



Article

# Complete Sequence, Multichromosomal Architecture and Transcriptome Analysis of the *Solanum tuberosum* Mitochondrial Genome

Jean-Stéphane Varré <sup>1,†</sup>, Nunzio D'Agostino <sup>2,†,‡</sup> , Pascal Touzet <sup>3</sup> , Sophie Gallina <sup>3</sup>,  
Rachele Tamburino <sup>4</sup> , Concita Cantarella <sup>2</sup>, Elodie Ubrig <sup>5</sup>, Teodoro Cardi <sup>2</sup> ,  
Laurence Drouard <sup>5</sup> , José Manuel Gualberto <sup>5,\*</sup> and Nunzia Scotti <sup>4,\*</sup>

<sup>1</sup> Univ. Lille, CNRS, Centrale Lille, UMR 9189—CRISTAL—Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France; jean-stephane.varre@univ-lille.fr

<sup>2</sup> CREA Research Centre for Vegetable and Ornamental Crops, 84098 Pontecagnano Faiano, SA, Italy; nunzio.dagostino@crea.gov.it (N.D'.A.); concita.cantarella@gmail.com (C.C.); teodoro.cardi@crea.gov.it (T.C.)

<sup>3</sup> Univ. Lille, CNRS, UMR 8198—Evo-Eco-Paleo, F-59000 Lille, France; pascal.touzet@univ-lille.fr (P.T.); sophie.gallina@univ-lille.fr (S.G.)

<sup>4</sup> CNR-IBBR, National Research Council of Italy, Institute of Biosciences and BioResources, 80055 Portici, NA, Italy; rachele.tamburino@gmail.com

<sup>5</sup> Institut de Biologie Moléculaire des Plantes-CNRS, Université de Strasbourg, 67084 Strasbourg, France; elodie.ubrig@ibmp-cnrs.unistra.fr (E.U.); laurence.drouard@ibmp-cnrs.unistra.fr (L.D.)

\* Correspondence: jose.gualberto@ibmp-cnrs.unistra.fr (J.M.G.); nscotti@unina.it (N.S.)

† These authors contributed equally to this work.

‡ Present address: Department of Agricultural Sciences, University of Naples Federico II, 80055 Portici, NA, Italy.

Received: 1 August 2019; Accepted: 24 September 2019; Published: 26 September 2019



**Abstract:** Mitochondrial genomes (mitogenomes) in higher plants can induce cytoplasmic male sterility and be somehow involved in nuclear-cytoplasmic interactions affecting plant growth and agronomic performance. They are larger and more complex than in other eukaryotes, due to their recombinogenic nature. For most plants, the mitochondrial DNA (mtDNA) can be represented as a single circular chromosome, the so-called master molecule, which includes repeated sequences that recombine frequently, generating sub-genomic molecules in various proportions. Based on the relevance of the potato crop worldwide, herewith we report the complete mtDNA sequence of two *S. tuberosum* cultivars, namely Cicero and Désirée, and a comprehensive study of its expression, based on high-coverage RNA sequencing data. We found that the potato mitogenome has a multi-partite architecture, divided in at least three independent molecules that according to our data should behave as autonomous chromosomes. Inter-cultivar variability was null, while comparative analyses with other species of the Solanaceae family allowed the investigation of the evolutionary history of their mitogenomes. The RNA-seq data revealed peculiarities in transcriptional and post-transcriptional processing of mRNAs. These included co-transcription of genes with open reading frames that are probably expressed, methylation of an rRNA at a position that should impact translation efficiency and extensive RNA editing, with a high proportion of partial editing implying frequent mis-targeting by the editing machinery.

**Keywords:** mitochondria; potato; Solanaceae family; mtDNA; multichromosomal structure; repeated sequences; RNA editing; comparative genomics

## 1. Introduction

Mitochondrion is a semi-autonomous organelle that supplies cells with ATP through oxidative phosphorylation. In higher plants, the mitochondrial DNA (mtDNA) displays unique features when compared with animal and fungi counterparts. It is larger and highly variable in size, ranging between 208 kb in *Brassica hirta* [1] and 11.3 Mb in *Silene conica* [2]. The mitochondrial genome (mitogenome) also varies greatly in size across species belonging to the same family, such as in Cucurbitaceae, where the mtDNA of *Citrullus lanatus* is 379 kb [3], while in *Cucumis melo* it is 2740 kb in size [4]. However, the extreme size variability is not reflected in the gene content, which remains quite constant among species and includes less than 50 protein-coding genes, mainly components of the electron transport chain, and a few rRNA and tRNA genes, accounting for about 10% of the mitogenome [5]. Most of the remaining DNA consists of non-coding sequences commonly of unknown origin, but also acquired from nuclear, chloroplast or viral DNA by horizontal transfer [6].

The plant mitogenomes are conventionally described as circular structures, and in most species they could be mapped into a single circular chromosome often called a “master circle”. However, in vivo observation often failed to recover circular molecules [7–9]. Rather, most studies demonstrated that the mitochondrial genome structure is more complex and dynamic, being a mixture of inter-convertible linear and circular DNA molecules that result from homologous recombination events involving large, intermediate-size (ISR) and small repeated sequences scattered throughout the genome [10–12]. The relative ratio of alternative structures depends on the balance between their recombination and replication rates. Specifically, the recombination events that involve ISRs are often asymmetrical and have been linked to intra-specific mtDNA variation [13]. The single circle map of the genome is also a concept that had to be revised, because of the recent identification of mitogenomes with multiple molecules, constituted by assembled identities autonomous from each other, with no evidence that they can be inter-convertible to a single molecule by recombination [11,12,14–16]. An extreme example was found in certain *Silene* species, which can have more than 50 autonomous molecules [2], and that evolve fast by gain or loss of entire molecules [17].

Furthermore, mitochondria of many plant species also contain circular and linear plasmid DNA molecules that exist as standalone extra-chromosomal elements that can range from 0.7 to over 20 kb [10,18,19]. Multimeric forms of these molecules, likely resulting from rolling-circle replication processes have been observed [20]. The pattern of mitochondrial plasmids can be species-specific and contribute to the complexity of mitochondrial genetics. Most of them do not show sequence similarity with the main chromosome, and thus cannot be considered as sub-genomes [10,18]. Despite the high variability in size and structure, in most plant species, the mtDNA displays a slow sequence evolution rate likely due to efficient DNA repair systems, most probably by copy-correction by recombination [21,22]. Finally, recombination events may lead to the generation of new chimeric open reading frames (orfs) that, in several cases, have been associated with cytoplasmic male sterility (CMS) [22]. CMS is a maternally inherited inability to produce or shed functional pollen, an economically important trait that can be exploited in plant breeding for the production of hybrid seeds [23–25].

A particularity of plant mitochondrial gene expression is extensive RNA editing by C to U deamination, required for the correct expression of its genes [26]. In all flowering plants, RNA editing is an essential step of mitochondrial RNA maturation without which no functional mitochondria can be assembled and maintained in the cell [27]. Numerous studies on RNA editing have been carried out, but how a functional editosome is assembled in plant organelles has remained elusive [28,29]. Very recently, Oldenkott et al. [30] demonstrated that single Pentatricopeptide Repeat (PPR) proteins with a terminal DYW domain from *Physcomitrella patens* can edit their corresponding target in *Escherichia coli*. Although new sequence approaches allowed the sequencing of the mitogenomes from many plant species, the complete profile of mitochondrial editing sites has only been determined for relatively few species. Comparison between species allows good prediction of sites required for correct protein expression, but the in silico predictions are blind to editing sites that are partially edited or that result in silent substitutions.

Potato (*Solanum tuberosum* L.) is the third most important food crop in the world grown for human consumption (<http://www.fao.org/faostat/en>). It originates from Southern America and is nowadays cultivated in all continents except Antarctica, with China being the greatest producer (<http://www.fao.org/faostat/en>). The key event in potato domestication was adaptation to a long-day photoperiod, enabling global cultivation at non-equatorial latitudes [31]. Domestication also selected for enlarged tubers with a reduced content in toxic glycoalkaloids and reduced sexual fertility [31]. The potato nuclear sequence was published in 2011 [32] and since then numerous genomic analyses have been performed to characterize nuclear genome composition, to assess chromosome structure organization as well as the genetic variability of potato germplasm collections [33–38]. Furthermore, several studies have included comparisons among closely related species aimed at uncovering the evolutionary history of the Solanaceae family. Intra- and inter-specific variation in mtDNA organization was highlighted in common potato and related species [39–42], showing co-evolution patterns of the chondriome and the other cellular genomes. In this regard, studies on the mitogenome composition, structure and organization may help elucidate the genetic diversity in potato and reveal molecular mechanisms underlying nuclear-cytoplasmic interactions responsible for male fertility and expression of other agronomic relevant traits [41,43].

Herein we report the complete mtDNA sequence of two *S. tuberosum* cultivars, namely Cicero and Désirée. From PacBio RS long-reads we found that the potato mitogenome is divided in a minimum of three autonomous molecules that according to our data should behave as autonomous chromosomes, and that there is no variability between the two potato cultivars investigated. Comparative analyses with available mitogenomes for species belonging to the Solanaceae family allowed us to identify syntenic blocks and investigate the evolutionary history of its mitogenomes. Finally, from RNA sequencing data, we identified putative promoters and transcription processing sites, and characterized the extensive RNA editing pattern of potato mitochondria.

## 2. Results and Discussion

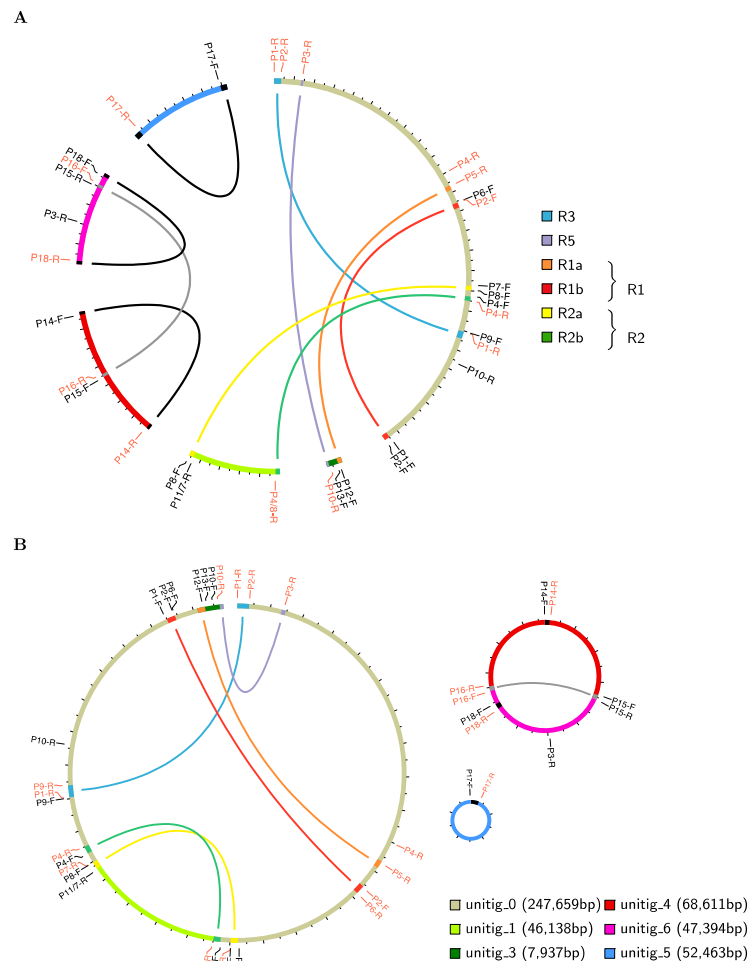
### 2.1. *S. tuberosum* Mitochondrial Genome Assembly

The *S. tuberosum* cv. Cicero was assembled into 6 linear unitigs (Table S1), with a total size of 470 kb. The *S. tuberosum* cv. Désirée was assembled into 5 linear scaffolds, with a total size of 475 kb (Table S1). The extraction of DNAs from purified and DNase treated mitochondria should avoid the inclusion of any NUMTs in the final assemblies. Despite the fact that the two assemblies originated from different data sets and by the application of different bioinformatic tools, they are almost identical in the primary sequence (99.99% identity), as well as in the multi-partite architecture. Therefore, hereinafter we will refer to a single assembly, namely that of *S. tuberosum* cv. Cicero (Figure 1A).

Genome assembly was complicated by the presence of direct or inverted repeats of large/medium size (colored blocks in Figure 1A, Table S2) responsible for the multi-partite configuration of the mitogenome. We initially found 23 pairs of repeats, 10 being larger than 1,000 bp. Several repeats were located at the ends of raw contigs (Figure 1A), indicating that they might exist as independent sub-genomes in vivo. After manual scaffolding the number of repeats larger than 100 bp was reduced to 18, ranging from 111 to 11,915 bp in length, (Figure 1, Table S2). The assembly graph provided by Canu gave us insights into putative scaffolding scenarios (Figure S1). By mapping repeats on reads in this graph, we were able to explore which rearrangements had to be considered. PCR experiments designed to verify alternative configurations (Table S3, Figure S2) allowed us to propose the best assembly, and to recover missing sequences filtered out during the assembly process.

We found that the potato mitogenome can be assembled into two autonomous circular molecules of 49,229 bp and 112,797 bp, plus a third linear sequence of 312,491 bp, for a final genome size of 474,520 bp (Figure 1B). The same organization was inferred to both Cicero and Désirée cultivars. Other conformations are possible, and alternative circular rearrangements can be described for the linear sequence (Figure S3). One possible configuration uses repeat R3 (blue repeat in Figure 1) leading to a

circular sequence of 229,571 bp by removing the end of unitig 0 (Figure S3A). A second alternative configuration uses repeat R5 (purple repeat in Figure 1) leading to a slightly different circular sequence of 296,789 bp, by removing the beginning of unitig 0 (Figure S3B). Thus, it is possible that in vivo there is co-existence of two alternative circular forms, and no linear molecules. However, the existence of the two is required, because essential genes are present in the small regions differentiating the two alternative circular forms. PCR experiments confirmed the existence of these alternative configurations.



**Figure 1.** *S. tuberosum* mitochondrial genome assembly. (A) The 6 unitigs output by HGAP for the Cicero mitogenome. Colored blocks on the unitigs show repeats. Labeled ticks give primers positions, with those in black in forward orientation and those in light red in reverse orientation. (B) The 3 contigs obtained after PCR validation. Colored blocks on the contigs show repeats. The larger contig could not be circularized.

## 2.2. Gene Content of the Potato mtDNA

The analysis of the *S. tuberosum* mtDNA sequence revealed a standard set of protein-coding genes usually found in the mitogenomes of other dicot species (Table 1, Figure 2). An *rps14*-related sequence located downstream *rpl5* is a pseudogene that contains frame-shifts resulting in in-frame stop codons. This pseudogene was previously described both in potato and Arabidopsis [44–46], and it has been shown that a functional copy of the gene has been transferred to the nucleus [47]. This transfer occurred roughly 80 million years ago, and it is surprising that the sequence of the mitochondrial pseudogene has remained virtually unchanged [48].

**Table 1.** List of transcribed protein-encoding genes and orfs. Genes are clustered according to transcription units.

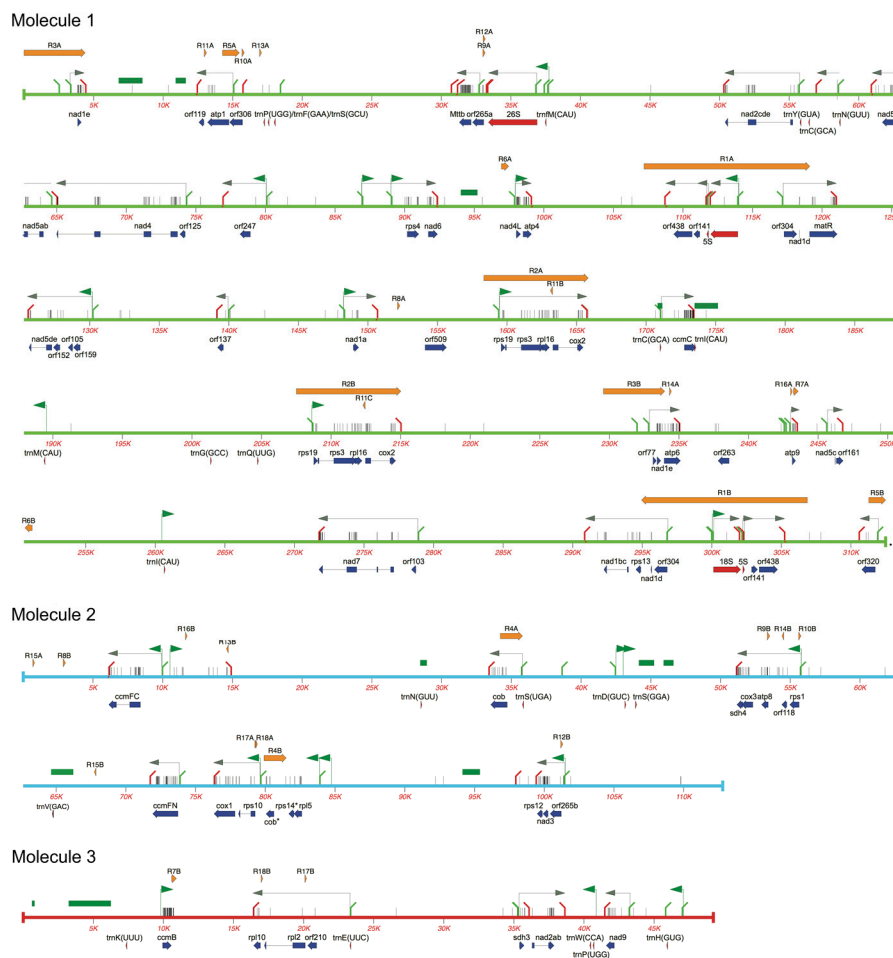
| Molecule 1 |        |        |        |  |
|------------|--------|--------|--------|--|
| Gene       | Start  | Stop   | Strand | Comments   |
| nad1e      | 3860   | 4118   | +      |  |
| orf119     | 12570  | 12929  | –      |  |
| atp1       | 13197  | 14732  | –      |  |
| Mttb       | 31292  | 32122  | –      |  |
| orf265a    | 32244  | 33041  | –      | N-term atp8 (20 codons), ELF-domain (pfam03317).<br>Co-transcribed with MttB   |
| 26S        | 33385  | 36879  | –      |  |
| nad2cde    | 50394  | 50581  | –      |  |
|            | 52050  | 52622  | –      |  |
|            | 55094  | 55254  | –      |  |
| nad5ab     | 61695  | 62910  | –      |  |
|            | 63755  | 63984  | –      |  |
|            | 64989  | 65077  | –      |  |
| nad4       | 67705  | 68127  | –      |  |
|            | 71259  | 71773  | –      |  |
|            | 73185  | 73645  | –      |  |
| orf125     | 73849  | 74226  | –      |  |
| orf247     | 78169  | 78912  | –      |  |
| rps4       | 90192  | 91013  | +      | No evidence of TAG created by editing. Possible<br>non-canonical initiation at GTG codon   |
| nad6       | 91707  | 92360  | +      | Transcript is processed upstream of stop codon   |
| nad4L      | 98027  | 98329  | +      |  |
| atp4       | 98518  | 99114  | +      |  |
| orf438     | 109344 | 110660 | –      | RdRp-like  |
| orf141     | 110784 | 111209 | –      |  |
| 5S         | 111720 | 111838 | –      |  |
| 18S        | 112001 | 113946 | –      |  |
| orf304     | 117275 | 118189 | +      | N-term pfam12725. C-term has 59% identity to<br>hypothetical protein RirG_027070   |
| nad1d      | 118391 | 118449 | +      |  |
| matR       | 119111 | 121087 | +      |  |
| nad5de     | 125647 | 125793 | –      |  |
|            | 126889 | 127283 | –      |  |
| orf152     | 127397 | 127855 | –      |  |
| orf105     | 128459 | 128776 | –      |  |
| orf159     | 128839 | 129318 | –      | RdRp-like  |
| orf137     | 139197 | 139610 | –      | CMS-associated protein   |
| nad1a      | 148955 | 149339 | +      |  |
| rps19      | 159585 | 159869 | +      |  |
| rps3       | 159883 | 159956 | +      |  |
|            | 161026 | 162643 | +      |  |
| rpl16      | 162534 | 163049 | +      | Editing site 162570 (96%) creates internal stop codon.<br>Editing is conserved in Arabidopsis. It implies that<br>there is no re-initiation of translation inside rps3.<br>Possible initiation at GTG codon. |
| cox2       | 163299 | 163680 | +      |  |
|            | 165066 | 165466 | +      |  |
| ccmC       | 172753 | 173583 | +      | ORF overlaps with tRNA and transcript is processed<br>at the tRNA 5', without stop codon.  |
| rps19      | 208754 | 209038 | +      |  |
| rps3       | 209052 | 209125 | +      |  |
|            | 210195 | 211812 | +      |  |
| rpl16      | 211703 | 212218 | +      |  |
|            | 212468 | 212849 | +      |  |
| cox2       | 214235 | 214635 | +      |  |

Table 1. Cont.

| Molecule 1 |        |        |        |   |
|------------|--------|--------|--------|---|
| Gene       | Start  | Stop   | Strand | Comments  |
| orf77      | 233142 | 233375 | +      |   |
| nad1e      | 233431 | 233689 | +      |   |
| atp6       | 233940 | 235106 | +      | Editing site 235065 creates stop codon making orf 13 aa shorter, with the same C-term as Arabidopsis atp6.  |
| atp9       | 243146 | 243379 | +      | Editing site 243368 creates stop codon making protein 3 aa shorter. With editing it is the same C-term as Arabidopsis atp9.   |
| nad5c      | 246225 | 246246 | +      |   |
| orf161     | 246314 | 246799 | +      |   |
|            | 271765 | 272026 | -      |   |
|            | 273770 | 274477 | -      |   |
| nad7       | 275936 | 276004 | -      |   |
|            | 276920 | 277062 | -      |   |
| orf103     | 278419 | 278730 | -      |   |
| nad1bc     | 292250 | 292441 | -      |   |
|            | 293925 | 294007 | -      |   |
| rps13      | 294547 | 294897 | -      |   |
| nad1d      | 295623 | 295681 | -      |   |
| orf304     | 295883 | 296797 | -      | hypothetical protein RirG ( <i>Rhizophagus irregularis</i> )  |
| 18S        | 300126 | 302071 | +      |   |
| 5S         | 302234 | 302352 | +      |   |
| orf141     | 302863 | 303288 | +      | Similarities to region 3' UTR of orf247   |
| orf438     | 303412 | 304728 | +      | RdRp-like   |
| orf320     | 310786 | 311748 | -      | Chimeric orf: 5' of atp1, 3' region upstream nad5c. Promoter of atp1 that is present in repeat R5   |
| Molecule 2 |        |        |        |   |
| Gene       | Start  | Stop   | Strand | Comments  |
| ccmFC      | 6136   | 6685   | -      |   |
|            | 7635   | 8401   | -      |   |
| Cob        | 33524  | 34705  | -      |   |
| sdh4       | 51176  | 51589  | -      | Overlaps cox3. Real ATG might be at codon 24. The transcript is processed about 8 codons before stop codon. Internal stop codon created by partial editing (26%) at codon 93. |
| cox3       | 51517  | 52314  | -      |   |
| atp8       | 52947  | 53417  | -      |   |
| orf118     | 54391  | 54747  | -      | Chimeric orf: C-term is from atp6   |
| rps1       | 54964  | 55635  | -      |   |
| ccmFN      | 71926  | 73737  | -      |   |
| cox1       | 76329  | 77825  | -      | Initiation codon created by editing   |
|            | 78075  | 78187  | -      | Stop codon created by editing (78107)   |
| rps10      | 78963  | 79212  | -      | Initiation codon created by editing (79211).  |
| rps14 *    | 81682  | 82051  | -      | Pseudo-gene   |
| rp15       | 82053  | 82613  | -      |   |
| rps12      | 99496  | 99867  | -      |   |
| nad3       | 99916  | 100272 | -      |   |
| orf265b    | 100423 | 101220 | -      | N-term atp8, ELF-domain (pfam03317)   |

Table 1. Cont.

| Molecule 3 |       |       |        |  |
|------------|-------|-------|--------|--|
| Gene       | Start | Stop  | Strand | Comments   |
| ccmB       | 9945  | 10565 | +      | Editing site 10238 creates stop codon in about 50% of the transcripts, in the middle of the ORF. |
| rpl10      | 16440 | 16919 | -      |  |
| rpl2       | 17204 | 17320 | -      |  |
| orf210     | 20294 | 20926 | -      |  |
| sdh3       | 35400 | 35726 | +      |  |
| nad2ab     | 36302 | 36453 | +      |  |
| nad9       | 41589 | 42161 | -      |  |



**Figure 2.** Organization of the potato mtDNA. Molecules 1, 2 and 3 are represented in green, light blue and red respectively. Gene sequences are shown below the sequence line, with protein genes in dark blue and rRNA and tRNA genes in red. Repeated sequences larger than 100 bp are shown above the sequence as orange arrows. Green bars indicate sequences of plastidial origin. Bent green and red lines indicate 5' and 3' transcript boundaries, respectively. Grey horizontal arrows represent major transcripts. Green horizontal arrows indicate consensus promoters found upstream of transcripts. Thin vertical lines indicate editing sites.

The *rpl5-ψrps14* sequences were located upstream of a truncated copy of *cob* (i.e., *ψcob*), contained in the R4B repeat, and *rps10-cox1* genes (Figure 2). An alternative arrangement with *rpl5-ψrps14* in

front of *cob* was reported by [45], and its occurrence in common potato and other species was later investigated [42,49]. Both arrangements, theoretically possible in *S. tuberosum* due to recombination between the R4 repeats were confirmed by PCR experiments and analysis of long reads spanning the R4 sequence.

As with all other higher plant mitogenomes, that of the potato also codes for the ribosomal RNAs 5S, 18S and 26S. A limited set of 23 tRNA genes, corresponding to 15 amino acids, was found in the genome (Table 2). Based on sequence similarity it can be inferred that nine are of plastidial origin, from the promiscuous import and insertion in the mtDNA of plastidial genomic sequences. A comprehensive two-dimensional polyacrylamide gel study of potato mitochondrial tRNAs showed that six of them could be detected in mitochondria [50]. However, sequences for trnC(GCA), trnV(GAC) and trnI(CAU) of plastidial origin are present in the genome (Table 2). By northern blot hybridization, we confirmed for the first two that the corresponding tRNAs cannot be detected in purified mitochondrial RNA and are apparently pseudogenes (Figure S4). The missing tRNA species are likely imported from the cytosol into mitochondria [50].

**Table 2.** List of tRNA genes. Plastid-like tRNA genes are tagged by '\*'. Expression is according to [46]. Absence of expression for plastid-like Cys, and Val tRNA genes has been checked by northern blot (Figure S3). nd = not determined.

| tRNA Gene         | Start  | Stop   | Strand | Editing NGS | Expression |
|-------------------|--------|--------|--------|-------------|------------|
| <b>Molecule 1</b> |        |        |        |             |            |
| trnP(UGG)         | 17220  | 17294  | -      |             | +          |
| trnF(GAA)         | 17545  | 17618  | -      | +           | +          |
| trnS(GCU)         | 17981  | 18068  | -      |             | +          |
| trnMf(CAU)        | 37458  | 37531  | -      |             | +          |
| trnY(GUA)         | 55772  | 55854  | -      |             | +          |
| trnN(GUU) *       | 56390  | 56461  | -      |             | +          |
| trnC(GCA)         | 58606  | 58676  | -      | +           | +          |
| trnC(GCA) *       | 171005 | 171076 | +      |             | -          |
| trnI(CAU) *       | 173488 | 173561 | +      |             | nd         |
| trnMe(CAU) *      | 189363 | 189435 | -      |             | +          |
| trnG(GCC)         | 201326 | 201397 | +      |             | +          |
| trnQ(UUG)         | 204702 | 204773 | +      |             | +          |
| trnI(CAU)         | 260636 | 260709 | +      |             | +          |
| <b>Molecule 2</b> |        |        |        |             |            |
| trnN(GUU) *       | 28511  | 28582  | +      |             | +          |
| trnS(UGA)         | 35798  | 35884  | -      |             | +          |
| trnD(GUC) *       | 43141  | 43214  | +      |             | +          |
| trnS(GGA) *       | 43901  | 43987  | +      |             | +          |
| trnV(GAC) *       | 64733  | 64804  | -      |             | -          |
| <b>Molecule 3</b> |        |        |        |             |            |
| trnK(UUU)         | 7326   | 7398   | -      |             | +          |
| trnE(UUC)         | 23306  | 23377  | -      |             | +          |
| trnW(CCA) *       | 40411  | 40484  | -      |             | +          |
| trnP(UGG)         | 40642  | 40715  | -      |             | +          |
| trnH(GUG) *       | 45897  | 45971  | -      |             | +          |

### 2.3. Transcriptome of the Potato mtDNA

Many orfs larger than 100 codons were found in the mitogenome. Among these, several are closely associated with known mitochondrial genes that could be potentially co-transcribed and expressed. It is also known that extensive RNA editing alters the coding sequences of most mitochondrial genes, including the creation of initiation and stop codons. Thus, a comprehensive study of the mtDNA genes expression requires transcriptomic studies that gives information on the sequences that are transcribed, potential promoters, transcript processing sites and editing. We have therefore sequenced three independent Illumina RNA-seq libraries (2 × 100 bp paired-end). Respectively, 61, 29 and



39 million reads could be aligned to the potato mtDNA. The very high coverage of the transcribed sequences allowed us to identify with precision the boundaries of most expressed transcripts and all editing sites, including those that are edited with low efficiency and those found in non-coding sequences that are transcribed at background level.

Based on the RNA-seq data, we were able to define 33 transcription units containing protein-coding genes and orfs (Table 1). Many of them contained co-transcribed genes, such as the one that includes co-transcribed genes *rps1*, *atp8*, *cox3* and *sdh4*, while others are mono-cistronic, like in the case of the transcripts of genes *cob*, *atp6* and *atp9*. Quite surprisingly, a highly expressed transcript (coordinates 56998–58596 in the (–) strand of molecule 1) does not contain any putative coding sequence. An analysis of its sequence did not reveal any particular putative secondary structure, but this long non-coding RNA (lncRNA) might assume regulatory functions in mtDNA maintenance or expression, as described for the mitochondria of several other non-plant systems [51].

Several unidentified orfs are transcribed. Among them, a few are most likely expressed, because their retention in the transcription unit does not seem fortuitous. That is the case of *orf247* and *orf137*, which are individually transcribed in their own specific transcripts (Figure S5A). An ortholog of *orf137* was previously described as potentially being involved in the CMS of chili pepper [52]. Other transcribed orfs deserve further study, as is the case of *orf265b*, which is co-transcribed with *nad3-rps12* genes (Figure S5B) and contains a characteristic ELF-domain (Pfam accession PF03317). Several of the transcribed orfs have sequence similarities to RNA-dependent RNA polymerases (RdRp) and might be of retroviral origin.

Based on RNA-seq data, it was possible to define RNA boundaries with relatively good precision. Genome annotation indicates the borders of the regions that are covered, and the true RNA boundaries should be a few additional nucleotides upstream and downstream of 5' and 3' ends, respectively. In the mitochondria of dicots, poorly conserved core promoter sequences have been described, with primary transcripts initiating a few nucleotides downstream of a CRTA conserved element, or of TATA-like sequences, such as TATTA [53–55]. We looked for putative promoters upstream of the 5' ends of transcripts. Eleven putative promoters were so identified, containing CRTA or TATATAA core elements (Table S4). The other 5' transcript ends that were identified are probably processing sites.

The presence of RNA structures that could be involved in transcript processing and or stability was also investigated. It is known that tRNA sequences and tRNA-like RNA structures (t-elements) that are processed by RNases P and Z are often used as processing sites for protein-coding transcripts. In potato mitochondria we found that processing of five tRNAs by RNase Z is used as a signal to define the 5'-ends of as many transcripts (rRNA 26S precursor, *nad2cde*, the lncRNA, *cob*, and *ccmC*) (Table 3). A particular case is the 3' processing of the *ccmC* transcript. A tRNACys gene of plastidial origin overlaps the C-terminus of the *ccmC* orf. The processing of this tRNA sequence by RNase P results in a truncated transcript of 92 nucleotides with no stop codon. The processing of this tRNA suggests that it is a functional tRNA potentially involved in translation. However, as described above, it has not been detected in the total population of potato mitochondrial tRNAs [50]. Northern blot analysis (Figure S4) shows a very weak signal in mitochondria as compared with chloroplasts. This signal is likely due to plastid contamination, although we cannot exclude it corresponding to a very weak mitochondrial expression, in agreement with *ccmC* processing. In Arabidopsis, processing of the *ccmC* transcript also results in a truncated orf, but in Arabidopsis the tRNACys sequence has evolved to become a pseudogene that just acts as a t-element for RNA processing [56]. Thus, it is likely that the plastid-like tRNACys of potato mitochondria is dispensable for mitochondrial translation and has been retained just as an RNA processing signal. Surprisingly, apart from tRNACys involved in *ccmC* 3' processing, none of the other events of processing involving a tRNA sequence in potato is conserved in Arabidopsis. The reciprocal is also true, and none of the tRNA and t-elements identified as signals for processing of Arabidopsis transcripts are conserved in potato (Table 3). Thus, the evolution of transcription units and of the signals required for their processing is a very rapid process. This is in striking contrast to the very low synonymous substitution rates of the gene coding sequences [57].

It has been shown that the 3'-end of several plant mitochondrial transcripts can have stable double or single stem-loop structures, probably required for their stability or processing by RNase Z [56]. Analysis of predicted RNA structures at the transcripts 3' ends revealed eight of such putative structural elements (Table 3). These elements are also not conserved between potato and Arabidopsis mitochondria.

**Table 3.** List of stem-loops and T-elements.

| <i>Arabidopsis thaliana</i> |              |                              |                   |                        |
|-----------------------------|--------------|------------------------------|-------------------|------------------------|
| Gene                        | mRNA End     | Kind of Secondary Structure  | Putative Nuclease | Conservation in Potato |
| ccmFC                       | 5'           | trnG                         | RNaseZ            | No                     |
| rps3                        | 5'           | trnK                         | RNaseZ            | No                     |
| rps4                        | 5'           | t-element                    | RNaseZ            | No                     |
| ccmFN1                      | 5'           | t-element                    | RNaseZ            | No                     |
| cox1                        | 5'           | t-element                    | RNaseZ            | No                     |
| rpl5                        | 5'           | Acceptor stem-like stem-loop | RNaseZ            | No                     |
| rpl5                        | 5'           | Acceptor stem-like stem-loop | RNaseZ            | No                     |
| atp6-2                      | 5'           | Acceptor stem-like stem-loop | RNaseZ            | No                     |
| nad7                        | 5'           | Acceptor stem-like stem-loop | RNaseP            | No                     |
| atp6-1                      | 3'           | trnS                         | RNaseP            | No                     |
| atp6-2                      | 3'           | trnS                         | RNaseP            | No                     |
| atp9                        | 3'           | Double stem-loop             | RNaseZ            | No                     |
| nad1e                       | 3'           | Double stem-loop             | RNaseZ            | No                     |
| cox2                        | 3'           | Stem-loop                    | RNaseZ            | No                     |
| ccmC                        | 3'           | t-element                    | RNaseP            | trnI                   |
| nad6                        | 3'           | t-element                    | RNaseP            | Yes                    |
| <i>Solanum tuberosum</i>    |              |                              |                   |                        |
| rrrn26S                     | 5' precursor | trnfM                        | RNaseZ            |                        |
| nad2cde                     | 5'           | trnY                         | RNaseZ            |                        |
| non-coding RNA              | 5'           | trnC                         | RNaseZ            |                        |
| Cob                         | 5'           | trnS                         | RNaseZ            |                        |
| ccmC                        | 5'           | trnC                         | RNaseZ            |                        |
| ccmC                        | 3'           | trnI                         | RNaseP            |                        |
| atp1                        | 3'           | Double stem-loop             | RNaseZ            | 12397–12441            |
| mittB                       | 3'           | Stem-loop                    | RNaseZ            | 31111–31148            |
| nad5ab                      | 3'           | Double stem-loop             | RNaseZ            | 60956–60986            |
| orf247                      | 3'           | Double stem-loop             | RNaseZ            | 76895–76938            |
| nad6                        | 3'           | t-element                    | RNaseZ            | 92309–92353            |
| atp4                        | 3'           | Stem-loop                    | RNaseZ            | 99142–99166            |
| nad1a                       | 3'           | Stem-loop                    | RNaseZ            | 150600–150637          |
| orf438                      | 3'           | Stem-loop                    | RNaseZ            | 305247–305291          |

#### 2.4. RNA Editing

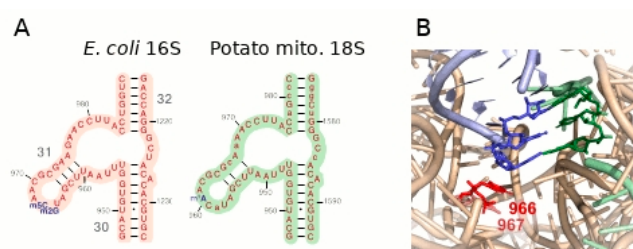
To identify editing sites, a conservative approach was used by only counting sites (C-to-T or G-to-A differences with respect to the genomic sequence, depending on the transcripts orientations) identified in at least 10% of the reads and in at least two of the three RNA-seq libraries. These were validated by visual inspection of the mapped reads. A total of 799 RNA editing sites were thus identified, a much larger number than that found in other flowering plants. That is because the very high coverage enabled the identification of many partially edited sites (Table S5). Of these, 510 (64%) were edited with an efficiency above 95%, while 149 (18%) with an efficiency below 25%. Six hundred and seventy-five (84%) of the editing sites fell into gene coding sequences (Table S5). Editing sites found in intragenic regions and in 5'- and 3'-UTRs were, in their majority, edited with low efficiency.

A possible explanation is that these editing sites are off-targets poorly recognized by PPR proteins specific for other more essential editing sites in gene coding sequences.

While most editing sites result in changes in the identity of the coded amino acids, in several cases, editing is also responsible for the creation of initiation and/or termination codons. Thus, in potato, editing creates the initiation codons of *cox1* and *rps10*, although in the latter case it was shown that translation initiates at a genomic-encoded AUG and not at the conserved AUG codon created by RNA editing [58–60]. The stop codons of *rps10*, *atp6* and *atp9* are also created by editing, shortening the coding sequences by 10, 13 and 3 codons, respectively, as compared with the genomic orfs [58,61]. As a consequence, the N-terminal sequence of COX3 and the C-terminal sequences of ATP6 and ATP9 are the same as the corresponding Arabidopsis proteins. Surprisingly, an editing site creates a stop codon in the middle of the *ccmB*. This site was observed in 48% of the mapped reads, suggesting that about 50% of *ccmB* transcripts cannot be translated into functional proteins. A premature stop at codon 96 was also found in 26% of the reads of *sdh4*. These two cases might be additional examples of “sloppy” recognition of editing sites by PPR proteins. An interesting result concerned the *rpl16* transcript. In all higher plant mitogenomes already sequenced the *rpl16* sequence overlaps the 3'-end of *rps3* [62]. The overlapping of *rps3* and *rpl16* was also observed in the mtDNA of the liverwort *Marchantia polymorpha* [63]. It was therefore accepted that translation of *rpl16* initiated internal to the *rps3* sequence, in a different frame. However, in potato, an editing site (in 96% of the reads), that in *rps3* transforms an UCA serine codon into a UUA leucine codon, creates an internal UAG stop codon in the *rpl16* orf. This editing event was also found in Arabidopsis [64], but had not been discussed. As proposed for *Marchantia*, translation of *rpl16* might initiate at the level of the valine GUG codon at position 28 of the orf, which perfectly aligns with the N-term of bacterial Rpl16 [63]. Similarly, translation of *rps4* possibly initiates at the level of a GUG codon, because the initiation codon predicted in other species to be created by editing is not so, according to our experimental data. The protein would extend 84 codons upstream the first AUG and better align with *Rps4* from other species. Two editing sites found upstream of the genomic AUG are then internal to the gene sequence and required to code for conserved amino acids.

Editing also affects non-coding sequences. Thus, our data confirmed the editing of tRNAs trnF(GAA) and trnC(GCA) [65,66]. That we could detect such editing events in our libraries of 100 bp confirms that editing of tRNAs occurs at the level of the precursor transcripts. Intron sequences are also important targets of the editing machinery. There are 24 introns found in 10 potato mitochondrial genes, and as in all flowering plants these introns are members of the group II ribozyme family. As so, they share characteristic structural domains, in particular, stems V and VI at the intron 3' border. In eight of the introns, we found editing events required for proper base-pairing of stems V and/or VI (Figure S6). Five of these were already observed and described in wheat mitochondria [67]. The relative important number of introns that require editing for proper folding suggests that timely expression of many genes is controlled by editing factors (PPR proteins) that are indirectly needed for intron splicing.

Finally, in 55% of the reads a base in the 18S rRNA sequence is a T, while an A is found in the genomic sequence (Figure S7). Such mismatch is diagnostic of a mis-incorporation during cDNA synthesis in front of a m<sup>1</sup>A nucleotide [68]. This methylated residue is found in the loop of a conserved stem of the rRNA. In the 16S rRNA of *E. coli*, there are two methylated nucleotides in this loop (Figure 3A), in the close vicinity of the anticodon of the tRNA at the P site, with one of the methylated bases stacking with the nucleotide at the wobble position of the anticodon (Figure 3B). The structure of the plant mitochondrial ribosome has still not been determined with sufficient resolution, but it is reasonable to speculate that methylation of this adenosine of the plant mitochondrial 18S rRNA is important for proper tRNA positioning and efficient translation.



**Figure 3.** Methylation of the 18S rRNA. RNA-seq data revealed in about half of the reads an A-to-T mismatch at position 960 of 18S rRNA (position 968 in domain 31 of *E. coli* 16S rRNA). Such mismatch is diagnostic of mis-incorporation during cDNA synthesis because of  $m^1A$  methylation. (A) Comparison of the corresponding domains of the *E. coli* 16S rRNA (which is methylated at positions 966 and 967,  $m^2G$  and  $m^5C$  respectively) and of the potato 18S rRNA. Sequence differences are in lower case. (B) In the 3D structure of the bacterial ribosome, these nucleotides are close to the anticodon of the tRNA at position P (dark blue nucleotides in the tRNA structure pairing with the mRNA), with base 966 stacking with the first anticodon nucleotide.

### 2.5. Comparative Analysis of Mitochondrial Genomes among Solanaceae

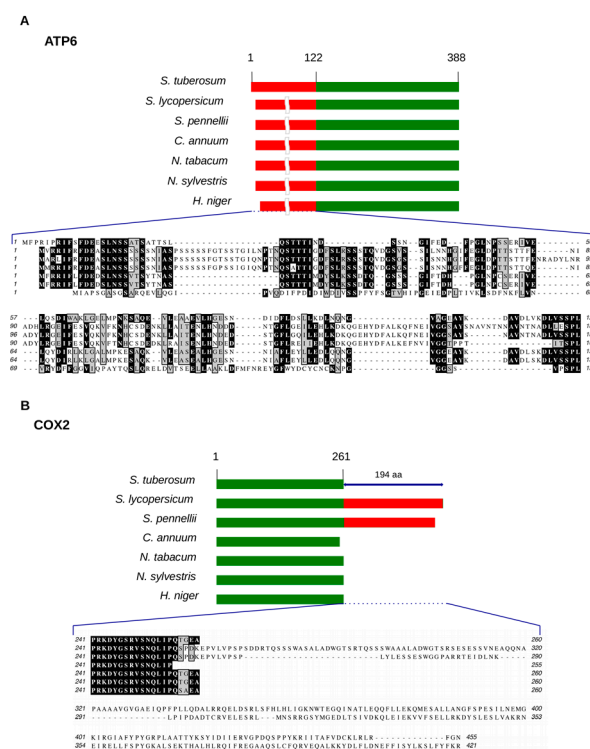
The *S. tuberosum* mitochondrial genome was compared with those of other Solanaceae species available in GenBank (see Methods). Solanaceae mtDNA varies in size from 423.60 kb (*S. pennellii*) to 511.53 kb (*C. annuum*) and has similar gene content (Table 4). Following the comparison of protein-coding genes between Solanaceae species and other Angiosperms [69], we found that several genes encoding ribosomal proteins (e.g., *rpl6*, *rps2*, *rps7*, *rps8* and *rps11*) were missing in all Solanaceae species under investigation, while *rps14* is a pseudogene in all species. Similarly, *rpl6* and *rps8* are missing in rice and *A. thaliana* mitogenomes [70,71]. Likewise, the presence of a *rps14* pseudogene has been previously described in several plant mitogenomes [44,72–77], suggesting that retention of ribosomal protein genes in the mitochondrial genome is not strictly necessary for organelle function [71]. The variability in ribosomal genes composition in Angiosperms is probably due to horizontal gene transfer [78] and led to several genome outcomes [79,80]. As already observed in rapeseed and maize mitogenomes [81,82], cultivated potato has an additional copy of *cox2*, *nad1e*, *rps3*, *rps19* and *rpl16*; the duplication of these latter two genes is shared with wild and cultivated tomato species. The absence of these duplications in *S. commersonii* is probably due to an incomplete assembly and annotation of its mitogenome. As shown in Table 4, we also surveyed the transcribed orfs identified in potato with the purpose of distinguishing highly conserved and functional ones. Three out of 18 orfs were present in all analyzed Solanaceae mitogenomes, six were common to species of the *Solanum* genus, six were present in all species except *H. niger*, one was potato-specific (wild and cultivated species), and two were specific of *S. tuberosum* (*orf125* and *orf137*). In silico analyses demonstrated that three orfs, namely *orf118*, *orf306* and *orf320*, code for chimeric proteins. In particular, ORF118 includes the C-terminus of ATP6, whereas ORF306 and ORF320 carry the N-terminus of RPS1 and ATP1, respectively. Furthermore, all chimeric ORFs have predicted transmembrane helices, a typical feature of CMS-associated proteins. Transmembrane helices were also predicted in other non-chimeric ORFs (e.g., ORF125, ORF137, ORF210, ORF247, etc.) that could potentially have a role in CMS as already observed in CMS-rice [83].

**Table 4.** Mitochondrial genes encoding proteins and transcribed open reading frames (*orfs*) among Solanaceae species available in GenBank. The symbol ● indicates the presence of the gene; ψ, a pseudogene; -, gene loss.

| Gene           | <i>S. tuberosum</i> | <i>S. commersonii</i> | <i>S. lycopersicum</i> | <i>S. pennellii</i> | <i>C. annuum</i> | <i>N. tabacum</i> | <i>N. sylvestris</i> | <i>H. niger</i> |
|----------------|---------------------|-----------------------|------------------------|---------------------|------------------|-------------------|----------------------|-----------------|
| <i>atp1</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>atp4</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>atp6</i>    | ●                   | ● <sup>a</sup>        | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>atp8</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>atp9</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>ccmB</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>ccmC</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>ccmFc</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>ccmFN</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>cob</i>     | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>cox1</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>cox2</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>cox3</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>matR</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>mttB</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad1</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad2</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad3</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad4</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad4L</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad5</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad6</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad7</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>nad9</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rpl2</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rpl5</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rpl10</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rpl16</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rps1</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rps3</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rps4</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rps10</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rps12</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rps13</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>rps14</i>   | ψ                   | ψ                     | ψ                      | ψ                   | ψ                | ψ                 | ψ                    | ψ               |
| <i>rps19</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>sdh3</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>sdh4</i>    | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>orf77</i>   | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | -               |
| <i>orf103</i>  | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>orf105</i>  | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | -               |
| <i>orf118</i>  | ●                   | ●                     | ψ                      | ψ                   | -                | -                 | -                    | -               |
| <i>orf119</i>  | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>orf125</i>  | ●                   | -                     | -                      | -                   | -                | -                 | -                    | -               |
| <i>orf137</i>  | ●                   | -                     | -                      | -                   | -                | -                 | -                    | -               |
| <i>orf141</i>  | ●                   | ●                     | ●                      | ●                   | -                | -                 | -                    | -               |
| <i>orf152</i>  | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | -               |
| <i>orf159</i>  | ●                   | ●                     | ●                      | ●                   | ψ                | -                 | -                    | -               |
| <i>orf161</i>  | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ψ               |
| <i>orf210</i>  | ●                   | ●                     | ●                      | ●                   | -                | -                 | -                    | -               |
| <i>orf247</i>  | ●                   | ●                     | ●                      | ●                   | -                | -                 | -                    | -               |
| <i>orf265a</i> | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ψ               |
| <i>orf265b</i> | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ψ               |
| <i>orf304</i>  | ●                   | ●                     | ●                      | ●                   | ψ                | -                 | -                    | -               |
| <i>orf320</i>  | ●                   | ●                     | ●                      | ●                   | ●                | ●                 | ●                    | ●               |
| <i>orf438</i>  | ●                   | ●                     | ●                      | ●                   | -                | -                 | -                    | -               |

<sup>a</sup> = *atp6* sequences available in GenBank (MF989960.1 and MF989961.1 accessions) were incomplete.

Potato mitochondrial proteins generally showed a highly conserved primary structure in comparison with other Solanaceae, with few exceptions (Figure 4). ATP6, a subunit of the  $F_0$  component of ATP synthase (respiratory complex V), differs in its N-terminal sequence, which corresponds to the leader peptide, cleaved by a metalloprotease after its translocation into the mitochondrial inner membrane [84–86]. This variability is because of DNA insertions and deletions, and in agreement with a faster evolution of functionally less important parts of a protein compared with those directly involved in the assembly of the ATP synthase complex [87,88]. Different N-terminal ATP6 sequences have been found in different species and even in the same species, as is the case with the two copies of ATP6 coded by the Arabidopsis mtDNA. Potato COX2, a subunit of the proton-pumping cytochrome c oxidase (respiratory complex IV), is shorter by 194 aa at the C-terminus, as compared to the predicted COX2 of *S. lycopersicum* and *S. pennellii*, and this feature is shared with *C. annuum*, *Nicotiana* species and *H. niger* (Figure 4). It is possible that in the other Solanaceae this C-terminal extension is post-translationally processed or is not translated due to the creation of a stop codon by editing. We cannot exclude, however, that variability in the primary structure of those proteins might be due to errors in the assembly process or mistakes in the annotation of the sequences.

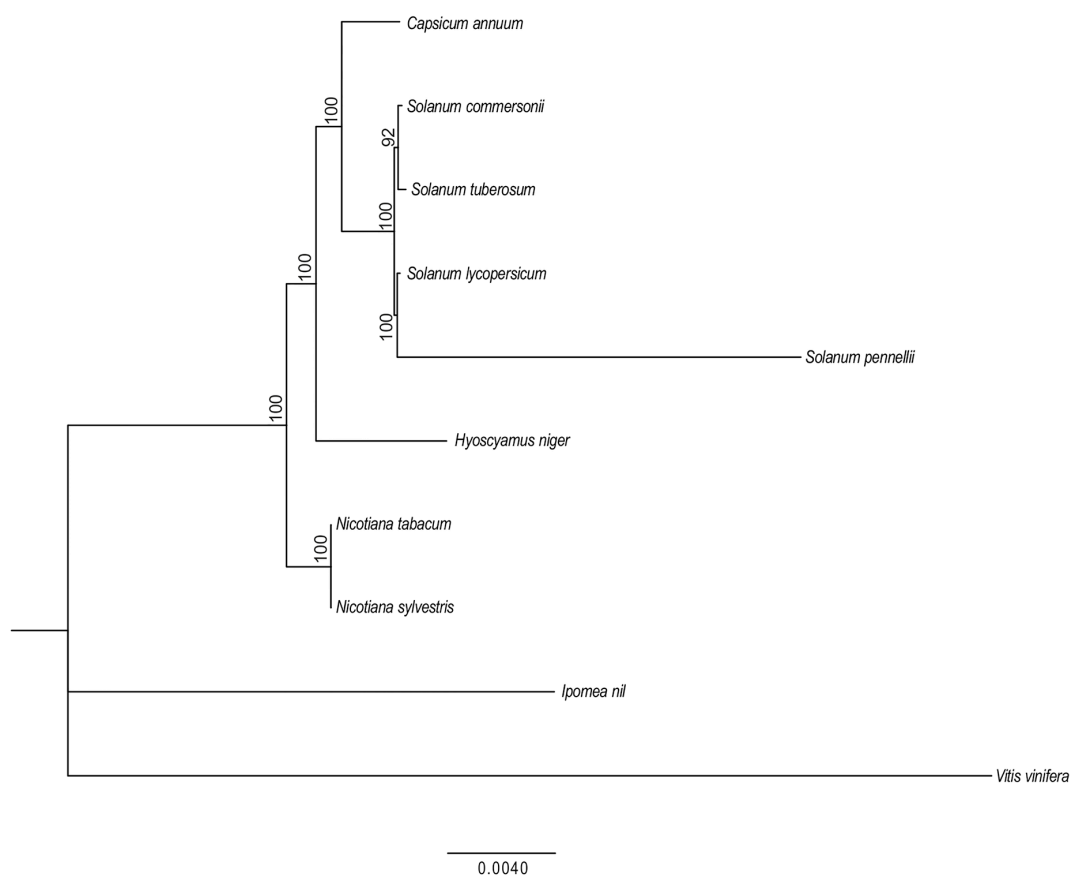


**Figure 4.** Representative examples of variable mitochondrial protein sequences in Solanaceae species. The primary structures of *S. tuberosum* ATP6 (A) and COX2 (B) proteins were compared with those of other Solanaceae species available in GenBank. (A) Red and green bars indicate non-conserved or conserved aa sequences in ATP6, respectively. (B) Cultivated (*S. lycopersicum*) and wild (*S. pennellii*) tomato have additional 194 aa at the C-terminus of the protein compared with potato and other Solanaceae species. The numbers above red and green bars indicate the start and end positions of potato mitochondrial proteins.

To investigate mitochondrial genome rearrangements across cultivated potato and the other Solanaceae species, we identified syntenic blocks  $\geq 5$  kb and sequence identity  $\geq 95\%$ . *S. tuberosum* mtDNA shared the highest number (up to 29) of syntenic blocks with cultivated and wild tomato species (Table S6), followed by *S. commersonii*, *C. annuum* and *Nicotiana* species. Comparison with *H. niger* resulted in only two syntenic blocks. Furthermore, the syntenic blocks identified within species

of the *Solanum* genus were bigger in size (between 6.7–58.3 kb) than those identified in other genera of the Solanaceae family. The largest syntenic block, about 58 kb, is shared with *S. commersonii* and *S. lycopersicum* (Figure S8). Within this syntenic block, two regions (i.e., *trnY*, *trnN*, *trnC*, *nad2* and *trnP*, *trnF*, *trnS*) were conserved with wild beet, *Cucurbita pepo* and diploid cotton *Gossypium raimondii*, and with wild beet, *G. raimondii* and coconut, respectively [3,89,90]. Two syntenic blocks containing *rps19*, *rps3*, *rpl16*, *cox2* and *trnW*, *trnP*, *nad9* and *trnH* genes are shared among species of the *Solanum* genus and *C. annuum*. Limited regions from these blocks (*rps3*, *rpl16* and *cox2* or *trnW*, *trnP*, *nad9*) were also conserved in *G. raimondii* and wild beet, respectively [89,91].

Since phylogenetic analysis of organellar genomes can identify evolutionary relationships accurately, a maximum likelihood phylogenetic tree was constructed by aligning the sequences of 15 protein-coding genes [92] from the available Solanaceae mitogenomes using *Ipomea nil* and *Vitis vinifera* as outgroup species. As expected, the sister relationships between wild and cultivated potato (*S. commersonii* and *S. tuberosum*), *S. lycopersicum* and *S. pennellii*, and between *N. tabacum* and *N. sylvestris* were all strongly supported (Figure 5). Furthermore, according to this phylogenetic reconstruction, potato species are closer to tomato species than to other Solanaceae and together form a strongly supported monophyletic lineage. The *Capsicum* genus is the closest relative to this lineage. Definitely, our results based on mtDNA sequences support the phylogenetic tree based on cpDNA sequences (i.e., combined *ndhF* and *trnL-trnF* regions) from 195 taxa [93].



**Figure 5.** Phylogenetic tree of Solanaceae species. Phylogram of the best maximum-likelihood (ML) tree as determined using the RAxML software from the concatenation of the coding sequences of 15 protein-coding genes. Numbers associated with branches are ML bootstrap support values. *Vitis vinifera* and *Ipomea nil* were used as outgroups.

### 3. Materials and Methods

#### 3.1. Plant Material

The *Solanum tuberosum* subsp. *tuberosum* cv. Désirée was micropropagated in vitro, transplanted in soil and grown in greenhouse. The *Solanum tuberosum* subsp. *tuberosum* cv. Cicero was obtained from a Strasbourg (France) local farmer.

#### 3.2. Isolation of mtDNA

Mitochondria were isolated from Désirée and Cicero tubers according to Pujol et al. [94] with some modifications. Briefly, after differential centrifugation mitochondria was purified on discontinuous Percoll gradient (14–28–45% *v/v*) and centrifuged at 70,000× *g* for 45 min (Beckman SW28 rotor). Mitochondria were collected at the 28–45% Percoll interface and washed twice with five volumes of washing buffer without BSA and centrifuged for 15 min at 10,000× *g*. Mitochondria were treated with DNase I for 45–60 min at 37 °C (1 mg/200 g of potato tubers, in 0.3 M sucrose, 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>), after which the reaction was diluted with three volumes of washing buffer without BSA and centrifuged for 20 min at 12,000× *g*. The mitochondrial pellet was washed twice with five volumes of washing buffer without BSA and the mtDNA extracted as previously described [95].

#### 3.3. Genome Sequencing and Assembly

Désirée. Sequencing was performed using both Illumina (Illumina, San Diego, CA, USA) and PacBio RS II single-molecule real-time (SMRT) sequencing technology (Pacific Biosciences, Menlo Park, CA, USA). A single sonication step was used for shear input DNA used for library construction with the TruSeq DNA kit (Illumina, San Diego, CA, USA). The sequencing (2 × 250 bp) was performed on a MiSeq device (estimated insert size ~500 bp).

FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) and Trimmomatic [96] were combined for assessing the overall quality of the sequencing run, to trim off poor quality bases, and to filter high-quality scores. High-quality reads were de novo assembled with Velvet (k-mer size 73) [97], filtering out contigs <200 bp in length. PacBio and Illumina data were used together to generate the final assembly. The hybrid scaffolding strategy was used. Briefly, using the existing assembly based on Illumina reads, PacBio sequences were used to join contigs with the AHA scaffolding algorithm [98] that is available through the SMRT analysis package (version 2.0, Pacific Biosciences). Assembled scaffolds were gap-filled and further scaffolded using GapCloser [99] with Illumina reads.

Cicero. Sequencing was performed using PacBio RS II single-molecule real-time (SMRT) sequencing technology (Pacific Biosciences, CA). PacBio data was assembled using the SMRT analysis package (version 2.2, Pacific Biosciences). The minimum coverage for correction was set to 35, and the minimum seed read length was set to 7 kb. The approximate genome size for the assembler module was set to 500 kb. Canu version 1.4 [100] was also used in order to inspect overlaps between corrected reads (extracted from the SMRT pipeline). Bandage [101] was used to draw overlapping read graphs. Contigs corresponding to plastid sequences were removed. Depth of sequencing coverage (Figure S9) was evaluated by mapping raw reads (prior to the correction step) to the assembled mitogenome. Possible configurations involving circularization of contigs or insertion of one into another by recombination were confirmed by PCR across the contig borders or across the repeated sequences involved in recombination, respectively.

#### 3.4. Genome Annotation

Auto-annotation was performed to identify most known genes by comparison with the tobacco mtDNA annotations (GenBank: NC\_006581.1). Annotations were then manually curated using the MacVector (MacVector, Inc., version 16.0.10) annotation tool.



### 3.5. Transcriptome Sequencing and Analysis

Total RNA extracted from pure mitochondria of the Cicero accession was sequenced as described elsewhere [102]. Briefly, RNAs were extracted with Tri Reagent<sup>®</sup> (Molecular Research Center, Cincinnati, OH, USA) according to the manufacturer's instructions and were DNase-treated with DNase RQ1 (Promega, Madison, WI, USA). The RNA-seq libraries, corresponding to three biological replicates, were prepared by the IGBMC Microarray and Sequencing Platform (Strasbourg, France) following Illumina protocols. The libraries were sequenced on an Illumina Genome Analyser IIX as paired-end 100 bp reads. A quality check on the raw data was performed using FastQC. Reads were aligned to the mtDNA sequence using Hisat2 version 2.0.5 [103]. Very high coverage was obtained: 61 out of 94, 29 out of 86 and 39 out of 76 million reads could be aligned to the mtDNA sequence.

### 3.6. Detection of Repeats

Repeats were searched by comparing the genome against itself using the Pustell DNA Matrix (Dot Plot) function of the MacVector package, parameters set to window size 30 bp, minimum score 90%, Hash value 8. Repeats of length greater than 100 bp were kept (Table S2). Primer3 included in the MacVector package was used for the design of PCR primers (Table S3).

### 3.7. Comparative and Evolutionary Analysis

The *S. tuberosum* mitogenome was compared with seven already published Solanaceae mitogenomes: *Solanum commersonii* (GenBank: MF989960.1, MF989961.1), *S. lycopersicum* (GenBank: NC\_035963.1), *S. pennellii* (GenBank: NC\_035964.1), *Nicotiana tabacum* (GenBank: NC\_006581.1), *N. sylvestris* (GenBank: NC\_029805.1), *Capsicum annuum* (GenBank: NC\_024624.1), *Hyoscyamus niger* (GenBank: NC\_026515.1). Syntenic blocks between mitogenomes were identified by BLASTn searches ( $e\text{-value} \leq 1 \times 10^{-5}$ ) based on the size of the alignment  $\geq 5$  kb and on sequence identity  $\geq 95\%$ . A maximum likelihood phylogenetic tree was constructed using the seven Solanaceae species aforementioned and two species, *Ipomea nil* (GenBank: NC\_031158) and *Vitis vinifera* (GenBank: NC\_012119), as outgroups. For comparison the coding sequences of 15 conserved protein-coding genes (*atp1*, *atp9*, *ccmB*, *cob*, *cox1*, *cox3*, *nad1*, *nad3*, *nad4*, *nad4L*, *nad6*, *nad7*, *nad9*, *rps3*, *rps4*) were extracted and a concatemer was assembled for each species. Sequences were then aligned in MAFFT v.7 [104] using default settings. RAxML v.8.2.12 [105] with the 'GTRGAMMA' evolutionary model under the rapid bootstrap algorithm with 1000 replicates was used to represent evolutionary relationships among species.

### 3.8. Data Deposition

The two mitogenomes reported in this paper were deposited in GenBank under accession numbers MN114537, MN114538, MN114539 (Cicero) and MN104801, MN104802, MN104803 (Désirée). Désirée raw sequences are available at the Sequence Read Archive (SRA) under the BioProject ID PRJNA544585. Raw RNA-seq data can be found in the Gene Expression Omnibus archive under the accession number GSE91388.

## 4. Conclusions

In the present study, we were able to resolve the mitogenome assembly of potato in 3 structural units by using long and short read sequencing data. The potato example adds to the few multichromosomal plant mitogenomes that cannot be resolved into a single circular map [2,11,12,14–16]. Isoforms due to recombination could be detected and confirmed by PCR. Long read sequences were thus very informative for better apprehending the recombination dynamics of plant mitogenomes. Comparative and phylogenetic analyses allowed us to identify large syntenic blocks among Solanaceae species.

We also provided a full picture of RNA processing using RNA-seq data, revealed numerous partially edited sites both in exon and intron sequences, potentially important methylation of the 18S rRNA and active transcription of several orfs of unknown function.

The detailed information obtained in this study for common potato will be very useful in future comparative analyses with mitogenomes of other tuber bearing species in order to better comprehend the co-evolution of nuclear and cytoplasmic genomes in this group of species. In particular, precise comparison of mitochondrial orfs, differentially present and/or expressed in various potato species as well as in rearranged mitogenomes of somatic hybrids, is of paramount importance to investigate molecular bases of nuclear-cytoplasmic interactions leading to cytoplasmic male sterility in some *Solanum* tuber-bearing interspecific hybrids [39]. The identification of mtDNA sequences involved in such a trait will be critical for developing a novel system to induce CMS also in other crops.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/19/4788/s1>.

**Author Contributions:** Conceptualization, T.C., L.D., J.M.G. and N.S.; Data curation, J.-S.V., N.D'.A., P.T., L.D., J.M.G., and N.S.; Formal analysis, J.-S.V., N.D'.A., S.G., R.T., C.C., E.U. and N.S.; Supervision, T.C., L.D., J.M.G. and N.S.; Writing—original draft preparation, J.-S.V., N.D'.A., P.T., R.T., J.M.G. and N.S.; Writing—review and editing, J.-S.V., N.D'.A., P.T., T.C., L.D., J.M.G. and N.S.

**Funding:** This research was partially funded by grants from the Italian Ministry of Research (MIUR) grant number: PON02\_00395\_3215002 “GenHORT” and by the Centre National de la Recherche Scientifique (CNRS) “Peps programme”. This work was also supported by the University of Strasbourg and is published under the framework of the LabEx (ANR-11-LABX-0057\_MITOCROSS) and benefits from the state managed by the French National Research Agency as part of the ‘Investments for the Future’ program.

**Acknowledgments:** We acknowledge Timothée Vincent for bioinformatic analysis of RNA-seq data and Anne-Marie Duchêne for providing the RNA-seq raw data. Technical assistance of G. Guarino and L. Sannino (CNR-IBBR, Portici, Italy) with artworks and plant growth is gratefully acknowledged.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Palmer, J.D.; Herbon, L.A. Unicircular structure of the *Brassica hirta* mitochondrial genome. *Curr. Genet.* **1987**, *11*, 565–570. [[CrossRef](#)] [[PubMed](#)]
2. Sloan, D.B.; Alverson, A.J.; Chuckalovcak, J.P.; Wu, M.; McCauley, D.E.; Palmer, J.D.; Taylor, D.R. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* **2012**, *10*, e1001241. [[CrossRef](#)] [[PubMed](#)]
3. Alverson, A.J.; Wei, X.; Rice, D.W.; Stern, D.B.; Barry, K.; Palmer, J.D. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.* **2010**, *27*, 1436–1448. [[CrossRef](#)] [[PubMed](#)]
4. Rodríguez-Moreno, L.; González, V.M.; Benjak, A.; Martí, M.C.; Puigdomènech, P.; Aranda, M.A.; Garcia-Mas, J. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genom.* **2011**, *12*, 424. [[CrossRef](#)] [[PubMed](#)]
5. Morley, S.A.; Nielsen, B.L. Plant mitochondrial DNA. *Molecules* **2017**, *15*, 17.
6. Richardson, A.O.; Palmer, J.D. Horizontal gene transfer in plants. *J. Exp. Bot.* **2006**, *58*, 1–9. [[CrossRef](#)] [[PubMed](#)]
7. Bendich, A.J. Reaching for the ring: The study of mitochondrial genome structure. *Curr. Genet.* **1993**, *24*, 279–290. [[CrossRef](#)]
8. Oldenburg, D.J.; Bendich, A.J. Size and structure of replicating mitochondrial DNA in cultured tobacco cells. *Plant Cell* **1996**, *8*, 447–461. [[CrossRef](#)]
9. Manchekar, M.; Scissum-Gunn, K.; Song, D.; Khazi, F.; McLean, S.L.; Nielsen, B.L. DNA recombination activity in soybean mitochondria. *J. Mol. Biol.* **2006**, *356*, 288–299. [[CrossRef](#)]
10. Gualberto, J.M.; Milesina, D.; Wallet, C.; Niazi, A.K.; Weber-Lotfi, F.; Dietrich, A. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* **2014**, *100*, 107–120. [[CrossRef](#)]

11. Alverson, A.J.; Rice, D.W.; Dickinson, S.; Barry, K.; Palmer, J.D. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* **2011**, *23*, 2499–2513. [[CrossRef](#)] [[PubMed](#)]
12. Sloan, D.B. One ring to rule them all? Genome sequencing provides new insights into the ‘master circle’ model of plant mitochondrial DNA structure. *New Phytol.* **2013**, *200*, 978–985. [[CrossRef](#)] [[PubMed](#)]
13. Davila, J.I.; Arrieta-Montiel, M.P.; Wamboldt, Y.; Cao, J.; Hagmann, J.; Shedje, V.; Xu, Y.-Z.; Weigel, D.; Mackenzie, S.A. Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. *BMC Biol.* **2011**, *9*, 64. [[CrossRef](#)] [[PubMed](#)]
14. Shearman, J.R.; Sonthirod, C.; Naktang, C.; Pootakham, W.; Yoocha, T.; Sangrakru, D.; Jomchai, N.; Tragoonrung, S.; Tangphatsornruang, S. The two chromosomes of the mitochondrial genome of a sugarcane cultivar: Assembly and recombination analysis using long PacBio reads. *Sci. Rep.* **2016**, *6*, 31533. [[CrossRef](#)] [[PubMed](#)]
15. Kazama, T.; Toriyama, K. Whole mitochondrial genome sequencing and re-examination of a cytoplasmic male sterility-associated gene in Boro-taichung-type cytoplasmic male sterile rice. *PLoS ONE* **2016**, *11*, e0159379. [[CrossRef](#)] [[PubMed](#)]
16. Tsujimura, M.; Kaneko, T.; Sakamoto, T.; Kimura, S.; Shigyo, M.; Yamagishi, H.; Terachi, T. Multichromosomal structure of the onion mitochondrial genome and a transcript analysis. *Mitochondrion* **2019**, *46*, 179–186. [[CrossRef](#)] [[PubMed](#)]
17. Wu, Z.; Cuthbert, J.M.; Taylor, D.R.; Sloan, D.B. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10185–10191. [[CrossRef](#)]
18. Halldén, C.; Lind, C.; Bryngelsson, T. Minicircle variation in Beta mitochondrial DNA. *Theor. Appl. Genet.* **1989**, *77*, 337–342. [[CrossRef](#)] [[PubMed](#)]
19. Warren, J.M.; Simmons, M.P.; Wu, Z.; Sloan, D.B. Linear plasmids and the rate of sequence evolution in plant mitochondrial genomes. *Genome Biol. Evol.* **2016**, *8*, 364–374. [[CrossRef](#)] [[PubMed](#)]
20. Backert, S. R-loop-dependent rolling-circle replication and a new model for DNA concatemer resolution by mitochondrial plasmid mp1. *EMBO J.* **2002**, *21*, 3128–3136. [[CrossRef](#)]
21. Christensen, A.C. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. *Genome Biol. Evol.* **2013**, *5*, 1079–1086. [[CrossRef](#)] [[PubMed](#)]
22. Gualberto, J.M.; Newton, K.J. Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* **2017**, *68*, 225–252. [[CrossRef](#)] [[PubMed](#)]
23. Horn, R.; Gupta, K.J.; Colombo, N. Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* **2014**, *19*, 198–205. [[CrossRef](#)] [[PubMed](#)]
24. Chen, Z.; Zhao, N.; Li, S.; Grover, C.E.; Nie, H.; Wendel, J.F.; Hua, J. Plant mitochondrial genome evolution and cytoplasmic male sterility. *Crit. Rev. Plant Sci.* **2017**, *36*, 55–69. [[CrossRef](#)]
25. Kim, Y.-J.; Zhang, D. Molecular control of male fertility for crop hybrid breeding. *Trends Plant Sci.* **2018**, *23*, 53–65. [[CrossRef](#)] [[PubMed](#)]
26. Takenaka, M.; van der Merwe, J.A.; Verbitskiy, D.; Neuwirt, J.; Zehrmann, A.; Brennicke, A. RNA Editing in Plant Mitochondria. In *RNA Editing*; Göringer, H.U., Ed.; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2008; pp. 105–122. [[CrossRef](#)]
27. Takenaka, M.; Verbitskiy, D.; van der Merwe, J.A.; Zehrmann, A.; Brennicke, A. The process of RNA editing in plant mitochondria. *Mitochondrion* **2008**, *8*, 35–46. [[CrossRef](#)] [[PubMed](#)]
28. Sun, T.; Bentolila, S.; Hanson, M.R. The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant Sci.* **2016**, *21*, 962–973. [[CrossRef](#)]
29. Tang, W.; Luo, C. Molecular and Functional Diversity of RNA Editing in Plant Mitochondria. *Mol. Biotechnol.* **2018**, *60*, 935–945. [[CrossRef](#)]
30. Oldenkott, B.; Yang, Y.; Lesch, E.; Knoop, V.; Schallenberg-Rüdinger, M. Plant-type pentatricopeptide repeat proteins with a DYW domain drive C-to-U RNA editing in *Escherichia coli*. *Commun. Biol.* **2019**, *2*, 85. [[CrossRef](#)]
31. Hardigan, M.A.; Laimbeer, F.P.E.; Newton, L.; Crisovan, E.; Hamilton, J.P.; Vaillancourt, B.; Wiegert-Rininger, K.; Wood, J.C.; Douches, D.S.; Farré, E.M.; et al. Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E9999. [[CrossRef](#)]

32. The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* **2011**, *475*, 189. [[CrossRef](#)] [[PubMed](#)]
33. McGregor, C.E.; Lambert, C.A.; Greyling, M.M.; Louw, J.H.; Warnich, L. A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. *Euphytica* **2000**, *113*, 135–144. [[CrossRef](#)]
34. Ghislain, M.; Andrade, D.; Rodríguez, F.; Hijmans, R.J.; Spooner, D.M. Genetic analysis of the cultivated potato *Solanum tuberosum* L. *Phureja* Group using RAPDs and nuclear SSRs. *Theor. Appl. Genet.* **2006**, *113*, 1515–1527. [[CrossRef](#)]
35. Spooner, D.M.; Núñez, J.; Trujillo, G.; del Rosario Herrera, M.; Guzmán, F.; Ghislain, M. Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19398–19403. [[CrossRef](#)] [[PubMed](#)]
36. Peters, S.A.; Bargsten, J.W.; Szinay, D.; van de Belt, J.; Visser, R.G.F.; Bai, Y.; de Jong, H. Structural homology in the Solanaceae: Analysis of genomic regions in support of synteny studies in tomato, potato and pepper. *Plant J.* **2012**, *71*, 602–614. [[CrossRef](#)]
37. Smyda-Dajmund, P.; Śliwka, J.; Wasilewicz-Flis, I.; Jakuczun, H.; Zimnoch-Guzowska, E. Genetic composition of interspecific potato somatic hybrids and autofused 4x plants evaluated by DARt and cytoplasmic DNA markers. *Plant Cell Rep.* **2016**, *35*, 1345–1358. [[CrossRef](#)] [[PubMed](#)]
38. Berdugo-Cely, J.; Valbuena, R.I.; Sánchez-Betancourt, E.; Barrero, L.S.; Yockteng, R. Genetic diversity and association mapping in the Colombian Central Collection of *Solanum tuberosum* L. Andigenum group using SNPs markers. *PLoS ONE* **2017**, *12*, e0173039. [[CrossRef](#)] [[PubMed](#)]
39. Cardi, T.; Bastia, T.; Monti, L.; Earle, E.D. Organelle DNA and male fertility variation in *Solanum* spp. and interspecific somatic hybrids. *Theor. Appl. Genet.* **1999**, *99*, 819–828. [[CrossRef](#)]
40. Lössl, A.; Adler, N.; Horn, R.; Frei, U.; Wenzel, G. Chondriome-type characterization of potato: Mt  $\alpha$ ,  $\beta$ ,  $\gamma\delta\epsilon$  and novel plastid-mitochondrial configurations in somatic hybrids. *Theor. Appl. Genet.* **1999**, *99*, 1–10. [[CrossRef](#)]
41. Scotti, N.; Monti, L.; Cardi, T. Organelle DNA variation in parental *Solanum* spp. genotypes and nuclear-cytoplasmic interactions in *Solanum tuberosum* (+) *S. commersonii* somatic hybrid-backcross progeny. *Theor. Appl. Genet.* **2003**, *108*, 87–94. [[CrossRef](#)]
42. Scotti, N.; Cozzolino, S.; Cardi, T. Mitochondrial DNA variation in cultivated and wild potato species (*Solanum* spp.). *Genome* **2007**, *50*, 706–713. [[CrossRef](#)] [[PubMed](#)]
43. Lössl, A.; Götz, M.; Braun, A.; Wenzel, G. Molecular markers for cytoplasm in potato: Male sterility and contribution of different plastid-mitochondrial configurations to starch production. *Euphytica* **2000**, *116*, 221–230. [[CrossRef](#)]
44. Brandt, P.; Unseld, M.; Eckert-Ossenkopp, U.; Brennicke, A. An *rps14* pseudogene is transcribed and edited in Arabidopsis mitochondria. *Curr. Genet.* **1993**, *24*, 330–336. [[CrossRef](#)] [[PubMed](#)]
45. Quiñones, V.; Zanolungo, S.; Moenne, A.; Gómez, I.; Holuigue, L.; Litvak, S.; Jordana, X. The *rpl5-rps 14-cob* gene arrangement in *Solanum tuberosum*: *rps14* is a transcribed and unedited pseudogene. *Plant Mol. Biol.* **1996**, *31*, 937–943. [[CrossRef](#)] [[PubMed](#)]
46. Scotti, N.; Maréchal-Drouard, L.; Cardi, T. The *rpl5-rps14* mitochondrial region: A hot spot for DNA rearrangements in *Solanum* spp. somatic hybrids. *Curr. Genet.* **2004**, *45*, 378–382. [[CrossRef](#)] [[PubMed](#)]
47. Figueroa, P.; Gómez, I.; Holuigue, L.; Araya, A.; Jordana, X. Transfer of *rps14* from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J.* **1999**, *18*, 601–609. [[CrossRef](#)]
48. Ong, H.C.; Palmer, J.D. Pervasive survival of expressed mitochondrial *rps14* pseudogenes in grasses and their relatives for 80 million years following three functional transfers to the nucleus. *BMC Evol. Biol.* **2006**, *6*, 55. [[CrossRef](#)]
49. Bastia, T.; Scotti, N.; Cardi, T. Organelle DNA analysis of *Solanum* and *Brassica* somatic hybrids by PCR with 'universal primers'. *Theor. Appl. Genet.* **2001**, *102*, 1265–1272. [[CrossRef](#)]
50. Maréchal-Drouard, L.; Guillemaut, P.; Cosset, A.; Arbogast, M.; Weber, F.; Weil, J.H.; Dietrich, A. Transfer RNAs of potato (*Solanum tuberosum*) mitochondria have different genetic origins. *Nucleic Acids Res.* **1990**, *18*, 3689–3696. [[CrossRef](#)] [[PubMed](#)]
51. Dietrich, A.; Wallet, C.; Iqbal, R.K.; Gualberto, J.M.; Lotfi, F. Organellar non-coding RNAs: Emerging regulation mechanisms. *Biochimie* **2015**, *117*, 48–62. [[CrossRef](#)] [[PubMed](#)]

52. Kim, D.H.; Kang, J.G.; Kim, B.-D. Isolation and characterization of the cytoplasmic male sterility-associated *orf456* gene of chili pepper (*Capsicum annuum* L.). *Plant Mol. Biol.* **2007**, *63*, 519–532. [[CrossRef](#)] [[PubMed](#)]
53. Fey, J.; Maréchal-Drouard, L. Compilation and Analysis of Plant Mitochondrial Promoter Sequences: An Illustration of a Divergent Evolution between Monocot and Dicot Mitochondria. *Biochem. Biophys. Res. Commun.* **1999**, *256*, 409–414. [[CrossRef](#)] [[PubMed](#)]
54. Kühn, K.; Weihe, A.; Börner, T. Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. *Nucleic Acids Res.* **2005**, *33*, 337–346. [[CrossRef](#)]
55. Kühn, K.; Bohne, A.-V.; Liere, K.; Weihe, A.; Börner, T. Arabidopsis phage-type RNA polymerases: Accurate *in vitro* transcription of organellar genes. *Plant Cell* **2007**, *19*, 959–971. [[CrossRef](#)] [[PubMed](#)]
56. Forner, J.; Weber, B.; Thuss, S.; Wildum, S.; Binder, S. Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: T-elements contribute to 5' and 3' end formation. *Nucleic Acids Res.* **2007**, *35*, 3676–3692. [[CrossRef](#)]
57. Palmer, J.D.; Herbon, L.A. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **1988**, *28*, 87–97. [[CrossRef](#)] [[PubMed](#)]
58. Zanlungo, S.; Quinones, V.; Moenne, A.; Holuigue, L.; Jordana, X. Splicing and editing of *rps10* transcripts in potato mitochondria. *Curr. Genet.* **1995**, *27*, 565–571. [[CrossRef](#)] [[PubMed](#)]
59. Zanlungo, S.; Quiñones, V.; Holuigue, L.; Jordana, X. Identification of the RPS10 polypeptide encoded by the mitochondrial genome in *Solanum tuberosum*: Translation initiates at a genomic-encoded AUG and not at a conserved AUG codon created by RNA editing. *Physiol. Plant* **2000**, *110*, 256–261. [[CrossRef](#)]
60. Quiñones, V.; Zanlungo, S.; Holuigue, L.; Litvak, S.; Jordana, X. The *cox1* initiation codon is created by RNA editing in potato mitochondria. *Plant Physiol.* **1995**, *108*, 1327. [[CrossRef](#)]
61. Dell'Orto, P.; Moenne, A.; Graves, P.V.; Jordana, X. The potato mitochondrial ATP synthase subunit 9: Gene structure, RNA editing and partial protein sequence. *Plant Sci.* **1993**, *88*, 45–53. [[CrossRef](#)]
62. Hunt, M.D.; Newton, K.J. The NCS3 mutation: Genetic evidence for the expression of ribosomal protein genes in *Zea mays* mitochondria. *EMBO J.* **1991**, *10*, 1045–1052. [[CrossRef](#)] [[PubMed](#)]
63. Takemura, M.; Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Nozato, N.; Akashi, K.; Ohshima, K. Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort, *Marchantia polymorpha*. *Nucleic Acids Res.* **1992**, *20*, 3199–3205. [[CrossRef](#)] [[PubMed](#)]
64. Giegé, P.; Brennicke, A. RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 15324–15329. [[CrossRef](#)] [[PubMed](#)]
65. Marechal-Drouard, L.; Cosset, A.; Remacle, C.; Ramamonjisoa, D.; Dietrich, A. A single editing event is a prerequisite for efficient processing of potato mitochondrial phenylalanine tRNA. *Mol. Cell. Biol.* **1996**, *16*, 3504–3510. [[CrossRef](#)] [[PubMed](#)]
66. Fey, J.; Tomita, K.; Bergdoll, M.; Maréchal-Drouard, L. Evolutionary and functional aspects of C-to-U editing at position 28 of tRNACys(GCA) in plant mitochondria. *RNA* **2000**, *6*, 470–474. [[CrossRef](#)] [[PubMed](#)]
67. Ngu, M.; Massel, K.; Bonen, L. Group II introns in wheat mitochondria have degenerate structural features and varied splicing pathways. *Int. J. Biochem. Cell Biol.* **2017**, *91*, 156–167. [[CrossRef](#)]
68. Cognat, V.; Morelle, G.; Megel, C.; Lalande, S.; Molinier, J.; Vincent, T.; Small, I.; Duchêne, A.-M.; Maréchal-Drouard, L. The nuclear and organellar tRNA-derived RNA fragment population in *Arabidopsis thaliana* is highly dynamic. *Nucleic Acids Res.* **2016**, *45*, 3460–3472. [[CrossRef](#)]
69. Guo, W.; Zhu, A.; Fan, W.; Mower, J.P. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. *New Phytol.* **2017**, *213*, 391–403. [[CrossRef](#)]
70. Adams, K.L.; Daley, D.O.; Whelan, J.; Palmer, J.D. Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell* **2002**, *14*, 931–943. [[CrossRef](#)]
71. Notsu, Y.; Masood, S.; Nishikawa, T.; Kubo, N.; Akiduki, G.; Nakazono, M.; Hirai, A.; Kadowaki, K. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genom.* **2002**, *268*, 434–445. [[CrossRef](#)]
72. Aubert, D.; Bisanz-Seyer, C.; Herzog, M. Mitochondrial *rps14* is a transcribed and edited pseudogene in *Arabidopsis thaliana*. *Plant Mol. Biol.* **1992**, *20*, 1169–1174. [[CrossRef](#)] [[PubMed](#)]
73. Unseld, M.; Marienfeld, J.R.; Brandt, P.; Brennicke, A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **1997**, *15*, 57–61. [[CrossRef](#)] [[PubMed](#)]

74. Kubo, N.; Harada, K.; Hirai, A.; Kadowaki, K. A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: Common use of the same mitochondrial targeting signal for different proteins. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9207–9211. [[CrossRef](#)] [[PubMed](#)]
75. Mohanty, A.; Martín, J.P.; González, L.M.; Aguinagalde, I. Association between chloroplast DNA and mitochondrial DNA haplotypes in *Prunus spinosa* L. (Rosaceae) populations across Europe. *Ann. Bot.* **2003**, *92*, 749–755. [[CrossRef](#)] [[PubMed](#)]
76. Bartoszewski, G.; Malepszy, S.; Havey, M.J. Mosaic (MSC) cucumbers regenerated from independent cell cultures possess different mitochondrial rearrangements. *Curr. Genet.* **2004**, *45*, 45–53. [[CrossRef](#)] [[PubMed](#)]
77. Sandoval, P.; León, G.; Gómez, I.; Carmona, R.; Figueroa, P.; Holuigue, L.; Araya, A.; Jordana, X. Transfer of RPS14 and RPL5 from the mitochondrion to the nucleus in grasses. *Gene* **2004**, *324*, 139–147. [[CrossRef](#)]
78. Adams, K.L.; Qiu, Y.-L.; Stoutemyer, M.; Palmer, J.D. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9905–9912. [[CrossRef](#)]
79. Bergthorsson, U.; Adams, K.L.; Thomason, B.; Palmer, J.D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **2003**, *424*, 197–201. [[CrossRef](#)]
80. Park, S.; Grewe, F.; Zhu, A.; Ruhlman, T.A.; Sabir, J.; Mower, J.P.; Jansen, R.K. Dynamic evolution of Geranium mitochondrial genomes through multiple horizontal and intracellular gene transfers. *New Phytol.* **2015**, *208*, 570–583. [[CrossRef](#)]
81. Handa, H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): Comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.* **2003**, *31*, 5907–5916. [[CrossRef](#)]
82. Clifton, S.W.; Minx, P.; Fauron, C.M.R.; Gibson, M.; Allen, J.O.; Sun, H.; Thompson, M.; Barbazuk, W.B.; Kanuganti, S.; Tayloe, C.; et al. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* **2004**, *136*, 3486–3503. [[CrossRef](#)] [[PubMed](#)]
83. Luo, D.; Xu, H.; Liu, Z.; Guo, J.; Li, H.; Chen, L.; Fang, C.; Zhang, Q.; Bai, M.; Yao, N.; et al. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat. Genet.* **2013**, *45*, 573. [[CrossRef](#)] [[PubMed](#)]
84. Osman, C.; Wilmes, C.; Tatsuta, T.; Langer, T. Prohibitins interact genetically with Atp23, a novel processing peptidase and chaperone for the F1Fo-ATP synthase. *Mol. Biol. Cell* **2007**, *18*, 627–635. [[CrossRef](#)] [[PubMed](#)]
85. Zeng, X.; Neupert, W.; Tzagoloff, A. The metalloprotease encoded by ATP23 has a dual function in processing and assembly of subunit 6 of mitochondrial ATPase. *Mol. Biol. Cell* **2007**, *18*, 617–626. [[CrossRef](#)] [[PubMed](#)]
86. Fox, T.D. Mitochondrial protein synthesis, import, and assembly. *Genetics* **2012**, *192*, 1203–1234. [[CrossRef](#)] [[PubMed](#)]
87. Kimura, M.; Ohta, T. On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 2848–2852. [[CrossRef](#)] [[PubMed](#)]
88. Echave, J.; Spielman, S.J.; Wilke, C.O. Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **2016**, *17*, 109–121. [[CrossRef](#)] [[PubMed](#)]
89. Darracq, A.; Varré, J.-S.; Maréchal-Drouard, L.; Courseaux, A.; Castric, V.; Saumitou-Laprade, P.; Oztas, S.; Lenoble, P.; Vacherie, B.; Barbe, V. Structural and content diversity of mitochondrial genome in beet: A comparative genomic analysis. *Genome Biol. Evol.* **2011**, *3*, 723–736. [[CrossRef](#)]
90. Aljohi, H.A.; Liu, W.; Lin, Q.; Zhao, Y.; Zeng, J.; Alamer, A.; Alanazi, I.O.; Alawad, A.O.; Al-Sadi, A.M.; Hu, S.; et al. Complete Sequence and Analysis of Coconut Palm (*Cocos nucifera*) Mitochondrial Genome. *PLoS ONE* **2016**, *11*, e0163990. [[CrossRef](#)]
91. Bi, C.; Paterson, A.H.; Wang, X.; Xu, Y.; Wu, D.; Qu, Y.; Jiang, A.; Ye, Q.; Ye, N. Analysis of the Complete Mitochondrial Genome Sequence of the Diploid Cotton *Gossypium raimondii* by Comparative Genomics Approaches. *BioMed Res. Int.* **2016**, *2016*. [[CrossRef](#)]
92. Wang, X.; Cheng, F.; Rohlsen, D.; Bi, C.; Wang, C.; Xu, Y.; Wei, S.; Ye, Q.; Yin, T.; Ye, N. Organellar genome assembly methods and comparative analysis of horticultural plants. *Hortic. Res.* **2018**, *5*, 3. [[CrossRef](#)] [[PubMed](#)]
93. Olmstead, R.G.; Bohs, L.; Migid, H.A.; Santiago-Valentin, E.; Garcia, V.F.; Collier, S.M. A Molecular Phylogeny of the Solanaceae. *Taxon* **2008**, *57*, 1159–1181. [[CrossRef](#)]

94. Pujol, C.; Bailly, M.; Kern, D.; Maréchal-Drouard, L.; Becker, H.; Duchêne, A.-M. Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and chloroplastic Gln-tRNA<sup>Gln</sup> synthesis in plants. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6481–6485. [[CrossRef](#)] [[PubMed](#)]
95. Scotti, N.; Cardi, T.; Marechal-drouard, L. Mitochondrial DNA and RNA isolation from small amounts of potato tissue. *Plant Mol. Biol. Rep.* **2001**, *19*, 67. [[CrossRef](#)]
96. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, UK)* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
97. Zerbino, D.R.; Birney, E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821–829. [[CrossRef](#)] [[PubMed](#)]
98. Bashir, A.; Klammer, A.; Robins, W.P.; Chin, C.-S.; Webster, D.; Paxinos, E.; Hsu, D.; Ashby, M.; Wang, S.; Peluso, P.; et al. A hybrid approach for the automated finishing of bacterial genomes. *Nat. Biotechnol.* **2012**, *30*, 701–707. [[CrossRef](#)] [[PubMed](#)]
99. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **2012**, *1*, 18. [[CrossRef](#)] [[PubMed](#)]
100. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [[CrossRef](#)]
101. Holt, K.E.; Schultz, M.B.; Wick, R.R.; Zobel, J. Bandage: Interactive visualization of *de novo* genome assemblies. *Bioinformatics (Oxford, UK)* **2015**, *31*, 3350–3352.
102. Vincent, T.; Vingadassalon, A.; Ubrig, E.; Azeredo, K.; Srouf, O.; Cognat, V.; Graindorge, S.; Salinas, T.; Maréchal-Drouard, L.; Duchêne, A.-M. A genome-scale analysis of mRNAs targeting to plant mitochondria: Upstream AUGs in 5' untranslated regions reduce mitochondrial association. *Plant J.* **2017**, *92*, 1132–1142. [[CrossRef](#)] [[PubMed](#)]
103. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357. [[CrossRef](#)] [[PubMed](#)]
104. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)] [[PubMed](#)]
105. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, UK)* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).