


Please cite the Published Version

Garcia, LE, Zubko, MK , Zubko, EI and Sanchez-Puerta, MV (2019) Elucidating genomic patterns and recombination events in plant cybrid mitochondria. *Plant Molecular Biology*, 100 (4-5). pp. 433-450. ISSN 0167-4412

DOI: <https://doi.org/10.1007/s11103-019-00869-z>

Publisher: Springer

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/623127/>

Usage rights:  In Copyright

Additional Information: This is an Author Accepted Manuscript of an article in *Plant Molecular Biology* published by Springer.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Elucidating genomic patterns and recombination events in plant cybrid mitochondria

Laura E. Garcia^{1,2}  · Mikhajlo K. Zubko³ · Elena I. Zubko³ · M. Virginia Sanchez-Puerta^{1,2}

Received: 2 January 2019 / Accepted: 1 April 2019
© Springer Nature B.V. 2019

Abstract

Key message Cybrid plant mitochondria undergo homologous recombination, mainly BIR, keep a single allele for each gene, and maintain exclusive sequences of each parent and a single copy of the homologous regions.

Abstract The maintenance of a dynamic equilibrium between the mitochondrial and nuclear genomes requires continuous communication and a high level of compatibility between them, so that alterations in one genetic compartment need adjustments in the other. The co-evolution of nuclear and mitochondrial genomes has been poorly studied, even though the consequences and effects of this interaction are highly relevant for human health, as well as for crop improvement programs and for genetic engineering. The mitochondria of plants represent an excellent system to understand the mechanisms of genomic rearrangements, chimeric gene formation, incompatibility between nucleus and cytoplasm, and horizontal gene transfer. We carried out detailed analyses of the mtDNA of a repeated cybrid between the solanaceae *Nicotiana tabacum* and *Hyoscyamus niger*. The mtDNA of the cybrid was intermediate between the size of the parental mtDNAs and the sum of them. Noticeably, most of the homologous sequences inherited from both parents were lost. In contrast, the majority of the sequences exclusive of a single parent were maintained. The mitochondrial gene content included a majority of *N. tabacum* derived genes, but also chimeric, two-parent derived, and *H. niger*-derived genes in a tobacco nuclear background. Any of these alterations in the gene content could be the cause of CMS in the cybrid. The parental mtDNAs interacted through 28 homologous recombination events and a single case of illegitimate recombination. Three main homologous recombination mechanisms were recognized in the cybrid mitochondria. Break induced replication (BIR) pathway was the most frequent. We propose that BIR could be one of the mechanisms responsible for the loss of the majority of the repeated regions derived from *H. niger*.

Keywords Mitochondrial DNA · Cybrid · DNA recombination · BIR · Protoplast fusion

Introduction

Mitochondria are semi-autonomous cellular compartments, in which numerous essential metabolic pathways take place (Logan 2006; Taanman 1999). This organelle contains its own genome and molecular machinery for gene expression. However, the maintenance of the mitochondrial genome (mtDNA) and the transcription of mitochondrial-encoded genes are controlled by the nucleus. Therefore, mtDNA replication, structural organization and gene expression depend on the coordinated interactions between the mitochondrial and the nuclear genomes (Chase 2007). The effects of the co-evolution of these genomes are relevant for the human health (Chou and Leu 2015), for breeding programs (Orczyk et al. 2003), and genetic engineering (Greiner and Bock 2013). Angiosperm mtDNAs have particular characteristics

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11103-019-00869-z>) contains supplementary material, which is available to authorized users.

✉ Laura E. Garcia
lauraevgarcia@gmail.com

¹ Facultad de Ciencias Agrarias, IBAM, Universidad Nacional de Cuyo, CONICET, Almirante Brown 500, M5528AHB Chacras de Coria, Argentina

² Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, 5500 Mendoza, Argentina

³ Centre for Bioscience, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester M1 5GD, UK

that distinguish them from those in fungi and animals: very low substitution rates, highly scrambled gene order, relatively large and variable genome sizes (between 222 kb and 11.3 Mb), and long intergenic regions that encompass up to 98% of the genome (Kubo and Mikami 2007; Skippington et al. 2015; Sloan et al. 2012). Part of these intergenic regions could be attributed to repeats and to sequences acquired from the nuclear and chloroplast genomes by intracellular gene transfer or from other plant mitochondria via horizontal gene transfer (Rice et al. 2013; Sanchez-Puerta et al. 2008; Xi et al. 2013), while the major part remains to be characterized (Gualberto and Newton 2017). Frequent homologous recombination events across repeats make plant mtDNA prone to rearrangements, evolving rapidly in structure (Arrieta-Montiel and Mackenzie 2011; Palmer and Herbon 1988). The mixing of mtDNA regions by homologous recombination can modify gene organization and create chimeric genes that play a role in evolution (Gualberto and Newton 2017). In particular, the cytoplasmic male sterility (CMS) is an economically important trait that can result from the formation of chimeric mitochondrial genes (Chase 2007; Touzet and Meyer 2014). The generation of somatic hybrids via protoplast fusions has been used for years to introduce CMS to crop species, as well as to accelerate and improve various aspects of plant breeding (Gleba and Sytnik 1984; Orczyk et al. 2003).

Somatic hybridization between different plant species represents a powerful method to generate novel traits or to transfer organelle-encoded features to cultivars. Cybrids (cytoplasmic hybrids) are somatic hybrids that contain the nuclear genome of one parent while the cytoplasmic genomes could be inherited from the other parent or from both parents. The production of cybrids, in particular, is an important tool for crop improvement and to study nuclear-cytoplasmic interactions and evolutionary dynamics (Greiner and Bock 2013; Woodson and Chory 2008). Somatic hybrid production has been widely used in Solanaceae breeding programs (Austin et al. 1988; Brown et al. 1996; Iovene et al. 2007; McGrath et al. 2002; Orczyk et al. 2003). Interspecific and intergeneric somatic hybrids (including cybrids) have been produced between *Nicotiana* spp. and related species to introduce relevant traits, such as CMS (Carlson et al. 1972; Fitter et al. 2005; Ilcheva et al. 2000; Sun et al. 2005).

It is well known that plant somatic hybridization often results in recombinant mitochondrial genomes (Belliard et al. 1979; Gleba and Sytnik 1984; Vedel et al. 1986; Zubko et al. 1996). Characterization of these chimeric mitogenomes has been quite limited, including the identification of a small number of parental-specific markers by Southern blot hybridizations (Aviv et al. 1984; Belliard et al. 1979; Kofler et al. 1991; Morgan and Maliga 1987; Nagy et al. 1983; Rothenberg and Hanson 1988; Scotti et al. 2004) and the

complete sequence of two cybrid mtDNAs (Arimura et al. 2018; Sanchez-Puerta et al. 2015). Genomic comparisons between parental and cybrid cytoplasmic genomes showed that the cybrid mitochondrial genomes were larger than each parental mtDNA and remarkably chimeric. Despite their increased size, the cybrid mtDNAs contained single alleles for most of the mitochondrial genes. Retaining a single form of each gene could be a selective mechanism to minimize intracellular incompatibilities or a neutral process that preferentially eliminates duplicated regions (Sanchez-Puerta et al. 2015). Further analysis of hybrid mitochondria would help to elucidate the nature and evolutionary dynamics of the molecular processes in plant mitochondria, such as genomic recombination, nuclear-cytoplasmic interactions, and the incorporation of foreign sequences through horizontal gene transfer.

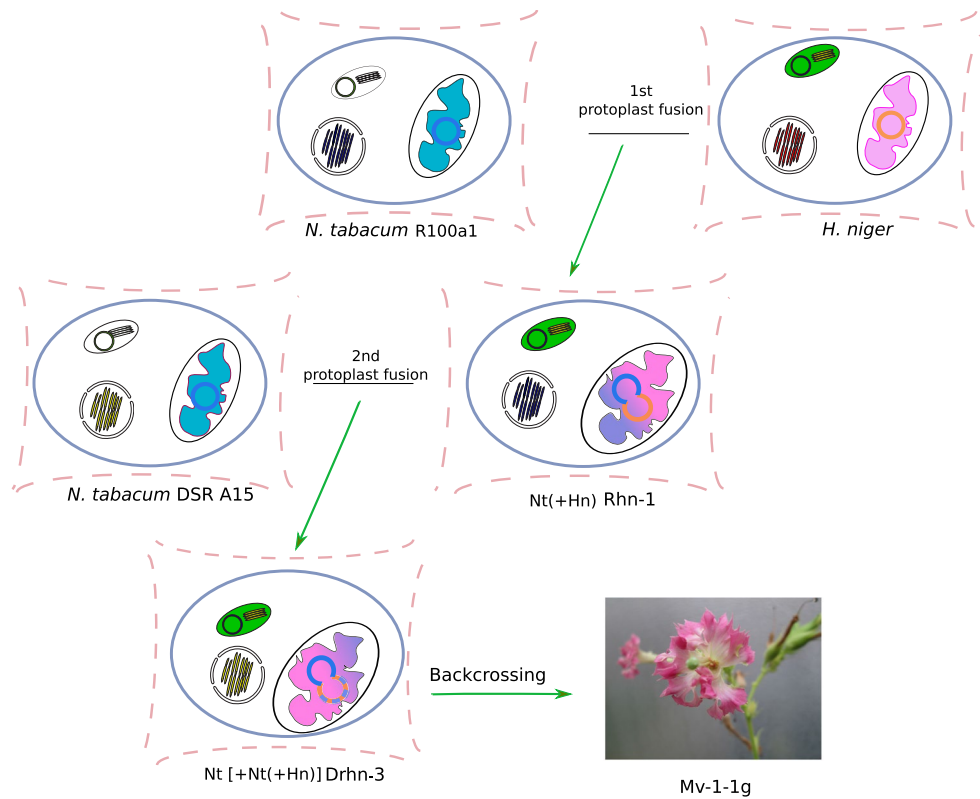
In this study, we aim to gain insight into the mitochondrial genome dynamics of cybrid plants through sequencing and analysis of the mtDNA of a repeated cybrid between *Nicotiana tabacum* and *Hyoscyamus niger*. This cybrid combination was chosen because both parental mtDNAs were sequenced (Gurdon et al. 2016; Sanchez-Puerta et al. 2015; Sugiyama et al. 2005) and a previous study on a cybrid produced from the same parent combination was reported (Sanchez-Puerta et al. 2015), allowing a direct comparison of the evolutionary path taken by this independent cybrid. We wish to address the following questions: Were the homologous sequences from each parent mtDNA lost from the cybrid mitochondria as observed in the *N. tabacum* (+*H. niger*) cybrid previously analyzed? Did the cybrid mtDNA maintain the sequences that are exclusive of each parental mitochondria? Has the initial co-existence of the two parental alleles in the cybrid mtDNA been maintained? Is there a strong preference for tobacco-derived intergenic and/or genic sequences given that the nuclear genome was inherited from tobacco?

Materials and methods

Genome sequencing, assembly, and validation

Total DNA was extracted from leaves of an individual plant of a repeated cybrid between *Nicotiana tabacum* and *Hyoscyamus niger* (Fig. 1) following the Dellaporta method (Dellaporta et al. 1983). This plant was denoted as Mv-1-1g. DNA from the plant was used to construct a paired end library with insert size of ~800 bp that was sequenced at the Beijing Genomics Institute using the Illumina HiSeq 2500 sequencing system, generating 70 million clean paired-end reads of 125 bp. For the reconstruction of the mtDNA, we carried out the following methodology. First, we performed de novo assembly using Velvet v.1.2.03 (Daniel Zerbino,

Fig. 1 Schematic production of the repeated cybrid plant Mv-1-1g in a two-step protoplast fusion experiment. In the first step, protoplasts from an albino mutant line of *Nicotiana tabacum* and a wild-type line of *Hyoscyamus niger* were fused. Among green hybrid plants there was a cybrid line Rhn-1 manifesting a pistilloid form of CMS. Protoplasts of Rhn-1 were inactivated by gamma irradiation before the second protoplast fusion with tobacco plastome mutant DSR A15. Mv-1-1g is a descendant of the line Drhn-3 backcrossed with wild-type *N. tabacum*



European Bioinformatics Institute Cambridge, UK) on the Mason large-memory computer cluster at Indiana University-Bloomington (IN, USA). We ran eight independent Velvet de novo assemblies without scaffolding and using different hash lengths each time (k-mer lengths: 41, 73, 81, 97, 99, 103, 107, 111). The best assembly (k-mer 103) resulted in 387 contigs larger than 1 kb with N50 contig length of 5249 bp and 35 contigs in the N50. Using BLAST we selected 66 contigs that matched the mitochondrial genomes of the parents (*H. niger* and *N. tabacum*) and had query coverage (qcov) greater than 50%. These putative-mitochondrial contigs were used as trusty contigs in the assembler SPAdes (Bankevich et al. 2012), producing 51 contigs that were later extended using SSAKE (Warren et al. 2007). The extended contigs were joined using SSPACE (Boetzer et al. 2011) and Gapfiller (Boetzer and Pirovano 2012). Finally the resulting contigs were manually assembled based on consistent paired-end reads using Consed v.26 (Gordon and Green 2013). The quality of mitochondrial genome assembly was evaluated by mapping all Illumina reads until the read depth was as expected and the high-quality mismatches were very low or zero. The DNA library had a mean size estimated by Consed v.26 of 769 bp, SD 139. Repeats shorter than 750 bp and with >80% identity were identified with cross-match in Consed v.26. To identify repeats larger than the insert size, we took into account the total read depth and

paired-end read information using Consed v.26. To confirm the absence of complete mtDNA sequences from the parents (even at low stoichiometries) in the cybrid mitochondria, we mapped all Illumina reads to the *H. niger* or *N. tabacum* mitochondrial genome sequences using Consed v.26. The total and paired-end read depth across the parental mtDNAs was irregular, showing regions that lack reads and regions with inconsistent paired-end reads. Thus, we confirmed that the cybrid mitochondria did not contain the full-length genomes from its parents but it was composed of a subset of sequences from both parents. Chloroplast contigs were identified through BLAST and the chloroplast genome was assembled using Novoplasty (Dierckxsens et al. 2017).

All raw sequence data are available from the NCBI Sequence Read Archive database as accession number PRJNA512112. The annotated cybrid mitochondrial genome was deposited in the GenBank data libraries under accession numbers MH168702-MH168706.

Genome annotation

Mitochondrial contigs were annotated using the software Geneious v.11.0.4, BLAST (Altschul et al. 1990), and tRNAscan (Lowe and Chan 2016). Graphical genome maps were generated using OGDRAW (Lohse et al. 2007).

Unknown ORFs (> 300 bp and with ATG as start codon) were identified using Geneious v.11.0.4.

Comparisons between genomes

BLASTn comparisons of the mitochondrial genomes of the cybrid and its parents (*N. tabacum*: KR780036, NC_006581 and *H. niger*: NC_026515) were visualized in Genome Workbench and Geneious v.11.0.4. Putative intergenomic recombination events in the cybrid mtDNA were inferred by visual inspection of BLAST results. Gene conversion events (without crossover) were inferred when three or more polymorphisms in a short tract (< 400 bp) were shared by the cybrid and one parent inside a region that was otherwise clearly derived from the other parent. Genes were aligned, and differences between *N. tabacum* and *H. niger* in protein-coding genes were scored as synonymous or nonsynonymous substitutions using Geneious v.11.0.4.

Results

Cybrid production

The Mv-1-1g repeated cybrid was constructed by a two-step protoplast fusion involving two species of the family Solanaceae (Fig. 1). In the first experiment, protoplasts from a recipient line of *Nicotiana tabacum* L. ($2n = 4x = 48$) plastome albino mutant (R100a1) were fused with those from a donor line of wild type *Hyoscyamus niger* L., $2n = 2x = 34$ (Zubko et al. 1996). Since albinism of the R100a1 line was controlled by plastome, any green tobacco-like clones recovered after fusions with green protoplasts of the donor should contain green chloroplasts of *H. niger* transferred to albino tobacco cells. This approach is quite efficient for selection of cytoplasmic hybrids, especially in remote combinations (Kushnir et al. 1987; Zubko et al. 1996). The fusion event resulted in the cybrid line Rhn-1 that was phenotypically similar to tobacco with the prevalence of the tobacco nuclear genome after 4 years of vegetative propagation. The cybrid exhibited cytoplasmic male sterility (CMS): even mature flowers did not have corolla and stamens, and instead, they possessed 3–4 flower-like structures with pronounced pistilloidy (Zubko et al. 1996). In order to follow the genetic fate of this CMS trait and to associate it further with the composition of the mtDNA, the cytoplasm of Rhn-1 was re-transferred into the nuclear background of a plastome chlorophyll deficient-tobacco mutant (DSR A15) in a second experiment on protoplast fusion (involving inactivation of Rhn-1 protoplasts with gamma-irradiation) and posterior in vitro culture (Fig. 1). Methodological details of the cytoplasm re-transfer and

initial characterization of plants resulted from this experiment are described in the previous work (Zubko et al. 1996). Using gamma-irradiation in this experiment was the only efficient way to prevent cell divisions of Rhn-1 after protoplast fusions (otherwise, new cybrid lines would be phenotypically indistinguishable from the cybrid lines of Rhn-1). In addition, gamma-irradiation reduces transmission of nuclear genetic material from the donor of cytoplasm. Most cybrid lines recovered from these fusions inherited the pistilloid CMS phenotype of flowers, indicating that the trait is controlled by cytoplasm. Backcrossing these flowers was rarely successful. One of the resulting cybrid lines named Drhn-3 was reproducibly fertile in backcrosses. The plants of this line rooted in vitro were transferred to soil and placed in a greenhouse for repeated backcrosses with wild type tobacco in order to minimize the deviation of the cybrid nuclear genome from tobacco genome. Drhn-3 plants exhibited CMS, wrinkled corollas, no stamens, and one pistil. Here, we studied a cybrid plant (named Mv-1-1g) that was obtained after two backcrosses of the cybrid line Drhn-3 with wild type tobacco. This plant manifested more elongated leaves (in comparison to wild-type tobacco), funnel-like corolla (in contrast to tobacco bell-like corolla), crimped petals, and CMS (due to undeveloped stamens without pollen). The chloroplast genome of the cybrid Mv-1-1g was inherited from *H. niger* and 100% identical to the *H. niger* cpDNA.

The cybrid Mv-1-1g mitochondrial genome

We sequenced and assembled the mitochondrial genome of the Mv-1-1g repeated cybrid. The repeated cybrid was obtained after the second round of cybridization between an existing cybrid (a donor of the cytoplasm) and tobacco (a recipient of the cytoplasm), as explained above. The mtDNA assembled into 5 contigs (ranging between 9520 and 359,619 bp) with a total length of 707,034 bp (Fig. 2). The genome length was intermediate between the sizes of the parental mtDNAs (~501 and ~431 kb) and the sum of them (~932 kb). The total read depth was fairly even across the genome (Supplementary Fig. 1). All mitochondrial contigs could potentially recombine with each other through large and intermediate repeats. Based on the frequency of recombination, mitochondrial repeats have been classified as large repeats (> 1 kb) that mediate high-frequency recombination and shorter repeats (100–1000 bp) that recombine sporadically (Maréchal and Brisson 2010). We have found 18 large (> 1 kb) and 97 intermediate (100–1000 bp) repeats in the cybrid mtDNA. Also, reads at the end of each contig had their mate reads in a distant region of the same or other contig, indicating the existence of alternative configurations of the cybrid mitochondrial

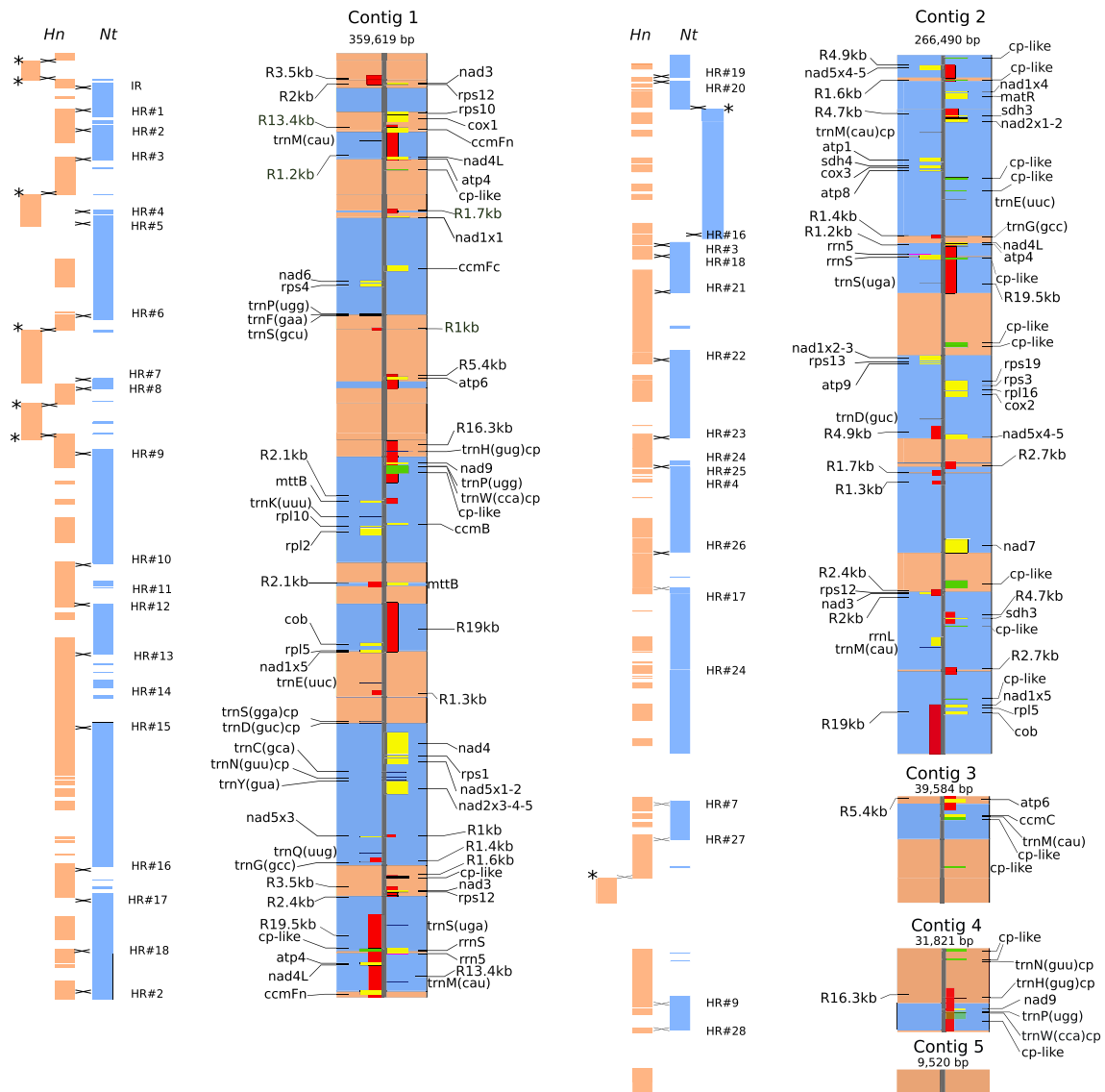


Fig. 2 Mitochondrial genome of the repeated cybrid Nt[+Nt(+Hn)]. Scale maps of the five mitochondrial contigs totaling 707,034 bp. The alternate wide orange and blue boxes on the contig maps represent regions inferred to derive from the parents *H. niger* or *N. tabacum*, respectively. The cybrid contig maps show full-length genes, repeats > 1 kb in length (red boxes labeled “R”, followed by the repeat length in kb), and chloroplast-derived regions > 300 bp in length (green boxes). Genes marked on different sides of the verti-

cal lines are transcribed from opposite strands of the genome (yellow boxes). The thinner orange and blue boxes to the left of the cybrid contig maps indicate regions of the cybrid mtDNA shared (as defined by BLAST hits) with the mitochondrial genomes of *H. niger* (*Hn*) or *N. tabacum* (*Nt*), respectively. Crossing lines between BLAST hits indicate the 28 homologous (HR) and one illegitimate (IR) intergenomic recombination events, respectively. Intragenomic homologous recombination events are indicated with asterisks

genome. The GC content of the genome was 44.6%, ranging between 44.2 and 44.9% in each individual contig.

Recombination events in the cybrid mitochondria

Once the parental mitochondrial genomes were united inside the cybrid mitochondria following mitochondrial fusion, they were able to undergo homologous recombination across the > 1000 regions of 100–9000 bp with high

identity between the two genomes. These homologous regions formed non-identical repeats (95–100% identity) in the initial cybrid mitochondria.

By comparing the cybrid and parental mtDNAs, we first identified the regions inherited from each parent based on sequence identity. We found 47 interspersed tracts of tobacco and *H. niger* with 99–100% identity to the parental sequences (Fig. 2). The analyses of the cybrid BLASTn hits to each parental mtDNA (orange and blue thin boxes on the

left of each cybrid contig in Fig. 2) indicated that the recombination events took place at homologous regions (with one exception) of variable lengths between the parental mtDNAs. We inferred 28 intergenomic homologous recombination events (HR#) giving rise to the chimeric Mv-1-1g mtDNA (Fig. 2). The estimated length of the parental recombining regions was 2500 bp in average, ranging from 205 to 9850 bp and 94–100% sequence identity (Table 1). Based on the estimated minimal length of homology required by the enzyme RecA (Lovett et al. 2002; Shen and Huang 1986), one case of illegitimate recombination was inferred, in which the parental recombining regions shared only 19 bp (Fig. 2 and Table 1).

Overall, the recombination events were unevenly distributed across the cybrid mtDNA (Fig. 2). The two farthest consecutive points of recombination were located at ~54 kb from each other (HR#15 and HR#16), and the closest ones were positioned at a distance of 1268 bp (Fig. 2 and Table 1). Of the 29 intergenomic recombination events, eleven took place within or near coding regions resulting in five chimeric genes (see below).

In addition, we identified eight intragenomic homologous recombination events in the cybrid mtDNA, *i.e.* recombination between two distant regions derived from the same parent (asterisks in Fig. 2). The recombining regions are repeats of estimated average length 510 bp, ranging from 73 to 2623 bp and 86–100% identity. Noticeably, most of the intragenomic recombination events (87.5%) took place within sequences of the *H. niger* mtDNA. This might be explained by the high frequency of repeats in the *H. niger* mtDNA that could generate alternative genomic configurations through recombination events (Sanchez-Puerta et al. 2015). These intragenomic recombination events could have taken place in *H. niger* before protoplast fusion and cybrid production or within the cybrid mitochondria.

Parental contribution to the cybrid mitochondrial genome

Overall, 63% and 37% of the cybrid mtDNA were inherited from tobacco and *H. niger*, respectively. To examine the parental composition of the cybrid mtDNA in more detail, we distinguished three classes of sequences that contributed to the formation of the cybrid mitochondria: (i) those shared by the two parents, *i.e.* homologous regions that included all known genes and many intergenic sequences; these sequences turned into non-identical repeats when the parental mtDNAs recombined to form the cybrid; (ii) those exclusive to *H. niger*; and (iii) those exclusive to *N. tabacum*. The repeated cybrid genome conserved one copy of almost all the *N. tabacum* exclusive sequences (99%) and 82% of those exclusive of *H. niger*. Overall, only 56 kb (11%) of the non-repeated sequences were eliminated. In stark contrast,

194 kb (~48%) of the repeated sequences were lost (Fig. 3), leaving at least one copy of the homologous sequences in the cybrid Mv-1-1g. Of the 203 kb non-identical repeats initially present in the cybrid mitochondria (totaling 406 kb), only 10.5 kb sequences were retained from both parents (including 3 genes), and 172 kb and 19.1 kb were maintained from *N. tabacum* or *H. niger*, respectively (Fig. 3). Most coding sequences (yellow areas in Fig. 3) were retained from *N. tabacum*.

If we consider the double contribution of the tobacco mitochondria in the two successive protoplast fusions, the overall loss of repeated sequences increases significantly. The whole length of the tobacco mtDNA should be considered as a repeat, and because the tobacco-derived regions were not found twice in the cybrid mtDNA (with very few exceptions), then, an extra 431 kb of repeated regions were also lost along the whole process.

Mitochondrial gene content

Without counting repeats, the mtDNA of Mv-1-1g contained genes for 37 proteins, three rRNAs, and 21 tRNAs, *i.e.* the same number of mitochondrial genes as in the parental lines. The tRNA gene content is identical to tobacco, and not to *H. niger*, because the gene *trnD* is present while the gene *trnL-cp* is absent. In addition, the Mv-1-1g genome had two full-length copies of 11 genes (including the gene regions *nad1_exon5* and *nad5_exons4* and 5), and four genes were triplicates. Seven of these multi-copy genes resulted from the retention of full-length copies of each parent (*nad1* × 5, *nad3*, *rps12*) or through the formation of chimeric genes originated from both parents (*atp4*, *ccmFn*, *rpl5*, *rrnS*). The other duplicates were inherited from a single parent: *atp6*, *cob*, *nad9*, *mttB*, *rrn5*, and *sdh3* (Table 2).

The protein-coding chimeric genes in the mtDNA of Mv-1-1g, *ccmFn* (two identical copies), *atp4*, and *rpl5*, had nonsynonymous differences in respect of the parents, so the translated proteins would also be chimeric (Fig. 4A). The genes *ccmFn* and *atp4* produced chimeric proteins that had eight or two different amino acids compared to the proteins in either parent, respectively. Finally, *rpl5* showed two or three nonsynonymous changes compared to *H. niger* or tobacco, respectively (Fig. 4A). Two of the genes that were inherited from both parents (*nad3*, *rps12*) had nonsynonymous differences between them and similar hydrophobicity profiles (Fig. 4B), while *nad1x5* was identical in both parents.

Of those inherited from a single parent, 27 protein-, two rRNA-, and 14 tRNA-genes originated exclusively from *N. tabacum*, while only three protein and two tRNA genes derived exclusively from *H. niger* (Table 2). The three protein genes inherited only from *H. niger* (*atp6*, *co* × 1, and *rps10*) had nonsynonymous differences from those

Table 1 Events of recombination between the parental mitochondrial genomes (*Nicotiana tabacum* and *Hyoscyamus niger*) in the cybrid Mv-1-g

Recombination event	Length of the recombining regions (bp)	Number of recombination products	Location of recombination			Estimated mechanism of recombination
			Contig coordinates	Within a repeat > 1 kb?	Within a coding region?	
Illegitimate recombination	19	1	Contig 1: 13,264–13,280		No	Illegitimate recombination through MMBIR
HR#1	2797	1	Contig 1: 22,281–22,322		No	BIR or dHJ resolved through CO
HR#2 (duplicated)	2784	1	Contig 1: 29,413–29,969 and 356,835–35,7391	Repeat 13.4 kb	ccmFn	BIR or dHJ resolved through CO
HR#3	1274;1268	2	Contig 1: 40,094–40,414 and Contig 2: 71,414–71,858	Repeat 1.2 kb	atp4; nad4L	dHJ resolved through CO
HR#4	1806;1796	2	Contig 1: 59,848–60,456 and Contig 2: 159,047–159,516	Repeat 1.7 kb	No	dHJ resolved through NCO
HR#5	4000	1	Contig 1: 62,209–62,755		nad1x1	BIR or dHJ resolved through CO
HR#6	2263	1	Contig 1: 99,430–99,589		trnF	BIR or dHJ resolved through CO
HR#7 (duplicated)	4035	1	Contig 1: 124,863–125,009 and Contig 3: 2760–2906	Repeat 5.4 kb	No	BIR or dHJ resolved through CO
HR#8	4035; 4035	2	Contig 1: 127,505–127,569	Repeat 5.4 kb	No	BIR
HR#9 (duplicated)	4461	1	Contig 1: 153,523–153,668 and Contig 4: 20,699–20,844	Repeat 16.3 kb	No	BIR or dHJ resolved through CO
HR#10	867	1	Contig 1: 193,785–194,096		No	BIR or dHJ resolved through CO
HR#11	2160; 2150	2	Contig1: 201,402–202,870	Repeat 2.1 kb	mttB	SDSA or dHJ dissolution.
HR#12	909; 909	2	Contig1: 209,565–209,681	Repeat 19 kb	No	BIR
HR#13	6515; 6532	2	Contig 1: 227,466–227,797	Repeat 19 kb	rpl5	BIR
HR#14	1323; 1301	2	Contig 1: 244,829–245,456	Repeat 1.3 kb	No	SDSA or dHJ dissolution
HR#15	11521	1	Contig 1: 255,104–255,156		No	BIR or dHJ resolved through CO
HR#16	1457;1462	2	Contig 1: 309,198–309,283 and Contig 2: 68,948–69,033	Repeat 1.4 kb	No	dHJ resolved through CO
HR#17	2271;2477	2	Contig 1: 320,931–320,984 and Contig 2: 204,640–204,666	Repeat 2.4 kb	No	dHJ resolved through CO
HR#18 (duplicated)	5166	1	Contig 1: 341,742–342,696 and Contig2: 76,106–77,060	Repeat 19.5 kb	rrnS	SDSA, dHJ resolved through NCO, or dHJ dissolution

Table 1 (continued)

Recombination event	Length of the recombining regions (bp)	Number of recombination products	Location of recombination			Estimated mechanism of recombination
			Contig coordinates	Within a repeat > 1 kb?	Within a coding region?	
HR#19	205; 207	2	Contig 2: 8543–8624	Repeat 1.6 kb	No	BIR
HR#20	226; 219	2	Contig 2: 9992–10,107	Repeat 1.6 kb	No	BIR
HR#21	8986; 8986	2	Contig 2: 90,553–90,835	Repeat 19.5 kb	No	BIR
HR#22	1989	1	Contig 2: 113,811–114,537		No	BIR or dHJ resolved through CO
HR#23	1861; 1868	2	Contig 2: 145,963–146,237	Repeat 4.9 kb	nad5x5	BIR
HR#24	1682; 3220	2	Contig 2: 155,449–155,768 and 234,253–235,034	Repeat 2.7 kb	No	dHJ resolved through NCO
HR#25	678; 1185	2	Contig 2: 156,753–157,001	Repeat 2.7 kb	No	BIR
HR#26	6680	1	Contig 2: 189,075–189,866		nad7x4	BIR or dHJ resolved through CO
HR#27	2048	1	Contig 3: 15,707–16,024		No	BIR or dHJ resolved through CO
HR#28	1253; 1253	2	Contig 4: 30,891–30,974	Repeat 16.3 kb	No	BIR

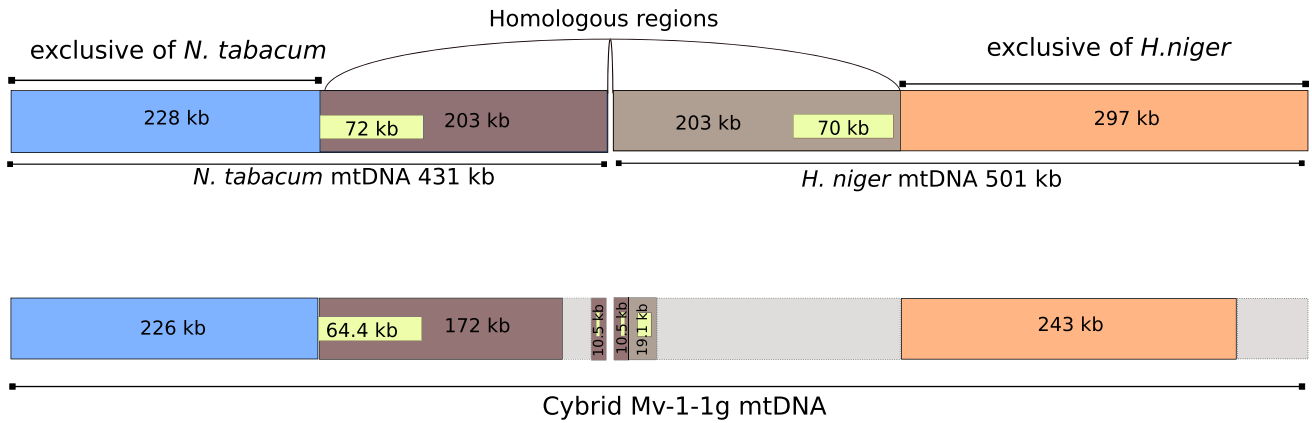


Fig. 3 Origin of the cybrid mtDNA with respect to its parental composition. Top: schematic representation of the parental mitochondrial genomes (*Nicotiana tabacum* and *Hyoscyamus niger*) that contributed to the formation of mtDNA in the repeated cybrid Nt [+Nt(+Hn)] through two protoplast fusions. Total amount of exclusive and shared parental sequences are drawn to scale. Bottom: summary status of the composition of the cybrid mtDNA. Colored regions indicate

sequences derived from tobacco (blue and dark brown) or *H. niger* (orange and light brown). Yellow regions represent coding sequences. Parental regions not present in the cybrid are shown in grey. Novel repeats in the cybrid mtDNA are not included, which explains the difference between the total length of the cybrid and the estimated amount of homologous and exclusive sequences derived from each parental mitochondria

in *N. tabacum* and different hydrophobicity patterns. In addition, the aforementioned genes had differences in the predicted editing sites but these differences are functionally irrelevant because *H. niger* genes had a subset of the tobacco editing sites. Also, the gene *coxI* has an intron in

H. niger but not in tobacco. In a previous work on a cybrid between *H. niger* and *N. tabacum* (Sanchez Puerta et al. 2015), we showed that the *H. niger*-derived intron-containing *coxI* gene was correctly spliced and edited in the tobacco nuclear background. This observation suggested that nuclear factors for *coxI* splicing are present in the

Table 2 Comparison of gene content in the mitochondrial genomes of *Nicotiana tabacum* (Nt), *Hyoscyamus niger* (Hn) and in the cybrids (Mv-1-1g and FH-4)

Gene name ^a	Length (in bp) ^b	Hn/Nt differences	Hn/Nt differences in coding regions		Number of intact genes				Origin of cybrid genes	
			Synonymous	Nonsynonymous	Hn	Nt	Cybrid Mv-1-1g	Cybrid FH-4 ^c	Mv-1-1g	FH-4 ^c
<i>atp1</i>	1530	25	19	6	1	1	1	2	Nt	Chimeric & Nt
<i>atp4</i>	597	5	1	4	1	1	3	1	2 Nt & 1 chimeric	Nt
<i>atp6</i>	1188	3	2	1	1	1	2	1	2 Hn	Nt
<i>atp8</i>	471	4	2	2	1	1	1	1	Nt	Nt
<i>atp9</i>	234	2	0	2	1	1	1	2	Nt	Nt & Hn
<i>ccmB</i>	621	2	1	1	1	1	1	2	Nt	2 Nt
<i>ccmC</i>	753	0	0	0	1	1	1	1	Nt	Hn
<i>ccmFc</i>	2271	11	0	0	1	1	1	1	Nt	Nt
<i>ccmFn</i>	1740	37	11	16	1	1	2	2	2 chimeric	2 Hn
<i>cob</i>	1182	3	2	1	1	1	2	1	2 Nt	Nt
<i>cox1</i>	2551	11	9	2	1	1	1	2	Hn	2 Hn
<i>cox2</i>	2143	9	2	4	1	1	1	1	Nt	Chimeric
<i>cox3</i>	798	2	1	1	1	1	1	1	Nt	Nt
<i>matR</i>	1977	5	1	4	1	1	1	1	Nt	Chimeric
<i>mttB</i>	840	1	0	1	1	1	2	1	2 Nt	Nt
<i>nad1*</i>										
x1	387	0	0	0	1	1	1	1	ND ^d	Nt
x2-3	1485	12	1	0	1	2	1	1	Nt	Nt
x4	59	0	0	0	1	1	1	1	Nt	Hn
x5	259	0	0	0	1	1	2	1	Nt & Hn	Nt
<i>nad2*</i>										
x1-2	1575	6	0	0	1	2	1	1	Nt	Nt
x3-5	4796	28	0	1	1	1	1	1	Nt	Nt
<i>nad3</i>	357	4	2	2	1	2	3	2	2 Hn & 1 Nt	2 Nt
<i>nad4</i>	8482	38	1	0	1	1	1	1	Nt	Chimeric
<i>nad4L</i>	303	0	0	0	1	1	3	1	2 Nt & ND ^d	Nt
<i>nad5*</i>										
x1-2	2290	8	1	1	1	1	1	1	Nt	Nt
x3	22	0	0	0	1	1	1	1	Nt	Nt
x4-5	1366	19	1	0	2	1	2	1	2 Nt	Chimeric
<i>nad6</i>	687	5	2	3	1	1	1	1	Nt	Nt
<i>nad7</i>	5291	46	0	2	1	1	1	1	Nt	Nt
<i>nad9</i>	573	2	0	2	2	1	2	1	2 Nt	Nt
<i>rpl2</i>	2898	47	4	4	1	1	1	1	Nt	Nt
<i>rpl5</i>	555	9	4	5	1	1	2	1	1 chimeric & 1 Nt	Nt
<i>rpl10</i>	480	4	2	2	1	1	1	2	Nt	2 Hn
<i>rpl16</i>	516	3	1	2	1	1	1	1	Nt	Hn
<i>rps1</i>	651	3	1	2	1	1	1	1	Nt	Nt
<i>rps3</i>	3167	27	6	4	1	1	1	1	Nt	Hn
<i>rps4</i>	1050	6	5	1	1	1	1	1	Nt	Nt
<i>rps10</i>	1138	4	1	2	1	1	1	2	Hn	2 Hn
<i>rps12</i>	378	2	1	1	1	1	3	1	2 Hn & 1 Nt	Nt
<i>rps13</i>	351	3	1	2	1	1	1	1	Nt	Nt
<i>rps19</i>	285	3	1	2	1	1	1	2	Nt	2 Hn
<i>rrn5</i>	119	1	NA	NA	1	1	2	1	2 Nt	Nt
<i>rrnS</i>	1947	4	NA	NA	1	1	2	1	2 chimeric	Nt

Table 2 (continued)

Gene name ^a	Length (in bp) ^b	Hn/Nt differences	Hn/Nt differences in coding regions		Number of intact genes				Origin of cybrid genes	
			Synonymous	Nonsynonymous	Hn	Nt	Cybrid Mv-1-1g	Cybrid FH-4 ^c	Mv-1-1g	FH-4 ^c
<i>rrnL</i>	3406	15	NA	NA	2	2	1	1	Nt	Chimeric
<i>sdh3</i>	327	3	1	2	1	3	2	3	2 Nt	3 Nt
<i>sdh4</i>	378	0	0	0	1	1	1	1	Nt	Nt

NA not applicable

^aIntron-containing genes are in bold font. Asterisks denote trans-spliced genes

^bThe gene lengths correspond to those in the cybrid Mv-1-1g mtDNA

^cThe FH-4 cybrid was described in Sanchez Puerta et al. (2015)

^dND The origin of a gene copy is not determined because it is located within a recombining region in which the parental sequences are identical

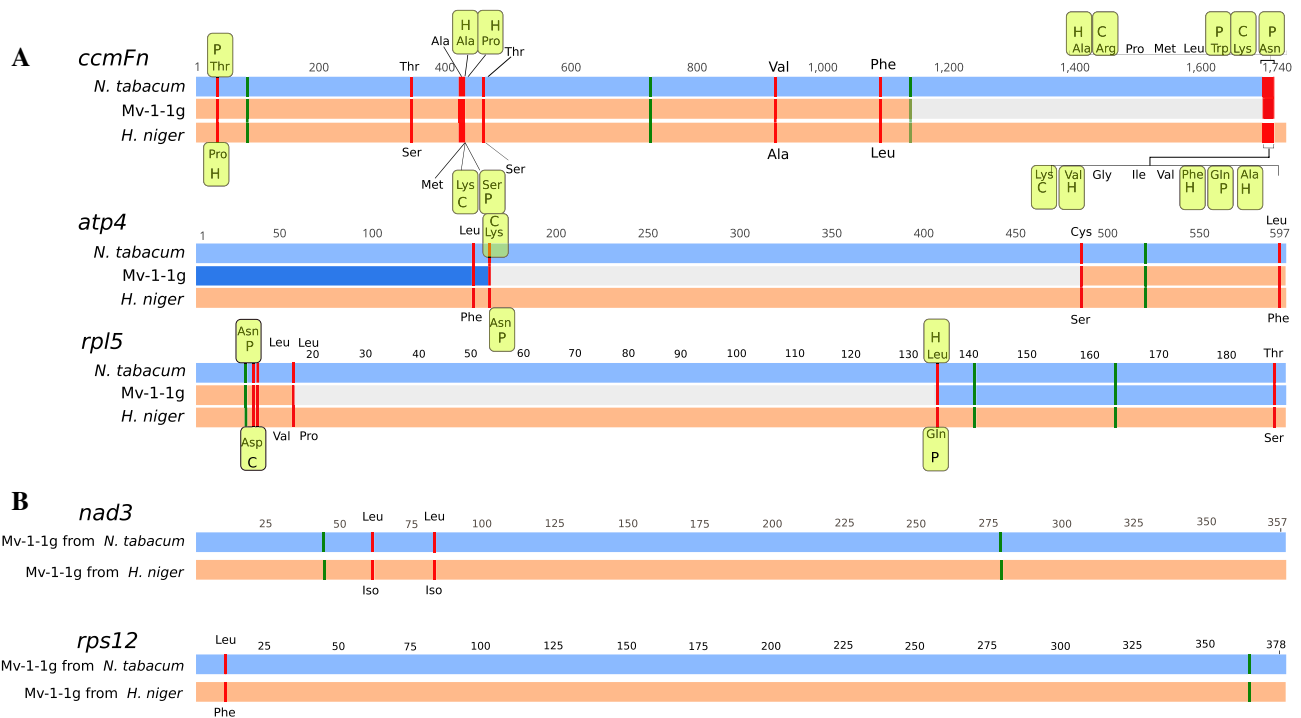


Fig. 4 Schematic representation of differences in chimeric and two parent-derived genes in the cybrid Mv-1-1g mtDNA with respect to the parental sequences. **A** Chimeric genes (*ccmFn*, *atp4*, and *rpl5*) in the cybrid Mv-1-1g formed by intergenomic recombination of parental mitochondria. **B** Genes in the cybrid (*nad3*, *rps2*) found in two copies, one derived from each parent. Blue and orange horizontal boxes indicate sequences derived (or from) *N. tabacum* and *H. niger*, respectively. Gray boxes depict crossover regions in which the paren-

tal sequences are identical. Red and green marks indicate nonsynonymous and synonymous differences, respectively, between the parental lines. The encoded amino acids are shown for the nonsynonymous differences. Amino acids with hydrophobic (H), polar (P), or charged (C) side chains are indicated within yellow rectangles where the nonsynonymous differences implicate a change in the amino acid chemical group. Numbers denote sequence length in bp

tobacco nuclear genome, and probably also widely found in plants (Sanchez Puerta et al. 2015).

ORFs in the cybrid mtDNA

We searched for uncharacterized ORFs in the cybrid mtDNA that could be related to the CMS condition and the atypical flower phenotype. In the cybrid mtDNA those ORFs inherited exclusively from *H. niger* are now in a foreign nuclear

Table 3 Comparisons between the mitochondrial genomes resulted from cybridization between *Nicotiana tabacum* (Nt) and *Hyoscyamus niger* (Hn)

Cybrids between Nt and Hn		Mv-1-1g	FH-4
Genome size (bp)		707,034	705,036
Genome composition	Derived from Nt	63.00%	58.00%
	Derived from Hn	37.00%	42.00%
Intergenic recombination events	Illegitimate recombination	1	0
	Homologous recombination	28	33
Coding sequences	Nt	41	30
	Hn	8 (<i>atp6</i> (x2), <i>cox1</i> , <i>nad3</i> (x2), <i>rps10</i> , <i>rps12</i> (x2))	14 (<i>atp9</i> , <i>ccmC</i> , <i>ccmFn</i> (x2), <i>cox1</i> (x2), <i>rpl10</i> (x2), <i>rpl16</i> , <i>rps3</i> , <i>rps10</i> (x2), <i>rps19</i> (x2))
	From both parents	<i>nad3</i> , <i>rps12</i>	<i>atp9</i>
	Chimeric	7 (<i>atp4</i> , <i>ccmFn</i> , <i>nad1</i> , <i>rpl5</i> , <i>rrnS</i>)	7 (<i>atp1</i> , <i>cox2</i> , <i>matR</i> , <i>nad1</i> , <i>nad4</i> , <i>nad5</i> , <i>rrnL</i>)

background, and could be involved in the CMS phenotype (Leino et al. 2005). We identified 66 ORFs derived from *H. niger* that did not show sequence similarity to ORFs in *N. tabacum* mtDNA. Two of those ORFs showed one of the most common features found in CMS-related genes, i.e. they start with partial sequences of known mitochondrial genes followed by an uncharacterized region (Chen and Liu 2014). One of them, ORF281 (located in contig 1) of 843 bp, begins with the first 66 bp of the gene *rps13*. The other, ORF289 (located in contig 2) of 867 bp, begins with the first 173 bp of the gene *nad9*.

In addition, new ORFs could arise in the cybrid mtDNA by intergenomic recombination. These newly formed ORFs are considered as potential candidates for determination of CMS (Chase 2007; Touzet and Meyer 2014). We found two ORFs located at the sites of recombination in the cybrid Mv-1-1g mtDNA. ORF326 of 978 bp was a new ORF with a domain from DNA polymerase B2 superfamily identified by HMMER (Finn et al. 2015). ORF117 (located in contig 1) was chimeric and showed nonsynonymous substitutions to similar ORFs in the parental mtDNAs.

Common patterns in the establishment of the mitochondrial genome in two cybrids produced independently in the same parental combination, *N. tabacum* + *H. niger*

Previously, the mtDNA of another cybrid between tobacco and *H. niger*, Nt(+Hn) FH-4, was reported (Sanchez-Puerta et al. 2015). By comparing the Mv-1-1g and FH-4 mitochondrial genomes, we identified several common features as a result of parallel molecular processes that took place in each cybrid (Table 3). In both cases the mtDNA was ~700 kb, that is 1.4–1.6 times greater than that of either parent but 24% less than the sum of both parents. Both cybrids had highly chimeric mitochondria emerged from 28 to 33 homologous

recombination events, and a single case of illegitimate recombination in Mv-1-1g (Tables 1, S1). We compared the location of the recombination events via non-identical repeats using *H. niger* mtDNA as a reference (Fig. 5). Half of the recombination sites involved the same non-identical repeats in both cybrids, and six of those occurred in the same location of the parental genome.

In terms of the parental origin of the mtDNAs, the majority of sequences (58–63%) derived preferentially from tobacco (Table 3). Even a more interesting bias involved the preferential retention (~89%) of those sequences that were exclusive to each parent and loss of a significant fraction of the sequences shared by both parents (48–49%). In detailed genomic comparisons, we found that both cybrids shared 70% of the regions of *H. niger* retained in their mitochondria (Fig. 5). This observation might indicate the existence of a universal mechanism directing the bias loss of those sequences shared by the parents.

The majority of the coding sequences were inherited from tobacco, although 22–38% of the protein-coding genes of *H. niger* were maintained in both cybrid mtDNAs. Interestingly, the mtDNAs contained single alleles for >90% of the genes, with only a few exceptions in which both parental forms were present (Table 3).

Discussion

Somatic hybridization is a useful tool for bypassing the incompatibility barriers in sexual crosses of distinct plant species, as well as the barriers imposed by the uniparental mode of transmission of mitochondrial genes in sexual crosses of plants (Greiner and Bock 2013). Protoplast fusion leads to de novo nuclear and cytoplasmic genome combinations, and plant hybrids almost invariably contain recombinant mitochondrial genomes (Akagi et al. 1995; Aleza et al.

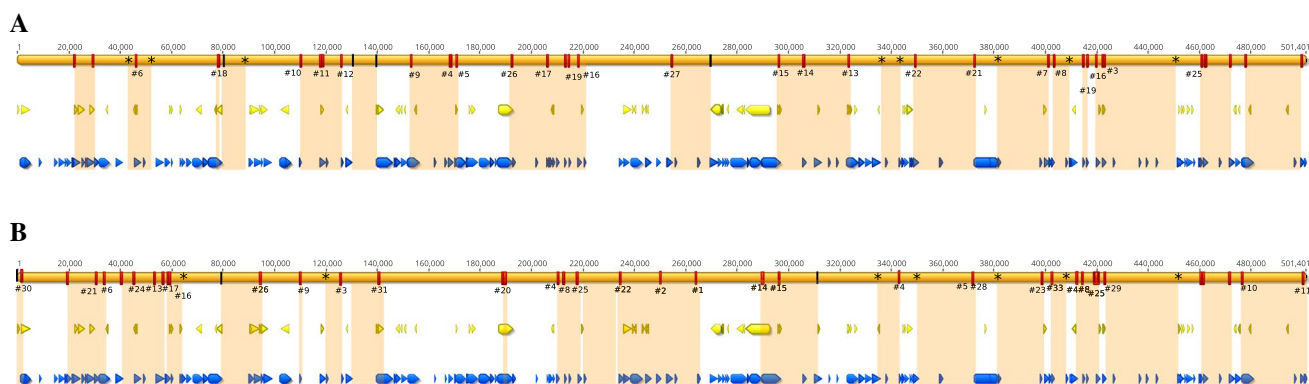


Fig. 5 Recombination events and mitochondrial DNA regions of *Hyoscyamus niger* retained in the cybrids, and recombination events. A linear map of the *H. niger* mitochondrial genome (in orange) shows the regions (in light orange) that were retained in the cybrids Mv-1-1g (A) and FH-4 (B). Yellow and blue arrows indicate coding

sequences and shared regions between tobacco and *H. niger* mitochondrial genomes, respectively. Red marks on the genome depict intergenomic recombination events identified in the cybrids Mv-1-1g (A) and FH-4 (B). Asterisks and black marks indicate intragenomic recombination events and the end of the contigs, respectively

2016; Belliard et al. 1979; Xiang et al. 2004). This turns somatic hybrids into an interesting subject for further studies of mitochondrial genome recombination, evolutionary dynamics, and nuclear-cytoplasmic coordination. Cybrids obtained via somatic hybridization might represent a valuable technological tool for transferring organelle-encoded traits, such as CMS, between cultivars (Gleba and Sytnik 1984; Izhar and Zelcer 1986; Liu et al. 2005). The CMS condition can result from the formation of chimeric mitochondrial genes, which are frequently formed in somatic hybrids and mitochondrial mutants (Chase 2007; Kim and Kim 2006). Little is known about the overall extent of mitochondrial recombination in somatic hybrids, the mechanisms involved, and the constraints operating on mtDNA recombination in the context of nuclear-cytoplasmic interactions and incompatibilities. The present study has been conducted in an attempt to advance our understanding of plant mitochondrial genome recombination, and their potential technological applications in plant breeding. In particular, we aimed to elucidate the mechanisms of mitochondrial genome recombination in plants by comparing cybrid and parental mitochondrial genomes through deep sequencing.

Evaluating the genetic basis of the CMS phenotype

The origin and mechanisms of CMS are still mysterious (Charlesworth 2017; Chase and Gabay-Laughnan 2004; Hanson and Bentolila 2004). Unusual ORFs determining CMS have been discovered in a wide range of plant species, showing considerable diversity of the proteins involved and the specific phenotypic effects (Chase 2007). Most of the mitochondrial-encoded loci responsible for CMS that were reported to date shared two features: they are ORFs formed

by sequences of unknown origin combined with pieces of standard mitochondrial genes, and they are expressed because they are fused directly to a mitochondrial promoter region or they are co-transcribed with upstream mitochondrial genes (Carlsson et al. 2008; Chase 2007; Chase and Gabay-Laughnan 2004; Hanson and Bentolila 2004; Schnable and Wise 1998). In other cases, CMS genes were not chimeric and, instead, contained sequences from a single source (Bonhomme et al. 1992; Ducos et al. 2001; Okazaki et al. 2013).

The cybrid analyzed in this work manifested CMS, and this condition was maintained after backcrosses with *N. tabacum* as well as after a second fusion event with a different nuclear background of *N. tabacum*. These experiments eliminated nuclear-factors as the cause of CMS (Zubko et al. 1996). Instead, the origin of such an important feature should be associated with changes in the mitochondrial genome, or resulted from nuclear-mitochondrial incompatibilities. In this work, we found unusual ORFs and features in the mitochondrial genome that could be responsible for the CMS phenotype in the Mv-1-1g repeated cybrid. For example, the mtDNA of the cybrid Mv-1-1g contained respiratory complexes I, IV, and V formed by subunits encoded by different combinations of *H. niger*-derived (i.e. *nad1*, *nad3*, *atp6*, *cox1*) and/or chimeric (i.e. *atp4*) mitochondrial genes, together with tobacco-derived genes (i.e. *nad2*, *nad4*, *nad5*, *cox2*, *cox3*, *atp1*, *atp8*, *atp9*) in a tobacco nuclear background. The assembly of protein complexes formed by genes of different origin, carrying new mutations, could cause steric incompatibility, reduction of substrate affinity, and prevention of activity of respiratory complexes, causing CMS (Ducos et al. 2001; Pineau et al. 2005; Rand et al. 2004).

Despite containing CMS genes, plants could be male fertile due to the presence of restorer genes in their nuclear genomes. When these pre-existing CMS genes are moved to a novel nuclear background, through genetic crossing or somatic hybridization, suppressors are removed and CMS is expressed (Carlsson et al. 2008; Leino et al. 2005). In the mtDNA of the Mv-1-1g cybrid, there are 66 ORFs derived from *H. niger* that do not have similarity with ORFs in *N. tabacum* mtDNA. Even though *H. niger* is male fertile, these ORFs could be involved in the CMS phenotype observed in the repeated cybrid. Of the 66 ORFs derived from *H. niger*, two were chimeric, beginning with a short piece of the genes *rps13* and *nad9*, respectively, and represent worthy candidates for CMS.

Variety of new ORFs and other reconstructed mtDNA regions implies a unique avenue for future work towards clarification of their roles in particular phenotypic peculiarities of CMS flower composition and other developmental features of cybrids. Extended investigation of links between mtDNA changes and various phenotypic traits in cybrids derived from the same remote parents might bring in depth understanding of nucleo-cytoplasmic communications as well as a potentially new source of genetic variation for evolution and breeding.

Recombination events in the cybrid mitochondria

Plant mtDNAs undergo frequent recombination processes that have central roles in DNA repair and replication (Backert and Börner 2000; Kajander et al. 2001; Ling et al. 1995). Studies on plant mitochondrial recombination and repair (focused mainly on *Arabidopsis* nuclear mutants) described the role of nuclear-encoded genes, such as OSBs (Zaegel et al. 2006) and homologs of the bacterial genes *MutS* and *recA* (Arrieta-Montiel et al. 2009; Davila et al. 2011; Miller-Messmer et al. 2012; Shedge et al. 2010). Rearrangements in plant mtDNAs can be explained by the action of homologous recombination (HR) pathways or through illegitimate recombination involving sequence microhomologies (Kubo and Newton 2008; Kühn and Gualberto 2012). Several studies suggested that the minimal length of the homology region required for HR is 24–40 bp (depending on the organism) although its efficiency increases rapidly with the size of the homologous regions that undergo recombination (Hua et al. 1997; Shen and Huang 1986; Sugawara et al. 2000).

Double strand breaks (DSBs) arise from exogenous and endogenous sources and can be repaired by three major HR mechanisms (Mehta and Haber 2014). A DSB leads to 5' → 3' degradation of the broken ends resulting in 3' single-stranded DNA tails (Fig. 6). The activation of different pathways to repair the DSB depends on the structure of the broken ends of the DNA and the factors available (Mehta and Haber 2014). In the double Holliday Junction

(dHJ, also known as Double Strand Break Repair DSBH) pathway, two Holliday junctions are formed and, depending on the resolution, it can result in crossover (CO) or non-crossover (NCO) products (Resnick 1976; Szostak et al. 1983). Alternatively, the dHJ can be dissolved via producing gene conversion products (Wu and Hickson 2003). In the Single-Dependent Strand Annealing (SDSA) pathway, strand invasion is rejected and the 3' ssDNA tail aligns with the other end of the DSB that results in a faithful repair without crossover products, leading to gene conversion (Pâques and Haber 1999). Break Induced Replication (BIR) plays an important role in the repair of the one-ended DSBs that can be formed as result of a collapse or arrest of the replication fork or when only one end of the DSB shares homology with a donor sequence (Anand et al. 2013). Strand invasion occurs and the synthesis proceeds to the end of the DNA molecule, resulting in a non-reciprocal crossover (Anand et al. 2013; Malkova et al. 1996, 2005).

At the time of protoplast fusion, the set of homologous sequences (> 1000 regions) between the two parental mtDNAs became non-identical repeats allowing the process of intergenomic HR. We analyzed the products of recombination between the two parental genomes to infer which mechanisms took place during cybrid formation (Table 1, Fig. 6). We identified three events of symmetric recombination produced by the dHJ pathway resolved towards CO (HR#3, HR#16 and HR#17), nine cases of non-reciprocal recombination (HR#8, HR#12, HR#13, HR#19, HR#20, HR#21, HR#23, HR#25, HR#28) indicating the involvement of the BIR pathway, and two events that could be explained by dHJ with NCO resolution (HR#4 and HR#24). Two cases (HR#11 and HR#14) could be explained either by SDSA or by dHJ dissolution. In addition, 11 single products of recombination were present in the cybrid mtDNA, for which the recombination pathway could not be determined (Table 1). To elucidate the mechanisms that gave rise to these single products, we tested the presence of alternative recombination products by aligning the paired-end reads to sequences designed to represent possible outcomes considering the main HR mechanisms: BIR and dHJ resolved as CO. We identified a few paired-end reads that mapped to one of the alternative products in 10 of the 12 events with single products. Interestingly, we detected only regions derived from tobacco, with one exception, and at very low stoichiometry (read depth equal to 0.3–3). The presence of those tobacco-derived regions suggests the implication of the BIR pathway.

Overall, BIR was the most frequent mechanism in agreement with previous observations in another cybrid (Table S1; Sanchez-Puerta et al. 2015). This could be due to frequent occurrence of stalled replication forks (Anand et al. 2013). Alternatively, a DSB close to the end of a homologous region in the parental genomes would lead to BIR because one end will not find a homology region to

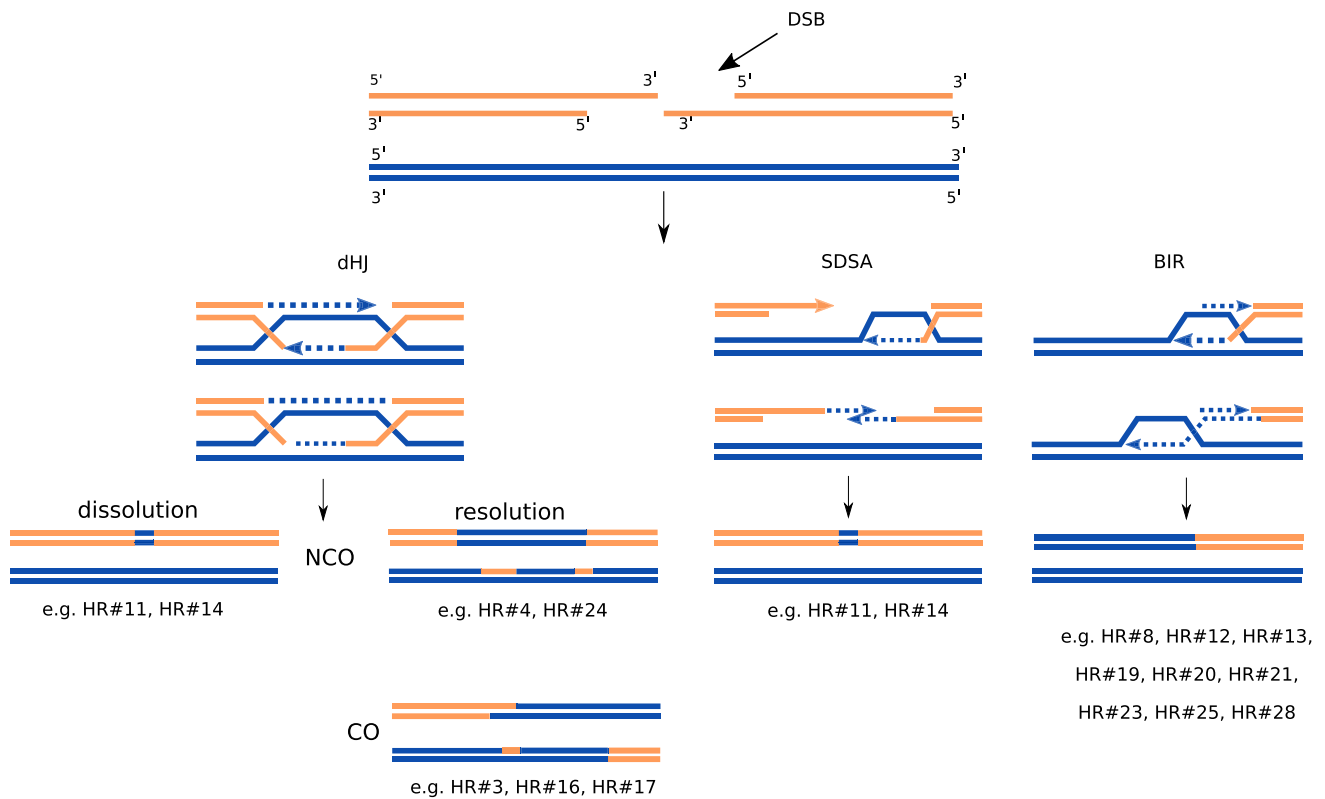


Fig. 6 Homologous recombination pathways and mechanisms of DSB repair. DSBs are 5' digested resulting in 3' single-stranded DNA tails. The double Holliday Junction (dHJ) pathway involves two strand invasions and can be resolved either by helicase and topoisomerase-mediated dissolution resulting in noncrossover (NCO) products, or cleaving with HJ resolvases to produce both crossover (CO) or NCO outcomes. In the Synthesis Dependent Strand Annealing (SDSA) pathway, the newly synthesized strand dissociates from the D-loop and results in a NCO outcome with no change to the tem-

plate DNA. The Break Induced Replication (BIR) pathway involves both leading and lagging strand synthesis and results in the loss of heterozygosity or a nonreciprocal translocation (if the template is located ectopically). Newly synthesized DNA is depicted as dashed lines in the same color as the template; arrowheads indicate direction of DNA synthesis. Examples of inferred homologous recombination pathways in the cybrid Mv-1-1g mtDNA are indicated. Blue and orange horizontal lines indicate sequences derived or from *N. tabacum* and *H. niger*, respectively

repair the DSB by gene conversion. BIR is a highly mutagenic pathway because it can result in rearrangements and the loss of a large piece of DNA (Sakofsky et al. 2012). It has been suggested that BIR is suppressed when a DSB has two ends and a more conservative HR pathway takes place (Llorente et al. 2008).

In addition to HR, there are mechanisms to repair DSB that require very short or no homology regions. Recombination at microhomologies causes illegitimate recombination through three molecular pathways: Non Homologous End Joining, (NHEJ), Microhomology-Mediated End Joining (MMEJ), and Microhomology-Mediated Break Induced Replication (MMBIR) (Hanson and Bentolila 2004; Hastings et al. 2009; Kubo and Newton 2008; Schnable and Wise 1998). NHEJ is the predominant repair mechanism for DNA DSB in the nuclear genome (Chang et al. 2017). This mechanism directly ligates broken DNA strands, without 5' resection, and requires microhomologies of 1–4 bp (Lieber 2010). This pathway can result in blunt joints or

small deletions or insertions in the breakpoint junction. MMEJ starts with the digestion of 5' DNA strand to obtain 3' DNA tails. The exposed microhomologies anneal, and the gaps are filled. This mechanism results in deletion of the DNA regions flanking the original break; it is an error prone repair mechanism (Bennardo et al. 2008). MMBIR is a replicative microhomology mediated mechanism. In this model, the replication fork stalls with the following resection of 5' ends, generating 3' strand overhangs, which can invade a region with microhomology on a different DNA template to establish a new replication fork (Hastings et al. 2009). In this study, a single recombination across a short repeat was found with no deletions suggesting that it is the result of MMBIR.

Parental content of the cybrid mtDNA

Several factors could shape the gene content, length, and parental origin of the mitochondrial genome of a cybrid

plant. For example, the interactions of the recombinant mitochondria with the chloroplast (derived from *H. niger*, in this case) and the nuclear (derived from tobacco) genomes could impose constraints on the number and origin of the mitochondrial genes. Furthermore, the recombination pathways that may take place between the parental mtDNAs would influence the structure, chimerism, repeat content, and length of the cybrid mitochondrial genome. In a previous study, the mtDNA of a cybrid between tobacco and *H. niger*, Nt(+Hn) FH-4, had preferentially eliminated repeated sequences and retained a single allele (mostly from tobacco) for each mitochondrial gene (Sanchez-Puerta et al. 2015). Recently, the mtDNA of a cybrid between *Brassica napus* and *Raphanus sativa* also showed similar features (Arimura et al. 2018).

In the cybrid under study, a total of three mitochondrial genomes (two of tobacco and one of *H. niger*) were joined through two protoplast fusion events. Instead of a potentially expected giant mtDNA, the resulting cybrid mitochondria contained a genome that was 1.4–1.6 times greater than that of either parent, in agreement with observations in the other two cybrids mentioned above (Arimura et al. 2018; Sanchez-Puerta et al. 2015). In all cases, the cybrid mtDNA lost a substantial fraction of the parental mitochondrial genomes but it was larger than either parental mtDNA. The three cybrids also shared a high similarity in the parental composition of the mtDNAs. They maintained preferentially mitochondrial sequences of the same parent as the nuclear genome.

In particular, we found that most coding sequences in the cybrid Mv-1-1g were inherited from *N. tabacum*, as expected, given that the nuclear genome derived from tobacco and that the cybrid received two doses of mtDNA from *N. tabacum*. However, the cybrid mtDNA retained six genes or gene regions from *H. niger*. Three of them were maintained as single copy genes (*atp6*, *cox1*, *rps10*), and the other three were multi-copy genes with one allele derived from each parent (*nad1x5*, *nad3*, *rps12*). Similarly, in the cybrid Nt(+Hn) FH-4 previously analyzed (Sanchez-Puerta et al. 2015), 10 genes or gene pieces from *H. niger* were retained, including the genes *cox1* and *rps10*.

The presence of the chloroplast genome from *H. niger* in both cybrids could influence the retention of the mitochondrial genes. It is possible that the transferred chloroplasts may need genetically related mitochondrial sequences to be functional in a foreign nuclear background. Several reports described the close relationship of the organellar genomes and the cross-talk between organelles and the nucleus to maintain a tight cellular organization. Chloroplast and mitochondria show a high level of metabolic interdependency and a coordinated work, and this communication has been established during evolution (Leister 2005; Raghavendra and Padmasree 2003; Sabar et al. 2000; Smith and Keeling 2015). Alternatively, the retention of a few genes from *H. niger* could be the result of genetic drift.

Besides the type of parental allele retained by each cybrid mtDNA, the three cybrids whose mitochondrial genomes were sequenced up to date, maintained a single allele of each gene, either from one parent or from the other, but rarely from both. According to Sanchez-Puerta et al. (2015), this observation could be explained by a neutral mechanism that eliminates preferentially repeated sequences, or by adaptive forces given the putative deleterious effect of keeping potentially conflicting copies of coding regions derived from both parents.

A comparison of homologous and non-homologous regions inherited by the cybrids between tobacco and *H. niger* (Mv-1-1g and FH-4) showed that almost all regions that were exclusive of each parent were maintained (~89%), in stark contrast to the limited retention of those homologous sequences contributed by both parents (~48%) (Figs. 5, 6). This is also true for the cybrid between *B. napus* and *R. sativa* (Arimura et al., 2018) that retained 100% and 45% of the exclusive mitochondrial sequences from *B. napus* and *R. sativa*, respectively, while homologous sequences were kept only from *B. napus* (except for 4 kb from *R. sativa* representing 2% of the shared sequences). We hypothesize that the non-homologous parental sequences were not eliminated in the cybrid mtDNA because they could not undergo intergenomic HR. Instead, DSBs in these regions were necessarily repaired using another copy of the same parental genome. These non-homologous regions were non-coding DNA (except for uncharacterized ORFs) and would be evolving mainly by genetic drift.

Of the homologous sequences in the parental mtDNAs, the cybrid Mv-1-1g maintained only a small fraction from both parents, while the majority was retained from a single parent (mostly tobacco). The homologous sequences include coding (34%) and non-coding (66%) sequences. Under the neutral model of evolution, we expected to find an almost unbiased parental composition of the non-coding homologous regions (derived randomly from *H. niger* or *N. tabacum*) in the cybrid mitochondria. Instead, we observed that, in all three cybrids mentioned above, the vast majority of the non-coding homologous sequences derived from the same parent as the nuclear genome (tobacco or *B. napus*, depending on the cybrid). This bias raises questions regarding the importance of non-coding sequences for the mitochondrial function/homeostasis/gene expression/nuclear cytoplasmic compatibility. Alternatively, the adaptive elimination of coding regions from *H. niger* could have dragged the surrounding non-coding regions along. In fact, most of the non-coding homologous sequences are interspersed among coding regions (Fig. 5). Therefore, the putatively advantageous loss of *H. niger* genes from the cybrid mtDNA described above could be responsible for the deletion of the homologous non-coding regions of *H. niger* surrounding those genes.

Several recombination pathways could cause the elimination of genomic sequences, such as SSA (Sugawara et al. 2000; de Zamaroczy et al. 1983), recombination across short

repeats, recombination via sub-stoichiometric intermediates (Small et al. 1989), and BIR (Malkova et al. 1996). Single-strand annealing (SSA) repairs DNA breaks that are flanked by long direct DNA repeats leading to a loss of DNA between the repeats (Sugawara et al. 2000). We did not find evidence for this DSB repair pathway in the cybrid mtDNA. Recombination via sub-stoichiometric intermediates could also explain deletions in plant mitochondrial genomes (Small et al. 1989), but we could not test this hypothesis directly. We propose that BIR could be one of the mechanisms responsible for the loss of the majority of the repeated regions derived from *H. niger*. Of the nine recombination events produced by BIR, seven implied the invasion of the *H. niger* strand into a tobacco-derived DNA, resulting in the loss of a homologous segment derived from *H. niger* (Fig. 6). In addition, half-crossover BIR, caused by aberrant processing of BIR intermediates, could result in the loss of one of the two chromosomes participating in the recombination events (Deem et al. 2008; Haber and Hearn 1985). This could explain several of the single product recombination events.

We conclude that *H. niger* genes and the flanking homologous regions were eliminated from the cybrid through natural selection and that BIR was the molecular pathway responsible for the loss of such sequences. Further investigation of these mechanisms after recombination of mtDNA from different species will allow answering questions regarding their dependence on taxonomic specificities of mtDNA, predictability of recombinant mtDNA composition and possibilities of developing methodologies of precise genetic engineering based on transformation of mitochondrial genomes.

Acknowledgements This work was supported by Universidad Nacional de Cuyo (Sectyp M033), Agencia Nacional de Promoción Científica y Tecnológica (Grant No. PICT1762) to M.V.S.P and by NSF (Grant No. 1062432) to Indiana University, which supports the computer cluster.

Author contributions MVSP and LEG designed the study and analyzed the data. MZ and EZ produced the repeated cybrid and performed the DNA extraction for sequencing. MVSP, LEG, MZ and EZ wrote the manuscript.

References

- Akagi H, Shimada H, Fujimura T (1995) High-frequency inter-parental recombination between mitochondrial genomes of rice cybrids. *Curr Genet* 29:58–65
- Aleza P, Garcia-Lor A, Juárez J, Navarro L (2016) Recovery of citrus cybrid plants with diverse mitochondrial and chloroplastic genome combinations by protoplast fusion followed by in vitro shoot, root, or embryo micrografting. *Plant Cell Tissue Organ Cult* 126:205–217
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Anand RP, Lovett ST, Haber JE (2013) Break-induced DNA replication. *Cold Spring Harb Perspect Biol* 5:a010397
- Arimura SI, Yanase S, Tsutsumi N, Koizuka N (2018) The mitochondrial genome of an asymmetrically cell-fused rapeseed, *Brassica napus*, containing a radish-derived cytoplasmic male sterility-associated gene. *Genes Genet Syst* 93:143–148
- Arrieta-Montiel MP, Mackenzie SA (2011) Plant mitochondrial genomes and recombination. In: Kempken F (ed) *Plant mitochondria*. Springer, New York, pp 65–82
- Arrieta-Montiel MP, Shedje V, Davila J, Christensen AC, Mackenzie SA (2009) Diversity of the *Arabidopsis* mitochondrial genome occurs via nuclear-controlled recombination activity. *Genetics* 183:1261–1268
- Austin S, Lojkowska E, Ehlenfeldt MK, Kelman A, Helgeson JP (1988) Fertile interspecific somatic hybrids of *Solanum*: a novel source of resistance to *Erwinia* soft rot. *Phytopathology* 78:1216–1220
- Aviv D, Bleichman S, Arzee-Gonen P, Galun E (1984) Intersectional cytoplasmic hybrids in *Nicotiana*: identification of plastomes and chondriomes in *N. sylvestris* + *N. rustica* cybrids having *N. sylvestris* nuclear genomes. *Theor Appl Genet* 67:499–504
- Backert S, Börner T (2000) Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant *Chenopodium album* (L.). *Curr Genet* 37:304–314
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijbelski AD et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477
- Belliard G, Vedel F, Pelletier G (1979) Mitochondrial recombination in cytoplasmic hybrids of *Nicotiana tabacum* by protoplast fusion. *Nature* 281:401–403
- Bennardo N, Cheng A, Huang N, Stark JM (2008) Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet* 4:e1000110
- Boetzer M, Pirovano W (2012) Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579
- Bonhomme S, Budar F, Lancelin D, Small I, Defrance MC, Pelletier G (1992) Sequence and transcript analysis of the Nco2.5 Ogura-specific fragment correlated with cytoplasmic male sterility in *Brassica* cybrids. *Mol Gen Genet* 235:340–348
- Brown CR, Yang CP, Mojtahedi H, Santo GS, Masuelli R (1996) RFLP analysis of resistance to Columbia root-knot nematode derived from *Solanum bulbocastanum* in a BC2 population. *Theor Appl Genet* 92:572–576
- Carlson PS, Smith HH, Dearing RD (1972) Parasexual interspecific plant hybridization. *Proc Natl Acad Sci USA* 69:2292–2294
- Carlsson J, Leino M, Sohlberg J, Sundström JF, Glimelius K (2008) Mitochondrial regulation of flower development. *Mitochondrion* 8:74–86
- Chang HHY, Pannunzio NR, Adachi N, Lieber MR (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* 18:495–506
- Charlesworth D (2017) Origins of rice cytoplasmic male sterility genes. *Cell Res* 27:3–4
- Chase CD (2007) Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends Genet* 23:81–90
- Chase CD, Gabay-Laughnan S (2004) Cytoplasmic male sterility and fertility restoration by nuclear genes. In: Daniell H, Chase C (eds) *Molecular biology and biotechnology of plant organelles: chloroplasts and mitochondria*. Springer, Dordrecht, pp 593–621

- Chen L, Liu Y-G (2014) Male sterility and fertility restoration in crops. *Annu Rev Plant Biol* 65:579–606
- Chou J-Y, Leu J-Y (2015) The red queen in mitochondria: cyto-nuclear co-evolution, hybrid breakdown and human disease. *Front Genet* 6:187
- Davila JJ, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu Y-Z, Weigel D, Mackenzie SA (2011) Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. *BMC Biol* 9:64
- de Zamaroczy M, Faugeron-Fonty G, Bernardi G (1983) Excision sequences in the mitochondrial genome of yeast. *Gene* 21:193–202
- Deem A, Barker K, VanHulle K, Downing B, Vayl A, Malkova A (2008) Defective break-induced replication leads to half-cross-overs in *Saccharomyces cerevisiae*. *Genetics* 179:1845–1860
- Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA miniprep: version II. *Plant Mole Biol Rep* 1:19–21
- Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45:e18
- Ducos E, Touzet P, Boutry M (2001) The male sterile G cytoplasm of wild beet displays modified mitochondrial respiratory complexes. *Plant J* 26:171–180
- Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR (2015) HMMER web server: 2015 update. *Nucleic Acids Res* 43:W30–38
- Fitter JT, Thomas MR, Niu C, Rose RJ (2005) Investigation of *Nicotiana tabacum* (+) *N. suaveolens* cybrids with carpeloid stamens. *J Plant Physiol* 162:225–235
- Gleba YY, Sytnik KM (1984) Protoplast fusion—genetic engineering in higher plants. Springer, Berlin, IllusX + p 220
- Gordon D, Green P (2013) Consed: a graphical editor for next-generation sequencing. *Bioinformatics* 29:2936–2937
- Greiner S, Bock R (2013) Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* 35:354–365
- Gualberto JM, Newton KJ (2017) Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol* 68:225–252
- Gurdon C, Svab Z, Feng Y, Kumar D, Maliga P (2016) Cell-to-cell movement of mitochondria in plants. *Proc Natl Acad Sci USA* 113:3395–3400
- Haber JE, Hearn M (1985) Rad52-independent mitotic gene conversion in *Saccharomyces cerevisiae* frequently results in chromosomal loss. *Genetics* 111:7–22
- Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16(Suppl):S154–169
- Hastings PJ, Ira G, Lupski JR (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 5:e1000327
- Hua SB, Qiu M, Chan E, Zhu L, Luo Y (1997) Minimum length of sequence homology required for in vivo cloning by homologous recombination in yeast. *Plasmid* 38:91–96
- Ilcheva V, San LH, Dimitrov B, Zagorska N (2000) Morphological and cytological characteristics of somatic hybrids of *Nicotiana tabacum* L. (+) *N. megalosiphon* Heurk. et Müll. *Vitro Cell Dev Biol Plant* 36:69–73
- Iovene M, Savarese S, Cardi T, Frusciante L, Scotti N, Simon PW, Carputo D (2007) Nuclear and cytoplasmic genome composition of *Solanum bulbocastanum* (+) *S. tuberosum* somatic hybrids. *Genome* 50:443–450
- Izhar S, Zelcer A (1986) Protoplast fusion and generation of cybrids for transfer of cytoplasmic male sterility. In: Vasil IK (ed) *Cell culture and somatic cell genetics of plants*. Academic Press, New York, pp 589–599
- Kajander OA, Karhunen PJ, Holt IJ, Jacobs HT (2001) Prominent mitochondrial DNA recombination intermediates in human heart muscle. *EMBO Rep* 2:1007–1012
- Kim DH, Kim BD (2006) The organization of mitochondrial *atp6* gene region in male fertile and CMS lines of pepper (*Capsicum annuum* L.). *Curr Genet* 49:59–67
- Kofer W, Glimelius K, Bonnett HT (1991) Modifications of mitochondrial DNA cause changes in floral development in homeotic-like mutants of tobacco. *Plant Cell* 3:759–769
- Kubo T, Mikami T (2007) Organization and variation of angiosperm mitochondrial genome. *Physiol Plantarum* 129:6–13
- Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8:5–14
- Kühn, K., and Gualberto, J.M. (2012). Recombination in the stability, repair and evolution of the mitochondrial genome. In: *Advances in botanical research*, vol 63, pp 215–252
- Kushnir SG, Shlumukov LR, Pogrebnyak NJ, Berger S, Gleba Y (1987) Functional cybrid plants possessing a *Nicotiana* genome and an *Atropa* plastome. *Mol Gen Genet* 209:159–163
- Leino M, Landgren M, Glimelius K (2005) Alloplasmic effects on mitochondrial transcriptional activity and RNA turnover result in accumulated transcripts of *Arabidopsis* orfs in cytoplasmic male-sterile *Brassica napus*. *Plant J* 42:469–480
- Leister D (2005) Genomics-based dissection of the cross-talk of chloroplasts with the nucleus and mitochondria in *Arabidopsis*. *Gene* 354:110–116
- Lieber MR (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 79:181–211
- Ling F, Makishima F, Morishima N, Shibata T (1995) A nuclear mutation defective in mitochondrial recombination in yeast. *EMBO J* 14:4090–4101
- Liu J, Xu X, Deng X (2005) Intergeneric somatic hybridization and its application to crop genetic improvement. *Plant Cell Tissue Organ Cult* 82:19–44
- Llorente B, Smith CE, Symington LS (2008) Break-induced replication: what is it and what is it for? *Cell Cycle* 7:859–864
- Logan DC (2006) The mitochondrial compartment. *J Exp Bot* 57:1225–1243
- 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. *Curr Genet* 52:267–274
- Lovett ST, Hurley RL, Suter VA Jr, Aubuchon RH, Lebedeva MA (2002) Crossing over between regions of limited homology in *Escherichia coli*: RecA-dependent and RecA-independent pathways. *Genetics* 160(3):851–859
- Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–57
- Malkova A, Ivanov EL, Haber JE (1996) Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci USA* 93:7131–7136
- Malkova A, Naylor ML, Yamaguchi M, Ira G, Haber JE (2005) RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. *Mol Cell Biol* 25:933–944
- Maréchal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytol* 186(2):299–317
- McGrath JM, Williams CE, Haberlach GT, Wielgus SM, Uchytel TF, Helgeson JP (2002) Introgression and stabilization of *Erwinia* tuber soft rot resistance into potato after somatic hybridization of *Solanum tuberosum* and *S. brevidens*. *Am J Potato Res* 79:19–24

- Mehta A, Haber JE (2014) Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb Perspect Biol* 6:a016428
- Miller-Messmer M, Kuhn K, Bichara M, Le Ret M, Imbault P, Gualberto JM (2012) RecA-dependent DNA repair results in increased heteroplasmy of the *Arabidopsis* mitochondrial genome. *Plant Physiol* 159:211–226
- Morgan A, Maliga P (1987) Rapid chloroplast segregation and recombination of mitochondrial DNA in *Brassica* cybrids. *Mol Gen Genet* 209:240–246
- Nagy F, Lázár G, Menczel L, Maliga P (1983) A heteroplasmic state induced by protoplast fusion is a necessary condition for detecting rearrangements in *Nicotiana* mitochondrial DNA. *Theor Appl Genet* 66:203–207
- Okazaki M, Kazama T, Murata H, Motomura K, Toriyama K (2013) Whole mitochondrial genome sequencing and transcriptional analysis to uncover an RT102-type cytoplasmic male sterility-associated candidate gene derived from *Oryza rufipogon*. *Plant Cell Physiol* 54:1560–1568
- Orczyk W, Przetakiewicz J, Nadolska-Orczyk A (2003) Somatic hybrids of *Solanum tuberosum*—application to genetics and breeding. *Plant Cell Tissue Organ Cult* 74:1–13
- Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol* 28:87–97
- Pâques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63:349–404
- Pineau B, Mathieu C, Gérard-Hirne C, De Paepe R, Chétrit P (2005) Targeting the NAD7 subunit to mitochondria restores a functional complex I and a wild type phenotype in the *Nicotiana sylvestris* CMS II mutant lacking nad7. *J Biol Chem* 280:25994–26001
- Raghavendra AS, Padmasree K (2003) Beneficial interactions of mitochondrial metabolism with photosynthetic carbon assimilation. *Trends Plant Sci* 8:546–553
- Rand DM, Haney RA, Fry AJ (2004) Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol* 19:645–653
- Resnick MA (1976) The repair of double-strand breaks in DNA: a model involving recombination. *J Theor Biol* 59:97–106
- Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW et al (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342:1468–1473
- Rothenberg M, Hanson MR (1988) A functional mitochondrial ATP synthase proteolipid gene produced by recombination of parental genes in a petunia somatic hybrid. *Genetics* 118:155–161
- Sabar M, De Paepe R, de Kouchkovsky Y (2000) Complex I impairment, respiratory compensations, and photosynthetic decrease in nuclear and mitochondrial male sterile mutants of *Nicotiana sylvestris*. *Plant Physiol* 124:1239–1249
- Sakofsky CJ, Ayyar S, Malkova A (2012) Break-induced replication and genome stability. *Biomolecules* 2:483–504
- Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD (2008) Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. *Mol Biol Evol* 25:1762–1777
- Sanchez-Puerta MV, Zubko MK, Palmer JD (2015) Homologous recombination and retention of a single form of most genes shape the highly chimeric mitochondrial genome of a cybrid plant. *New Phytol* 206:381–396
- Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci* 3:175–180
- Scotti N, Maréchal-Drouard L, Cardi T (2004) The *rpl5-rps14* mitochondrial region: a hot spot for DNA rearrangements in *Solanum* spp. somatic hybrids. *Curr Genet* 45:378–382
- Shedge V, Davila J, Arrieta-Montiel MP, Mohammed S, Mackenzie SA (2010) Extensive rearrangement of the *Arabidopsis* mitochondrial genome elicits cellular conditions for thermotolerance. *Plant Physiol* 152:1960–1970
- Shen P, Huang HV (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* 112:441–457
- Skippington E, Barkman TJ, Rice DW, Palmer JD (2015) Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc Natl Acad Sci USA* 112:E3515–E3524
- Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol* 10:e1001241
- Small I, Suffolk R, Leaver CJ (1989) Evolution of plant mitochondrial genomes via substoichiometric intermediates. *Cell* 58:69–76
- Smith DR, Keeling PJ (2015) Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci USA* 112:10177–10184
- Sugawara N, Ira G, Haber JE (2000) DNA length dependence of the single-strand annealing pathway and the role of *Saccharomyces cerevisiae* RAD59 in double-strand break repair. *Mol Cell Biol* 20:5300–5309
- Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiyama M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol Genet Genomics* 272:603–615
- Sun Y, Zhang X, Nie Y, Guo X (2005) Production of fertile somatic hybrids of *Gossypium hirsutum* + *G. bickii* and *G. hirsutum* + *G. stockii* via protoplast fusion. *Plant Cell Tissue Organ Cult* 83:303–310
- Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* 33:25–35
- Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410:103–123
- Touzert P, Meyer EH (2014) Cytoplasmic male sterility and mitochondrial metabolism in plants. *Mitochondrion* 19:166–171
- Vedel F, Chétrit P, Mathieu C, Pelletier G, Primard C (1986) Several different mitochondrial DNA regions are involved in intergenomic recombination in *Brassica napus* cybrid plants. *Curr Genet* 11:17–24
- Warren RL, Sutton GG, Jones SJM, Holt RA (2007) Assembling millions of short DNA sequences using SSAKE. *Bioinformatics* 23:500–501
- Woodson JD, Chory J (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* 9:383–395
- Wu L, Hickson ID (2003) The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* 426:870–874
- Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, Davis CC (2013) Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PLoS Genet* 9:e1003265
- Xiang F, Xia G, Zhi D, Wang J, Nie H, Chen H (2004) Regeneration of somatic hybrids in relation to the nuclear and cytoplasmic genomes of wheat and *Setaria italica*. *Genome* 47:680–688
- Zaegel V, Guermann B, Le Ret M, Andres C, Meyer D, Erhardt M, Canaday J, Gualberto JM, Imbault P (2006) The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *Plant Cell* 18:3548–3563
- Zubko MK, Zubko EI, Patskovsky YV, Khvedynich OA, Fisahn J, Gleba YY, Schieder O (1996) Novel “homeotic” CMS patterns generated in *Nicotiana* via cybridization with *Hyoscyamus* and *Scopolia*. *J Exp Bot* 47:1101–1110