of genome assembly. PCR and Sanger sequencing

The contigs obtained in *de novo* assembly were aligned with reference sunflower mitochondrial genome (NC_023337.1) using BLAST tool. Validation of revealed rearrangements was made by PCR analysis and Sanger sequencing. PCR reactions were performed with LongAmp Taq PCR Kit (New England BioLabs, USA) for reactions with expected amplicons more than 1.5 kb, and with Tersus Plus PCR kit (Evrogen, Russia) for other reactions, including Sanger sequencing. For 28-29 cycles of PCR, we used 0.4 uM of primers (Table 1) and 1 ng of extracted DNA. The direct sequencing of purified amplicons was performed using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA) and ABI Prism 3130xl Genetic Analyser (Applied Biosystems, USA).

RNA extraction and qRT-PCR

Total RNA from the leaves of five samples of each line was extracted with guanidinium thiocyanate-phenol-chloroform reagent kit – ExtractRNA (Evrogen, Russia). RNA quality and concentration were measured using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and the Qubit fluorimeter (Invitrogen, USA). 0.5 μg of total RNA was treated with DNAse I (Thermo Fisher Scientific, USA) according to the manufacturer's instruction. First-strand cDNA was synthesized using MMLV RT kit (Evrogen, Russia) and specific primers. The qPCR was performed with designed primers (Table 1) and PCR kit with EvaGreen dye (Syntol, Russia) on Rotor-Gene 6000 (Corbett Research, Australia). The relative expression level was calculated using 1+E^ΔCt formula, where E=PCR efficiency.

Results and Discussion

Organization of fertile line and PET1 mtDNA

De novo assembly of fertile line and PET1 mitochondrial genomes revealed 9-12 large contigs (10-115 kb), depending on K value. The optimum K value was 95. The obtained large contigs were covering up to 95% reference genome sequence. Wherein, the remaining 5% of the mitochondrial genome, are presented by repeats, so they were the breakpoints of contigs formation. Among numerous repeats in the mitochondrial genome, only one large (12933 bp) and six small (203-728 bp) played a crucial role in genome assembly and prevented single scaffold formation. As well four regions (415-1192 bp) with 99% chloroplast DNA identity affected mitochondrial genome assembly. Eventually, manual assembly based on predominantly contigs obtained by SPAdes supplemented by analysis of reads alignment by Bowtie2 and validation of controversial regions performed by PCR analysis and Sanger sequencing, allowed to summarize sequencing data in completed mitochondrial genomes. Circular mtDNA of HA89 and PET1 lines are presented in figure 1.

The mtDNA comparative analysis of sunflower fertile lines HA412 (NCBI accession NC_023337.1) and HA89 revealed two single nucleotide thymidine insertions: in positions 35690-35691 and 129368-129369 of NC_023337. Two SNP in noncoding part of 301 kb genome is a negligible difference. So we didn't append HA89 mitochondrion sequence in the NCBI databank and, further in this paper used the same positions of mtDNA for HA412 and HA89 lines for simplicity.

The PET1 mitochondrial genome had structural rearrangements as well as polymorphic sites compared to fertile lines' (HA412/HA89) mtDNA. Previously, using the restriction analysis and Sanger sequencing, the large mtDNA aberrations associated with the sterility of plants were detected in PET1 CMS type of sunflower – a 11 kb inversion and a 5 kb insertion (Kohler et al., 1991). The results of the current study not only confirmed the presence of these reorganizations in the HA89 PET1 mitochondrial genome but also allowed to detect more precise genome changes: 11852 bp inversion, 4732 bp insertion, 451 bp deletion. The revealed insertion had 98% identity to the PET1 insertion which is already have presented in NCBI (accession Z23137.1). We also have determined 18 variant sites in HA89(PET1) mtDNA as compared with fertile analog. Among nucleotide variations 8 were localized in SSR loci, 2 deletions (single and dinucleotide) and 7 SNP, including 1 transition and 6 transversions were predominantly located in noncoding regions (Table 2). The exceptions were nonsynonymous mutations in *orf777* (Asp251Glu), *nad6* (Ser232Tyr), *rpl16* (Lys32Gln). The HA89 (PET1) complete mitochondrial genome sequence has been deposited in NCBI databank (accession MG735191).

Organization of PET2 mtDNA

According to the described technique which was used for identification of HA89 fertile and PET1 complete mtDNA sequences we assembled the PET2 mitochondrial genome. The complete mitochondrion of HA89(PET2) made up 316586 bp (Figure 3) and in comparison with the HA89 fertile line contained both the large-scale reorganizations of the mtDNA structure and the minor changes represented by variant sites. Among significant rearrangements, 1 translocation, 2 deletions, 3 insertions and were determined.

Even in a single plant cell, the mitochondrial genome is presented by several DNA molecules with various structure (Sloan et al., 2012). So-called "master chromosome" – the single mtDNA molecule is a rare type of mitochondrial genome organization (Gualberto et al., 2014). More often mitochondrion includes a set of sub-genomic forms (Yang et al., 2015). Because subgenomes could form differing master chromosomes, the statement of translocation is equivocal. To compare complete mitochondrial genomes of HA89(PET2) and fertile HA89 lines the translocation of approximately 77 kb (positions 36393-114174 bp in HA89 fertile line mtDNA) could be established. However, it is important to note that using specific PCR primers (table 1) we identified that in sunflower mtDNA could form the 154 kb sub genome circle molecule (positions 36393-190650). In sunflower genome, there is a repeat region of 722 bp

(36393-37114 = 189929-190650 positions of the fertile line) with 100 % similarity, which makes possible the cyclization of such sub-genome molecule. In case of HA89(PET2) this sub-genome circle molecule has the wrong insertion point in the genome as compare with fertile analog and thus results in such translocation.

The deletion of 711 bp (35682-36393 positions in fertile line mtDNA) resulted in the absence of *orf777* in HA89(PET2) mitochondrial genome. The *orf777* is coding the putative protein with unknown function and have no similarities with other mitochondrial proteins. So its elimination can hardly be the molecular reason for CMS development. The other deletion of 3780 nucleotides (190659-194439 positions) affects only noncoding sequence. Notable that insertions are associated with sub-genome integration regions (36393-37114 and 189929-190650 positions). So there is a possibility, that deletions in these regions could impair master chromosome assembly, which resulted in translocation, which described above.

More significant mtDNA reorganizations are three revealed insertions – 1558 bp, 5050 bp and 14330 bp. Among them, the 5050 bp insertion is most likely not influence CMS phenotype origin. This insertion was found in the intergenic region *atp6-cox2* (275230-275231 positions of fertile line mtDNA), and it does not lead to the formation of new open reading frames (ORFs) directly in the place of insertion into the mitochondrial genome. Moreover, there are no large (more than 300 nucleotides) ORFs within the sequence of 5050 bp insertion. The exception is *orf645* putatively translating polypeptide of 215 amino acids. 23 amino acids on N-terminus of this 215 amino acid protein were similar to N-terminus of ribosomal protein S3. We determined *orf645* transcripts in PET2 CMS line, using specific primers (Table 1) and qPCR. In turn, mRNA was absent in fertile line and PET1 CMS line. Most often the molecular reason for CMS phenotype development is the emergence of chimeric ORFs with transmembrane domains, such as ATPase complex subunits, respiratory-related proteins, etc. (Gillman, Bentolila & Hanson, 2007; Yang, Huai & Zhang, 2009). Consequently, even if *orf645* is translated *in vivo,* its role in male sterility development most likely is negligible, as there are no similarities with respiratory/ATP synthesis-related proteins.

The other 14330 bp insertion in intergenic region *nad4L-atp9* is more complicated than 5050 bp insertion and includes different ORFs. First of all, it should be highlighted that most of the insertion (9482 bp) has 100% similarity to other PET2 mtDNA region (279734-289215 positions). This repeat could be divided into two parts – 6097 bp (35686-41782 positions of

PET2 mitochondrion) are common (99-100% identity) for helianthus mitochondrion: 269147-275243, 273418-279514 and 279734-285830 positions of fertile line, PET1 and PET2 CMS lines, respectively. Such repeat is predominantly consisting of noncoding sequence, except *atp6* gene. The other 3385 bp (41783-45167 positions of PET2 mitochondrion) have 100 % similarity to the part of 5050 bp insertion (285831-289215 positions of PET2 mitochondrion). However, this part of insertion does not contain the coding sequence. The unique part of 14330 bp insertion counts only 4848 bp – 45168-50015 positions of the PET2 mitochondrion. The *orf2565* is encoded in this part of insertion. The *orf2565* translates putative polypeptide of 855 amino acids. Homology search in NCBI database using BLAST pointed on similarity to DNA polymerase (type B). Thus 14330 bp insertions include two coding regions – duplicated *atp6* gene and *orf2565*. The impact of *orf2565* on CMS phenotype development is quite doubtful, taking into account that polypeptide has no transmembrane domains, and the opposite is *atp6* gene influence. *Atp6* chimeric ORFs or new ORFs co-transcribed with *atp6* are the quite common causes of CMS development in different plant species (Kim, Kang & Kim, 2007; Jing et al., 2011; Tan et al., 2017). So we proposed that *atp6* gene and its colocalized area are of particular interest as the candidate sequence for CMS phenotype development. The fact is that in the 5' adjoining sequence (150 bp) to *atp6* start codon there is the other one initiating codon (ATG). Translation from this codon (the same translation frame without any gaps) results in new ORF - *orf1053*. The *orf1053* is a putative chimeric protein consisting of *atp6* with additional 50 amino acids, wherein 37 of 50 amino acids are identical with N-terminus of *coxI*. In the HA89 fertile line most likely some mechanisms interfering the transcription from an alternative promoter. However, when gene colocalization changes, as in case of PET2, the extended transcript could be produced. The similar extension of protein was discovered in other CMS type of sunflower - ANT1 (Spassova et al., 1994), but the described protein had additional 87 aa on C' terminus of the *atp6* protein. To verify the assumption of putative *orf1053* influence on CMS phenotype we analyzed the expression level of *atp6* and *orf1053 (*5' elongated *atp6*) transcripts using the same reverse primer (Table 1). The elongated *atp6* transcript - *orf1053*, have been expressed *in vivo*, however, its expression level was 8-10 fold lower than *atp6* expression level and the ratio of expression (ΔCt) *orf1053/*atp6 was near the same in all three investigated lines of sunflower. Thus the *orf1053* could be a primary transcript of atp6 (pre-mRNA), and is not involved in PET2 type CMS development. Interesting to note that, the relative (to *atp1* gene) expression level of

atp6 in HA89(PET2) had no significant difference as compared with fertile and PET1 CMS analogs, in spite of *atp6* duplication in PET2 mitochondrion. It is also notable that 14330 bp insertion has the same region in mtDNA of PET2 as the other reorganizations - 3780 bp deletion and translocation.

The most important mitochondrial genome rearrangement, associated with PET2 CMS phenotype, is 1558 bp insertion localized between *nad5* and *ccmFc* genes. As well as 14330 bp insertion the 1558 bp insertion has the complement to other genome region part and the unique sequence. First 393 nucleotides of this insertion are similar to 114175-114568 positions of fertile line mtDNA, the next 271 bp are unique, then 500 bp complement to 114587-115087 positions of the HA89 fertile mitochondrion and the rest 394 nucleotides are also unique. From functional point this region presents duplication of *atp9* (114341-114601 positions of HA89 fertile mtDNA) gene impaired with 271 bp insertion and deletion of 12 bp, which resulted in 2 new ORFs formation – *orf228* (Reddemann, Horn 2018) and *orf246. Orf228* encodes 76 amino acids polypeptide 75 of which are identical to C-terminus of ATPase subunit 9, as well as 271 bp insertion, resulted in the start codon (AUG) for *orf228*. Naturally, ATPase subunit 9 has two transmembrane domains – near N- and C-terminus, in o*rf228* there also predicted two transmembrane domains (Reddemann, Horn 2018). However, it should be noted that N-terminus transmembrane domain of *orf228* encoded polypeptide lack of 4 amino acids as compared with *atp9* ones. This difference, in turn, could affect polypeptide (*orf228*) interaction with mitochondrial membrane and so resulting in mitochondrial membrane potential changes. The second new ORF revealed in 1558 bp insertion - *orf246* also encoding the polypeptide with the transmembrane domain. Thus both ORFs (o*rf228, orf246*) could be the main reason for development the PET2 type CMS. It should be pointed out that 1558 bp insertion has the same region in mtDNA of PET2 as the other reorganizations -711 bp deletion and translocation.

The comparative analysis of PET2 mitochondrion sequence with complete mtDNA of fertile line revealed 83 polymorphic sites – 14 SSR, 13 small indels (1-29 bp) and 56 SNP (Table 2). Among nucleotide variations only two SNP were localized in the coding sequence of genes *rps3* and *atp6*. However, these mutations are synonymous. Interesting to note, that PET1 and PET2 share 10 same polymorphic sites as compared with the fertile line. The obtained HA89(PET2) mitochondrial genome sequence has been deposited in NCBI databank (accessions

MG770607). Noteworthy that the sets of primers used for identification of PET2 insertions may be used for designing molecular markers for this type of CMS in sunflower.

Conclusions

The comparative analysis of HA89 fertile and PET1 CMS analog mitochondrial genomes revealed 11852 bp inversion, 4732 bp insertion, 451 bp deletion and 18 variant sites. In the mtDNA of HA89 (PET2) CMS line 77 kb translocation, 711 bp and 3780 bp deletions, as well as 1558 bp, 5050 bp, 14330 bp insertions and 83 polymorphic sites were determined. The 1558 bp insertion resulted in new open reading frames formation - o*rf228* and *orf246*, which could be the main reason for the development of PET2 CMS phenotype.

Acknowledgments

This research was supported by the Ministry of Education and Science of Russia project no. 6.929.2017/4.6 and Russian Foundation for Basic Research project no. 15-34-50294. Analytical work was carried out on the equipment of centres for collective use of Southern Federal University “High Technology”.

References

1. Balk J, Leaver CJ. 2001. The PET1-CMS mitochondrial mutation in sunflower is associated with premature programmed cell death and cytochrome c release. *Plant Cell*13:1803-1818.
2. Bohra A, Jha UC, Adhimoolam P, Bisht D, Singh NP. 2016. Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Reports* 35: 967-993. DOI: 10.1007/s00299-016-1949-3.
3. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for illumine sequence data. *Bioinformatics* 30: 2114-2120. DOI: 10.1093/bioinformatics/btu170.
4. Bruns HA. 2017. Southern Corn Leaf Blight: A Story Worth Retelling. *Agronomy Journal* 109:1218–1224. DOI:10.2134/agronj2017.01.0006.
5. Chen L, Liu YG. 2014. Male sterility and fertility restoration in crops. *Annual Review of Plant Biology* 65:579-606. DOI:10.1146/annurev-arplant-050213-040119.
6. Fujii S, Toriyama K. 2009. Suppressed expression of RETROGRADE-REGULATED MALE STERILITY restores pollen fertility in cytoplasmic male sterile rice plants.

Proceedings of the National Academy of Sciences 106: 9513–9518. DOI: 10.1073/pnas.0901860106.

7. Gagliardi D, Leaver CJ. 1999. Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. *EMBO Journal* 18:3757-66.
8. Garayalde AF, Presotto A, Carrera A, Poverene M, Cantamutto M. 2015. Characterization of a new male sterility source identified in an invasive biotype of *Helianthus annuus* (L.). *Euphytica* 206:579–595. DOI:10.1007/s10681-015-1456.
9. Gillman JD, Bentolila S, Hanson MR. 2007. The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. *The Plant Journal* 49:217–227. DOI: 10.1111/j.1365-313X.2006.02953.x.
10. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. 2013. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* 100: 107-120. DOI: 10.1016/j.biochi.2013.09.016.
11. Hanson MR, Bentolila S. 2004. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *The Plant Cell* 16:154-169. DOI: https://doi.org/10.1105/tpc.015966.
12. Horn R, Hustedt JE, Horstmeyer A, Hahnen J, Zetsche K, Friedt W. 1996. The CMS-associated 16 kDa protein encoded by *orfH522* in the PET1 cytoplasm is also present in other male-sterile cytoplasms of sunflower. *Plant Molecular Biology* 30:523-538.
13. Horn R. 2002. Molecular diversity of male sterility inducing and male-fertile cytoplasms in the genus Helianthus. Theoretical and Applied Genetics 104:562-570.
14. Horn R, Kusterer B, Lazarescu E, Prüfe M, Friedt W. 2003. Molecular mapping of the Rf1 gene restoring pollen fertility in PET1-based F1 hybrids in sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics* 106:599-606.
15. Horn R, Gupta KJ, Colombo N. 2014. Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* 19:198-205. DOI: 10.1016/j.mito.2014.04.004.
16. Ivanov MK, Dymshits GM. 2007. Cytoplasmic male sterility and restoration of pollen fertility in higher plants. *Russian Journal of Genetics* 43: 354–368. DOI: https://doi.org/10.1134/S1022795407040023.

17. Jing B, Heng S, Tong D, Wan Z, Fu T, Tu J, Ma C, Yi B, Wen J, Shen J. 2012. A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. *Journal of Experimental Botany* 63: 1285–1295. DOI:10.1093/jxb/err355.
18. Kim DH, Kang JG, Kim BD. 2007. Isolation and characterization of the cytoplasmic male sterility-associated orf456 gene of chili pepper (Capsicum annuum L.). *Plant Molecular Biology* 63:519–532. DOI 10.1007/s11103-006-9106-y.
19. Köhler RH, Horn R, Lössl A, Zetsche K. 1991. Cytoplasmic male sterility in sunflower is correlated with the co-transcription of a new open reading frame with the *atpA* gene. *Molecular and General Genetics* 227:369-376.
20. Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359. DOI: 10.1038/nmeth.1923.
21. Leclercq P. 1969. Une sterilite male chez le tournesol. *Ann Amelior Plant* 19:99–106.
22. Levings CS 3rd. 1990. The Texas cytoplasm of maize: cytoplasmic male sterility and disease susceptibility. Science 250:942-947.
23. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987-2993. DOI: 10.1093/bioinformatics/btr509.
24. Liu H, Cui P, Zhan K, Lin Q, Zhuo G, Guo X, Ding F, Yang W, Liu D, Hu S, Yu J, Zhang A. 2011. Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. *BMC Genomics* 12:163. DOI: 10.1186/1471-2164-12-163.
25. Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Current Genetics 52:267-274. DOI: 10.1007/s00294-007-0161-y.
26. Makarenko MS, Usatov AV, Markin NV, Azarin KV, Gorbachenko OF, Usatov NA. 2016. Comparative Genomics of Domesticated and Wild Sunflower: Complete Chloroplast and Mitochondrial Genomes *OnLine Journal of Biological Sciences* 16:71-75. DOI: 0.3844/ojbsci.2016.71.75.
27. Monéger F, Smart CJ, Leaver CJ. 1994. Nuclear restoration of cytoplasmic male sterility in sunflower is associated with the tissue-specific regulation of a novel mitochondrial gene. *EMBO Journal* 13:8-17.

28. Nizampatnam NR, Doodhi H, Kalinati Narasimhan Y, Mulpuri S, Viswanathaswamy DK. 2009. Expression of sunflower cytoplasmic male sterility-associated open reading frame, orfH522 induces male sterility in transgenic tobacco plants. *Planta* 229:987-1001. DOI: 10.1007/s00425-009-0888-4.

29. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, McLean J, LaskenR, Clingenpeel SR, Woyke T, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. *Research in Computational Molecular Biology. Lecture Notes in Computer Science* 7821: 158-170. DOI https://doi.org/10.1007/978-3-642-37195-0_13.

30. Reddemann A., Horn R. 2018. Recombination Events Involving the *atp9* Gene Are Associated with Male Sterility of CMS PET2 in Sunflower. *International Journal of Molecular Sciences* 19: 806. DOI:10.3390/ijms19030806.

31. Sabar M, Gagliardi D, Balk J, Leaver CJ. 2003. ORFB is a subunit of F1F(O)-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. *EMBO Reports* 4:381-386.

32. Siculella L, Palmer JD. 1988. Physical and gene organization of mitochondrial DNA in fertile and male sterile sunflower. CMS-associated alterations in structure and transcription of the *atpA* gene. *Nucleic Acids Research* 16:3787-3799.

33. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. 2012 Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates. *PLoS Biology* 10(1): e1001241. DOI: https://doi.org/10.1371/journal.pbio.1001241.

34. Spassova M, Moneger F, Leaver CJ, Petrov P, Atanassov A, Nijkamp HJ, Hille J. 1994. Characterisation and expression of the mitochondrial genome of a new type of cytoplasmic male-sterile sunflower. *Plant Molecular Biology* 26: 1819-1831. DOI https://doi.org/10.1007/BF00019495.

35. Tan GF, Wang F, Zhang XY, Xiong AI. 2017. 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, DNA Mapping, Sequencing, and Analysis 1-9.* DOI: 10.1080/24701394.2017.1303492.

425 36. Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV):
426 high-performance genomics data visualization and exploration. *Briefings in*
427 *Bioinformatics* 14:178-192. DOI: 10.1093/bib/bbs017.
428 37. Touzet P, Meyer EH. 2014. Cytoplasmic male sterility and mitochondrial metabolism in
429 plants. *Mitochondrion* 19:166-171. DOI: 10.1016/j.mito.2014.04.009.
430 38. Whelan EDP. 1980. A new source of cytoplasmic male sterility in sunflower. *Euphytica*
431 29:33-46.
432 39. Yang JH, Huai Y, Zhang MF. 2009. Mitochondrial atpA gene is altered in a new orf220-
433 type cytoplasmic male-sterile line of stem mustard (Brassica juncea). *Molecular Biology*
434 *Reports* 36:273–280. DOI 10.1007/s11033-007-9176-1.
435 40. Yang J, Liu1 G, Zhao1 N, Chen S, Liu D, Ma1 W, Hu Z, Zhan M. 2015. Comparative
436 mitochondrial genome analysis reveals the evolutionary rearrangement mechanism in
437 *Brassica. Plant Biology* 18:527-536. DOI: 10.1111/plb.12414.

Table 1(on next page)

Table 1 The primers sets used for HA89 (PET2) genome reorganizations validation and gene expression analysis.

All PCR reactions were held for HA89 fertile and CMS (PET2) lines. For simplicity primers were named according to their position in the HA89 fertile genome. The ‘F’ and ‘R’ letters denote the PCR strand orientation - forward (plus) and reverse (minus), respectively.

The purpose	Primer name	Primer sequence (5'-3')	Line	The expected amplicon size (kbp)	The real amplicon size (kbp)
Validation sub-genome structure 153,5 kb circle	189837F	CGTGAAGCCGGGATGGTATT	Fertile	-	0.9
	37192R	CAAGTGATCCCCCATCCAGG	PET1	-	0.9
			PET2	0.9	0.9
	189660F	AGGAGTGAGATGGACGCTCT	Fertile	-	1.8
	37954R	AAGTGTTGCACCCCCTTGAA	PET1	-	1.8
			PET2	1.8	1.8
The analysis of *orf645* expression	orf645F	GCCTTCCACCTCTCGTTTGA	Fertile	-	-
	orf645R	TCCGAAAGCCGGCCTAAAAT	PET2	162	162
The analysis of atp6/*orf1053* expression	atp6F	AGAACTGTAACTGACAACGC	Fertile	106	106
	atp6R	A	PET2	106	106
		CCTGAGTCCGAGTCTGCATC			
	orf1053F	TCCCATGCCTTTCTTGGTCG	Fertile	-	280
	atp6R	-||-	PET2	280	280
	orf1053F	-||-	Fertile	-	147
	orf1053R	CTCTCATAACCGGTGTGTGGT	PET2	147	147
The analysis of *atp1* expression	atp1F	CCCATGGCACAGCCAGAATA	Fertile	140	140
	atp1R	CAGAAACGCTCAACTGTGGC	PET2	140	140
Sanger resequencing the 5' and 3' ends of 5050 bp insertion	274655F	-||-	Fertile	-	-
	Pet2-seqR	GAAGGAACGAGACAGCACCA	PET2	0.7	0.7
	Pet2-seqF	AGGGAGAGGGACGAAGTGAC	Fertile	-	-
	275503R	TAACCGCTGCAAGAGTGAGG	PET2	0.7	0.7
Detection the 5' and 3' end of 14330 bp insertion	35202F	AGCTCTCCCCATCGGTAGTT	Fertile	-	-
	271194R	GGTCATCAGTTCGAGTGGCA	PET2	-	2.5
	Pet14kbF	AGGAAAAGACCCAACAGGCA	Fertile	-	-
	115189R	GGACACGCAGAAGCCAATTC	PET2	1.9	1.9
Detection of 1558 bp insertion	114439F	GAGCAAAGCCCAAAATGGCA	Fertile	-	-
	194675R	TAGCTCTTCCGGAGCACTCT			

1

Table 2 (on next page)

Table 2 Variation sites in mitochondrial DNA of HA89 CMS lines PET1 and PET2.

Nucleotide positions are specified according to fertile line mtDNA. IGR – an intergenic region. In case of indels, the deletions are indicated as "-" and the inserted nucleotide are in bold.

Position	Type	Fertile	PET1	PET2	Localization
3031	SSR	G5		**G6**	IGR *nad2-ccmC*
3107	SSR	T5		**T6**	IGR *nad2-ccmC*
3275-3276	INDEL	TA		T**TT**A	IGR *nad2-ccmC*
3281-3281	INDEL	AT		A**T**T	IGR *nad2-ccmC*
4715	SSR	T8		**T9**	IGR *nad2-ccmC*
6207	SSR	A8	**A7**	**A7**	IGR *nad2-ccmC*
6660	SNP	A		**G**	IGR *nad2-ccmC*
7404	SSR	G10		**G9**	IGR *nad2-ccmC*
7919	INDEL	A		**-**	IGR *nad2-ccmC*
9796	SNP	T		**C**	IGR *nad2-ccmC*
10467	SNP	A		**C**	IGR *nad2-ccmC*
10924	SNP	A		**C**	IGR *nad2-ccmC*
12314	SNP	T		**C**	IGR *nad2-ccmC*
19594	SNP	G		**A**	IGR *ccmC-atp4*
23917	SNP	G		**T**	IGR *ccmC-atp4*
31803	SNP	A		**C**	IGR *nad4L-orf259*
34099	SNP	A		**C**	IGR *nad4L-orf259*
34135-34136	INDEL	AT		A**TG**T	IGR *nad4L-orf259*
34162	SNP	T		**C**	IGR *nad4L-orf259*
35031	SNP	C		**A**	IGR *nad4L-orf259*
35114	SNP	C		**A**	IGR *nad4L-orf259*
35478	SNP	T		**C**	IGR *nad4L-orf259*
35511	SNP	G		**A**	IGR *nad4L-orf259*
35596	SNP	G		**C**	IGR *nad4L-orf259*
36360	SNP	T	**G**	**-**	orf259 Asp251Glu
42295	SNP	C		**A**	IGR *coxIII-rpl5*
46039	INDEL	A	**-**		IGR *rpl5-nad4*
49272	SSR	C11	**C9**	**C10**	IGR *nad4-ccmB*
50856	SNP	C		**A**	IGR *nad4-ccmB*
51678	SSR	G10	**G9**	**G9**	IGR *nad4-ccmB*
62360	SNP	T		**G**	IGR *nad4-ccmB*
62403	SNP	G		**A**	IGR *nad4-ccmB*
63433-63434	INDEL	TC		T**CC**	IGR *nad4-ccmB*
71497-71498	INDEL	GT		G**GGGC**T	IGR *rpl10-nad1*

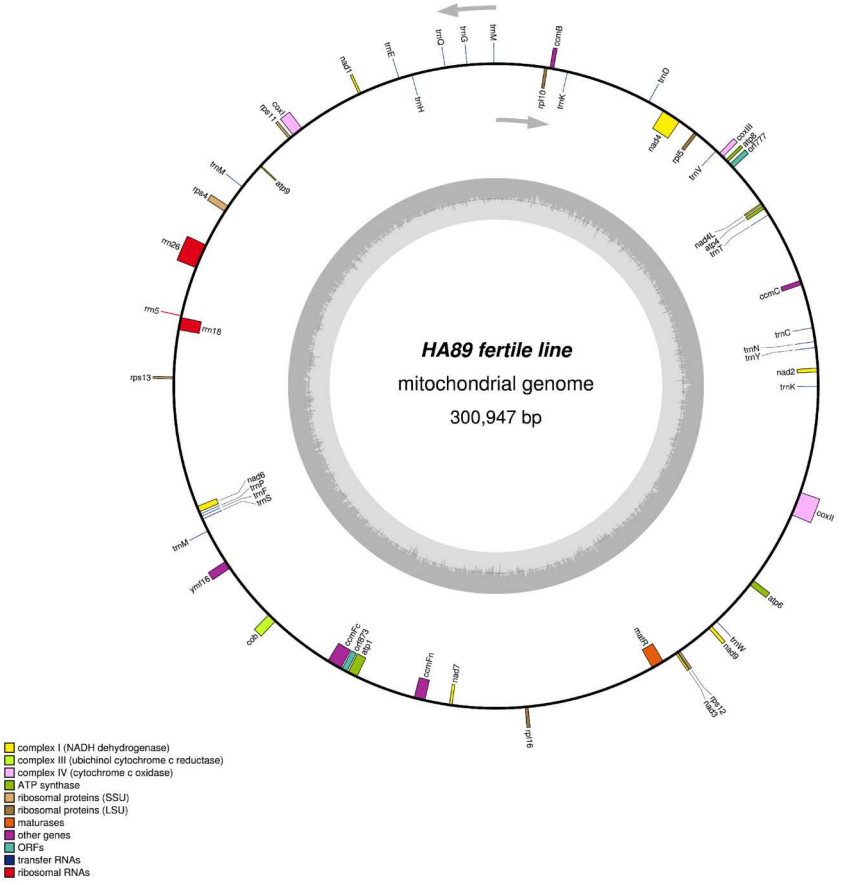
75332	SNP	A	**C**	**C**	IGR *rpl10-nad1*
91105	SNP	G		**T**	IGR *rpl10-nad1*
91106	SNP	A		**C**	IGR *rpl10-nad1*
105474	SSR	T35		**T25**	IGR *nad1-coxI*
108200	SNP	T		**G**	IGR *coxI-rps11*
115915	SNP	T		**G**	IGR *atp9-rps4*
116777	SNP	G	**T**	**G**	IGR *atp9-rps4*
119331	SNP	G		**A**	IGR *atp9-rps4*
121108-121109	INDEL	CC		**CTTC**	IGR *atp9-rps4*
122990	SNP	A		**C**	rps4 (synonymous)
133546	SNP	T		**A**	IGR *rrn26-rrn5*
133547-133548	INDEL	AT		**AGG**	IGR *rrn26-rrn5*
133548	SNP	T		**G**	IGR *rrn26-rrn5*
133549	SNP	A		**C**	IGR *rrn26-rrn5*
156213	SNP	C		**A**	IGR *rps13-nad6*
156621-156622	INDEL	CC		**CCTAC**	IGR *rps13-nad6*
157459	SNP	T		**G**	IGR *rps13-nad6*
169028	SNP	G	**T**		*nad6 (Ser232Tyr)*
170185	SSR	T14	**T12**	**T12**	IGR *nad6-ymf16*
174932-174933	INDEL	AC		**ACTCGACTGAA AGGAAAGGTA CGAAGTGGC**	IGR *nad6-ymf16*
175179	SNP	G		**T**	IGR *nad6-ymf16*
178406	SSR	T9	**T8**		*ymf16* intron
184739	SSR	A10	**A11**		IGR *ymf16-cob*
188363	SSR	T11	**T10**	**T10**	*cob* intron
189980	SNP	G		**T**	IGR *cob-ccmFc*
195008	SNP	G		**T**	IGR *cob-ccmFc*
195015	SNP	C		**A**	IGR *cob-ccmFc*
200174	SNP	G		**A**	*ccmfC* intron
200515	SNP	G		**A**	*ccmfC* intron
202672	SNP	T	**C**		IGR *orf873-atp1*
204990	SNP	C		**A**	IGR *atp1-ccmFn*
204846-204847	INDEL	AA		**ATA**	IGR *atp1-ccmFn*
207965	SSR	G10		**G12**	IGR *atp1-ccmFn*

209335-36	INDEL	AA	-		IGR *atp1-ccmFn*
209458	SNP	G		**A**	IGR *atp1-ccmFn*
212638	SSR	C9		**C12**	IGR *atp1-ccmFn*
215916	SNP	C		**T**	IGR *ccmFn-nad7*
223917	SNP	A		**C**	IGR *nad7-rps3*
223925-223926	INDEL	GA		GAA	IGR *nad7-rps3*
226977-226978	INDEL	AC		**ACGTTGTTTT**C	IGR *nad7-rps3*
230112	SNP	A	**C**		*rpl16* (Lys32Gln**)**
232826	SNP	G		**T**	IGR *rpl16-matR*
239880	SNP	G		**A**	IGR *rpl16-matR*
239988	SNP	A		**C**	IGR *rpl16-matR*
241035	SNP	G		**A**	IGR *rpl16-matR*
241475	SNP	A		**C**	IGR *rpl16-matR*
246053	SNP	C		**T**	IGR *rpl16-matR*
248266	SSR	A14	**A10**	**A9**	IGR *rpl16-matR*
249347	SSR	T8		**T9**	IGR *rpl16-matR*
249361	SNP	C		**A**	IGR *rpl16-matR*
260901	SNP	G		**T**	IGR *nad9-atp6*
262080	SNP	G		**A**	IGR *nad9-atp6*
269062	SNP	G	**C**	**C**	IGR *nad9-atp6*
269134	SNP	A		**C**	*atp6* (synonymous)
270676	SNP	G		**T**	IGR *atp6-coxII*
273344	SNP	C		**A**	IGR *atp6-coxII*
276834	SNP	T		**G**	IGR *atp6-coxII*

1

Figure 1(on next page)

Figure 1 Graphical mitochondrial genome maps of HA89 fertile and PET1 lines.



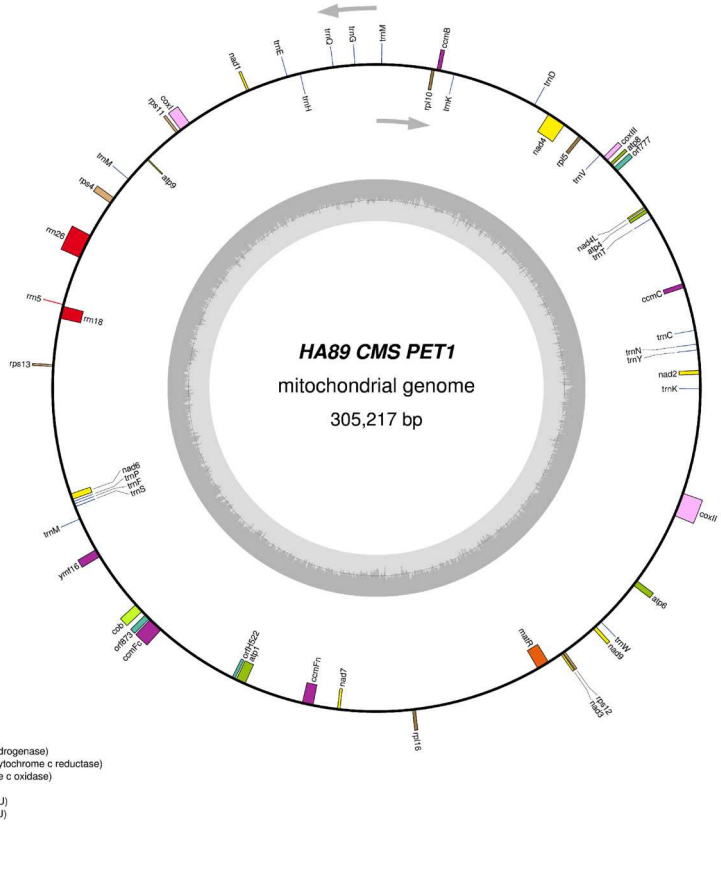
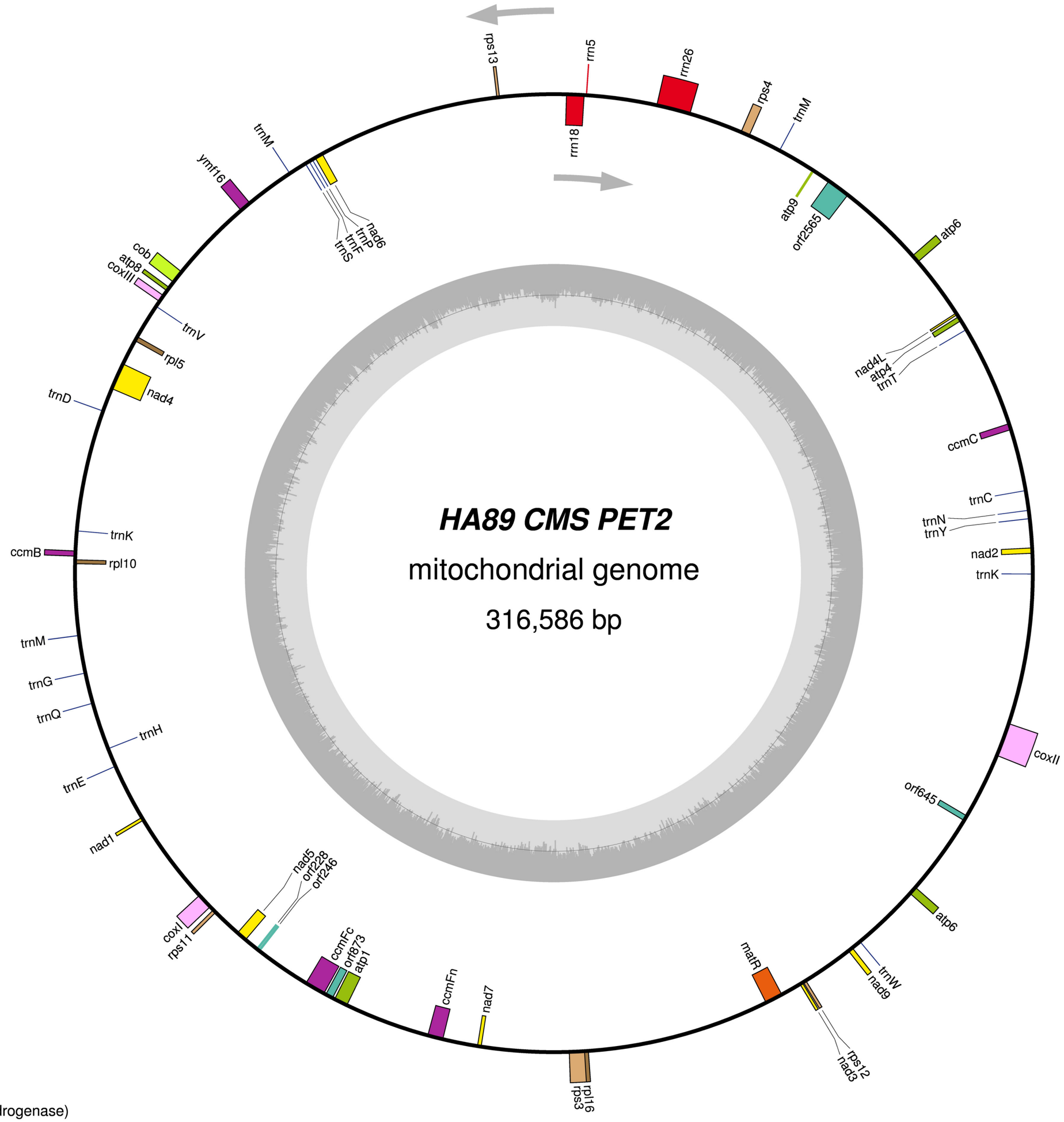


Figure 2(on next page)

Figure 2 Graphical mitochondrial genome map of PET2 line



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26438v2 | CC BY 4.0 Open Access | rec: 2 Apr 2018, publ: 2 Apr 2018