



Article

Comparative Mitogenome Analysis of the Genus *Trifolium* Reveals Independent Gene Fission of *ccmFn* and Intracellular Gene Transfers in Fabaceae

In-Su Choi ^{1,*}, Tracey A. Ruhlman ¹ and Robert K. Jansen ^{1,2}

¹ Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA; truhlman@austin.utexas.edu (T.A.R.); jansen@austin.utexas.edu (R.K.J.)

² Centre of Excellence in Bionanoscience Research, Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia

* Correspondence: 86ischoi@gmail.com

Received: 11 January 2020; Accepted: 17 February 2020; Published: 13 March 2020



Abstract: The genus *Trifolium* is the largest of the tribe Trifolieae in the subfamily Papilionoideae (Fabaceae). The paucity of mitochondrial genome (mitogenome) sequences has hindered comparative analyses among the three genomic compartments of the plant cell (nucleus, mitochondrion and plastid). We assembled four mitogenomes from the two subgenera (*Chronosemium* and *Trifolium*) of the genus. The four *Trifolium* mitogenomes were compact (294,911–348,724 bp in length) and contained limited repetitive (6.6–8.6%) DNA. Comparison of organelle repeat content highlighted the distinct evolutionary trajectory of plastid genomes in a subset of *Trifolium* species. Intracellular gene transfer (IGT) was analyzed among the three genomic compartments revealing functional transfer of mitochondrial *rps1* to nuclear genome along with other IGT events. Phylogenetic analysis based on mitochondrial and nuclear *rps1* sequences revealed that the functional transfer in Trifolieae was independent from the event that occurred in robinoid clade that includes genus *Lotus*. A novel, independent fission event of *ccmFn* in *Trifolium* was identified, caused by a 59 bp deletion. Fissions of this gene reported previously in land plants were reassessed and compared with *Trifolium*.

Keywords: legumes; clover; organelle genetics; mitochondria; endosymbiotic gene transfer; gene fission

1. Introduction

Plant cells comprise three genomic compartments (i.e., nucleus, mitochondrion and plastid). Unlike the typically conservative plastid genome (plastome) [1,2], plant mitochondrial genomes (mitogenome) display drastic evolutionary plasticity in size, content and structure, intracellular gene transfer (IGT) and interspecific horizontal gene transfer [3–8]. Substitution rates of mitochondrial protein coding genes, however, are the most conservative among the three genomic compartments [9]. In angiosperms, the relative rate of synonymous substitutions of mitogenome, plastome and nuclear genome is 1:3:16 [10].

Extensive gene loss and IGT of organelle DNA to the nucleus occurred in the early stages of endosymbiosis [11]. Nuclear genome sequences that originate from the mitogenome and plastome are referred to as nuclear mitochondrial DNA sequences (NUMTs) and nuclear plastid DNA sequences (NUPTs), respectively [12,13]. Transfer of mitochondrial DNA to the nuclear genome is an ongoing process in both of plants and animals but functional transfer of mitochondrial genes has almost ceased in animals [14]. Functional transfer of mitochondrial genes in plants has often involved ribosomal protein or succinate dehydrogenase genes [5]. Transfer of mitochondrial genes to the nuclear genome cannot substitute function of the original mitochondrial copy unless the nuclear copy

acquires the appropriate expression and targeting signals [15]. Before the acquisition of regulatory signals, NUMTs must survive mutational decay in nuclear genome, which limits the lifespan of the nonfunctional sequences [16]. Mitochondrial IGT events may be successful or unsuccessful in terms of functionality and the phylogenetic distribution of pseudogenization and deletion of mitochondrial genes can be assessed in descendant lineages [17].

Following functional transfer, NUMTs attain higher substitution rates than their mitochondrial counterparts [14] because of substantial differences in the synonymous substitution rate between mitochondrial and nuclear genomes in plants [10]. Hence, functional transfer of mitochondrial genes into the nucleus is often detected by the presence of intact but highly diverged copies in nuclear genome compared to mitochondrial copies [18,19]. On rare occasions, functional transfers of mitochondrial genes exhibit an intriguing situation in which the nuclear and mitochondrial genomes contain different portions of the coding region resulting from mitochondrial gene fission and IGT (e.g., *rpl2* in many of eudicots) [20]. Szafranski [21] named this process “intercompartmental piecewise gene transfer.” In plant mitogenome evolution, the protein that most commonly undergoes gene fission is cytochrome c maturation protein *ccmF* [22–25].

In *Escherichia coli*, the eight *ccm* genes (*ccmA-H*) are clustered in a single locus [26]. In most plants, three *ccm* genes (*ccmA*, *ccmE* and *ccmH*) have been transferred from the mitogenome to the nuclear genome, two (*ccmD* and *ccmG*) were lost and four (*ccmB*, *ccmC*, *ccmFc* and *ccmFn*) remain in the mitochondrion [27]. Since the fission of *ccmF* into *ccmFc* and *ccmFn* is shared by liverworts and seed plants [3], this event happened early in land plant evolution. In addition, there were independent fissions of *ccmF* in several lineages of land plants, including fission of *ccmFc* into *ccmFc1* and *ccmFc2* in *Marchantia* [22] and fissions of *ccmFn* into *ccmFn1* and *ccmFn2* in Brassicaceae [23,24] and *Allium* (Amaryllidaceae) [25].

Fabaceae are the third largest angiosperm family with approximately 20,000 species in six subfamilies [28]. Most species diversity occurs in subfamily Papilionoideae, which includes many economically important species [29]. The inverted repeat (IR) lacking clade (IRLC) is one of the major groups of Papilionoideae, which is defined by absence of the canonical plastome IR (~25 kb) [30]. Plastome studies of the IRLC elucidated several rare evolutionary phenomena, including high degree of genome rearrangement [31], localized hypermutation [32], genome size expansion with accumulation of dispersed repeats and unique sequences of unknown origin [33–35] and re-acquisition of a large IR [36]. However, mitogenome evolution in IRLC is poorly understood and represented by only two species, *Vicia faba* (tribe Fabeae) [37] and *Medicago truncatula* (tribe Trifolieae) [38]. In Trifolieae, a study of the mitochondrial *rps1* gene documented the existence of functional nuclear copies and putatively pseudogenized mitochondrial copies from three genera (*Medicago*, *Melilotus* and *Trigonella*) [39]. Deletion of mitochondrial *rps1* was also identified from another papilionoid species, *Lotus japonicus* [40]. The status of mitochondrial *rps1* across Trifolieae and related taxa has not been examined until recently. Parallel losses of several mitochondrial genes in Fabaceae were revealed in a previous study, however, whether the losses represent a single ancestral IGT or multiple IGTs was not determined [8].

Trifolium is the largest genus (ca. 250 species) of the tribe Trifolieae [41] and is divided into two subgenera (*Chronosemium* and *Trifolium*) [42]. Trifolieae belong to the IRLC and are closely related to Fabeae [30]. Several evolutionary studies of Trifolieae plastid [31,33,34] and nuclear [43–46] genomes have been conducted but mitogenome comparisons of *Trifolium* have been neglected. In this study, gene content, size and repeat structure of mitogenomes of four *Trifolium* species from the two subgenera *Chronosemium* (*T. aureum* and *T. grandiflorum*) and *Trifolium* (*T. meduseum* and *T. pratense*) were examined and compared to related papilionoid species.

2. Results

2.1. Mitogenome Features of Four *Trifolium* Species

For each of four *Trifolium*, a single chromosome was assembled that contained all expected mitochondrial coding sequences. The length of the four mitogenomes varied from 294,911 to 348,724 bp (Table 1). The GC content was conserved among the species at 44.9–45.2 %. Gene content was identical with three rRNAs, 16 tRNAs and 32 protein coding genes while gene order was distinct for each species (Figure 1).

Table 1. Assembly information for four *Trifolium* mitogenomes.

Species	Subgenus	Raw Reads	Mitogenome Coverage	Mitogenome Coverage	Mitogenome Length (bp)	GC (%)	NCBI Accession
<i>T. aureum</i>	<i>Chronosentium</i>	71,346,514	645	645	294,911	44.9	MT039392
<i>T. grandiflorum</i>	<i>Chronosentium</i>	48,390,678	197	197	347,723	45.1	MT039391
<i>T. meduseum</i>	<i>Trifolium</i>	68,712,286	207	207	348,724	45.0	MT039390
<i>T. pratense</i>	<i>Trifolium</i>	47,909,108	212	212	301,823	45.2	MT039389

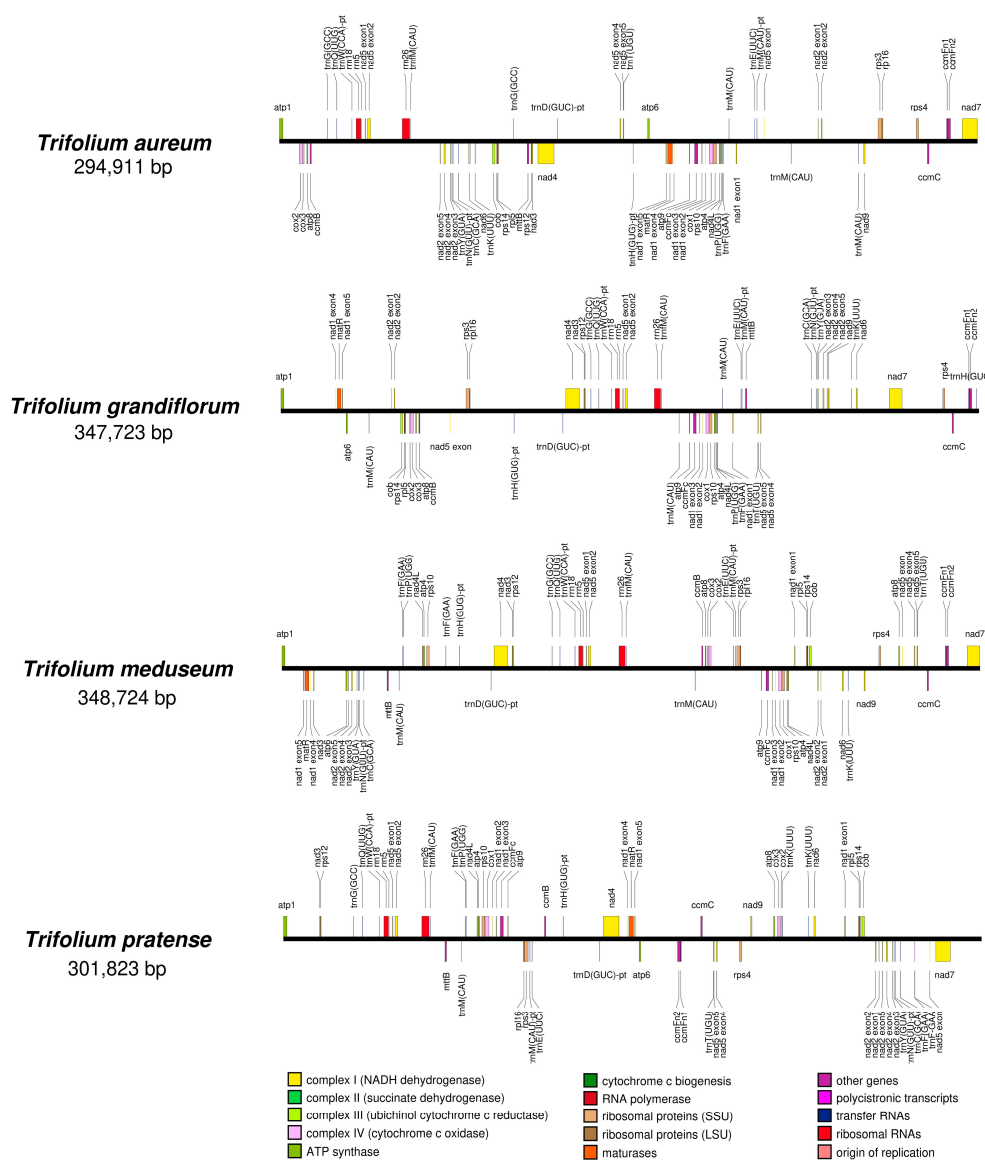


Figure 1. Linear mitogenome maps of four *Trifolium* species. Fragmented genes caused by duplication or pseudogenization are not depicted. pt indicates tRNAs of plastid origin.

Int. J. Mol. Sci. 2020, 21, 1959
 Figure 2. Alignment of six *ccmFn* sequences of mitogenomes from four *Trifolium* species. Fragmented genes caused by duplication or pseudogenization are not depicted. pt indicates tRNAs of plastid origin.

Gene and intron content comparison with other published mitogenomes revealed one gene loss (*rps1*) (Figure S1), which was shared with *Lotus* and two cis-spliced intron losses (*ccmFn*1829 and *rps31*174) that were exclusive to *Trifolium* (Figure S2). Sequence alignment of *ccmFn* from *Trifolium* with other IRLC genera revealed a 59 bp deletion that resulted in a frame shift and premature stop codon (Figure 2); a putative downstream start codon for a second open reading frame (ORF) (*ccmFn*2) was also identified.

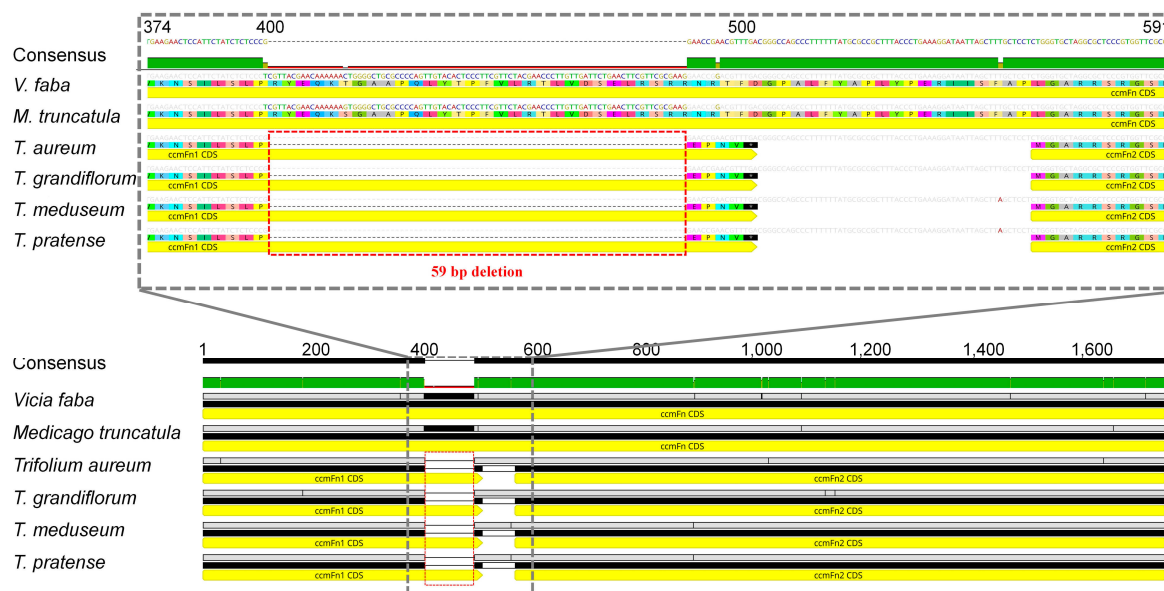


Figure 2: Fission of *ccmFn* in four *Trifolium*. Alignment of six *ccmFn* sequences of mitogenomes from species in inverted repeat lacking clade. The region (gray dashed rectangle, aligned positions 374–591) showing the 59 bp deletion (red dotted box) is enlarged above. Translated amino acid alignments are presented below corresponding nucleotide sequence alignments. Nucleotide coordinates are indicated above consensus of alignment. Sequence identity is shown below consensus (green = 100%, yellow = at least 30% or under 100%, red = below 30%).

2.2. Repeat Composition of Organelle Genomes in *Trifolium*

Repeat sequences were estimated for mitogenomes and thirteen plastomes (Table 2). The amount of repetitive sequences in mitogenomes was not highly variable (6.6–8.6%). In contrast, the amount of repetitive DNA in plastomes was highly variable (1.4–20.7%) and can be divided into two non-overlapping ranges that corresponded to two groups of two species (subgen. *Chronosemium* and subg. *Trifolium*; *Trifolium* *Brassicarum*, 1.4%–1.7% and five sects. of subg. *Trifolium* (*Lupinaster*, *Trichocephalum*, *Trifolium*, *Vesicarium* and *Trifoliarum*, 10.7%–20.7%). The contrasting repeat composition between organelle genomes was particularly evident in *T. pratense*, which had small amounts of repeat sequences in its mitogenome and the largest amount in its plastome (Figure 3; Table 2).

Table 2. Comparison of repeat percentage between organelle genomes in *Trifolium*.

Species	Subgenus	Section	Mitogenome			Plastome		
			Size (bp)	Repeat (%)	NCBI Accession	Size (bp)	Repeat (%)	NCBI Accession
<i>T. aureum</i>	<i>Chronosemium</i>	<i>Chronosemium</i>	294,911	8.6	MT039392	126,970	5.2	NC_024035

<i>glanduliferum</i>	<i>Trifolium</i>	<i>Parumesus</i>	Not applicable	9	4.8	44		
<i>T. lupinaster</i>	<i>Trifolium</i>	<i>Lupinaster</i>	Not applicable	135,049	10.9	KJ788287		
<i>T. subterraneum</i>	<i>Trifolium</i>	<i>Trichocephalum</i>	Not applicable	144,763	19.7	NC_011828		
<i>T. meduseum</i>	<i>Trifolium</i>	<i>Trichocephalum</i>	348,724	8.5	MT039390	142,595		
<i>T. pratense</i>	<i>Trifolium</i>	<i>Trifolium</i>	301,823	6.6	MT039389	146,573		
Species	Subgenus	Section	Size (bp)	Repeat (%)	NCBI Accession	Size (bp)	Repeat (%)	NCBI Accession
<i>T. hybridum</i>	<i>Trifolium</i>	<i>Vesicastrum</i>	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
<i>T. aureum</i>	<i>Chronosemium</i>	<i>Chronosemium</i>	294,911	8.6	MT039392	126,970	1	5.2
<i>T. grandiflorum</i>	<i>Chronosemium</i>	<i>Chronosemium</i>	347,723	8.4	MT039391	125,628	4.7	NC_024034
<i>T. boissieri</i>	<i>Chronosemium</i>	<i>Chronosemium</i>	Not applicable	Not applicable	Not applicable	125,740	5.1	NC_025743
<i>T. strictum</i>	<i>Chronosemium</i>	<i>Chronosemium</i>	Not applicable	Not applicable	Not applicable	125,834	4.4	NC_025745
<i>T. glanduliferum</i>	<i>Trifolium</i>	<i>Paramesius</i>	Not applicable	Not applicable	Not applicable	126,149	4.8	NC_025744
<i>T. lupinaster</i>	<i>Trifolium</i>	<i>Lupinaster</i>	Not applicable	Not applicable	Not applicable	135,049	10.9	KJ788287
<i>T. repens</i>	<i>Trifolium</i>	<i>Trifolium</i>	Not applicable	Not applicable	Not applicable	144,763	19.7	NC_011828
<i>T. meduseum</i>	<i>Trifolium</i>	<i>Trichocephalum</i>	348,724	8.5	MT039390	142,595	19.5	NC_024166
<i>T. pratense</i>	<i>Trifolium</i>	<i>Trifolium</i>	301,823	6.6	MT039389	146,573	20.7	MT039393
<i>T. hybridum</i>	<i>Trifolium</i>	<i>Vesicastrum</i>	Not applicable	Not applicable	Not applicable	134,831	13.1	KJ788286
<i>T. semipilosum</i>	<i>Trifolium</i>	<i>Vesicastrum</i>	Not applicable	Not applicable	Not applicable	138,194	15.8	KJ788291
<i>T. repens</i>	<i>Trifolium</i>	<i>Trifolium</i>	Not applicable	Not applicable	Not applicable	132,120	10.7	NC_024166
<i>T. occidentale</i>	<i>Trifolium</i>	<i>Trifolium</i>	Not applicable	Not applicable	Not applicable	133,806	11.1	KJ788289
<i>T. semipilosum</i>	<i>Trifolium</i>	<i>Vesicastrum</i>	Not applicable	Not applicable	Not applicable	138,194	15.8	KJ788291
<i>T. repens</i>	<i>Trifolium</i>	<i>Trifoliastrium</i>	Not applicable	Not applicable	Not applicable	132,120	10.7	NC_024036
<i>T. occidentale</i>	<i>Trifolium</i>	<i>Trifoliastrium</i>	Not applicable	Not applicable	Not applicable	133,806	11.1	KJ788289

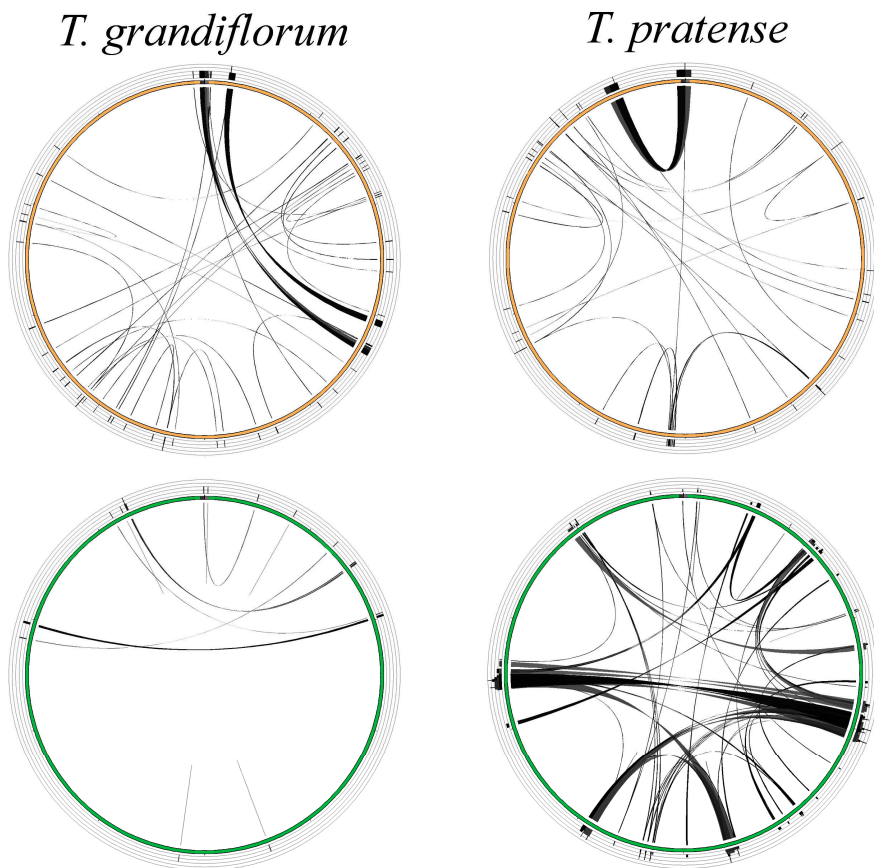


Figure 3. Distribution pattern of dispersed repeat sequences in circular representations of organelle genomes of two *Trifolium* species. Each ribbon represents a BLAST hit for a pair of dispersed repeats. Brown circles are mitogenomes and green circles are plastomes. Multiple hits in a single region are indicated by histograms in outer concentric rings. Data for repetitive sequences of all published organelle genomes of *Trifolium* is available in Table 2.

2.3. Intracellular Gene Transfer (IGT) in *Trifolium*

The extent of IGT among the three genomic compartments was analyzed in *T. pratense* by BLAST (Figure 4; Table 3). The amount of DNA shared between the two organelle genomes was very low (0.3 kb). The organelle genomes shared considerable DNA with the nuclear genome and GC content of shared DNA reflected the compartment of origin (45.8% for mitogenome and 35.1% for plastome). In general, BLAST hits between nuclear and organelle genomes were very short and had high sequence identity (Table 3).

BLAST (Figure 4; Table 3). The amount of DNA shared between the two organelle genomes was very low (0.3 kb). The organelle genomes shared considerable DNA with the nuclear genome and GC content of shared DNA reflected the compartment of origin (45.8% for mitogenome and 35.1% for plastome). In general, BLAST hits between nuclear and organelle genomes were very short and had high sequence identity (Table 3).

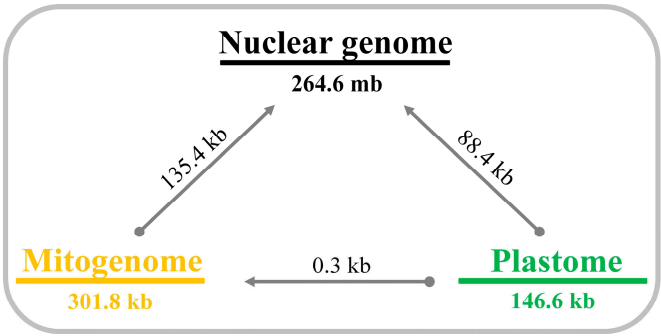


Figure 4. Shared DNA among three different genomic compartments in *Trifolium pratense*. Size of each genome is presented below the bar. Values on arrows represent the amount of shared DNA. Round end of arrow represents query sequence and pointed end represents subject sequence in BLAST analyses.

Table 3. BLAST hit statistics for shared DNA between nuclear and organelle genomes in *Trifolium pratense*.

Comparison	Number	Average Identity	GC (%)	Length (bp)				
				Min	Max	Mean	Median	
Comparison	Number	Average Identity	GC (%)	Min	Max	Mean	Median	
Mitochondria	1830	95.5	45.8	33	395	121.8	18.5	93.5
Nuclear vs. Mitochondria	1086	95.7	35.1	34	2027	144.2	118.5	
Nuclear vs. Plastid	1086	95.7	35.1	34	2027	144.2	118.5	

A long contiguous region (348.5 kb) was identified from chromosome 4 of *T. repens* (position: 72,476,623–72,825,180) that shared substantial DNA with the mitogenome of *Trifolium repens* (Fig. (Fig. 5)). This sequence had a high GC content (45.8%) compared to the entire chromosome 4 (33.2%).

2.4. Multiple Functional Transfers of Mitochondrial *rps1* in Papilionoideae

A phylogenetic analysis of nuclear and mitochondrial copies of *rps1* for papilionoid legumes was conducted (Figure 5). Nuclear genomes of two *Trifolium* species (*T. pratense* and *T. repens*) (Table S1) included multiple copies. Nuclear copies of *rps1* were placed in two separate positions, that included *L. bister* to the taxon in the tribes *Faboae* and *Trifolieae* and the second with four genera of the tribe *Trifolieae* (*Trigonella*, *Thilobium*, *Medicago* and *Trifolium*). Branch length for the nuclear copies of *rps1* was substantially longer than for the mitochondrial copies, indicating accelerated substitution rates. The trifolium was the phylogenetic branch leading to the tribe that was very short and the bootstrap value (BS = 42%) was low. In *Trifolium*, the mitochondrial *rps1* sequence formed a paraphyletic grade sister to a paraphyletic group of nuclear *rps1* (BS = 26%) group of nuclear *rps1* (BS = 96%).

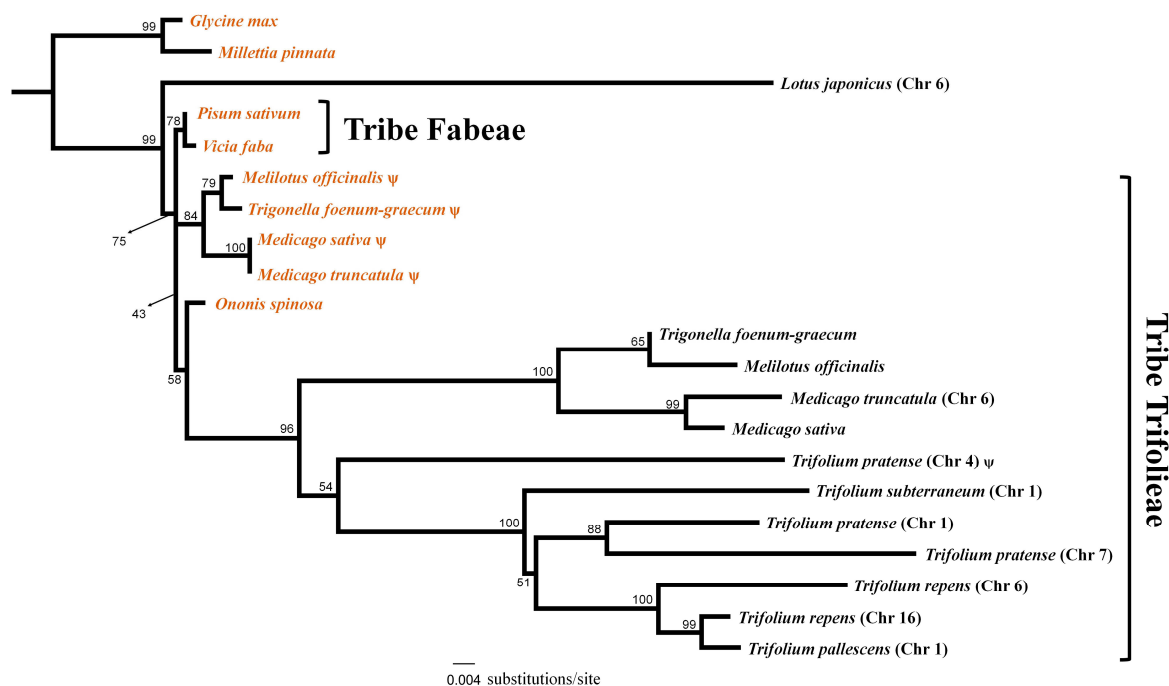


Figure 55. Maximum likelihood phylogeny of *rps1* gene. Bootstrap values are indicated at nodes. Sequences from mitochondrial genomes are indicated in brown. Sequences from nuclear genomes are indicated as black and chromosome (Chr) numbers are specified in the parentheses when available.

2.5. Fission of *ccmF* in Land Plants

To investigate the phylogenetic distribution of the fission of *ccmFn* and conservation of two ORFs *ccmFn1* and *ccmFn2* in *Trifolium*, mitochondrial *ccmF* sequences were assembled using available next-generation sequencing (NGS) reads (Table S2). The expanded taxon sampling confirmed the adjacency of the ORFs *ccmFn1* and *ccmFn2* and that the fission was restricted to *Trifolium*. All examined *Trifolium* species shared the *ccmFc* intron loss. Draft nuclear genome sequences of four species of *Trifolium* (*T. subterraneum*, *T. pratense*, *T. pallescens* and *T. repens*) were examined for intact copies of *ccmFn1* and *ccmFn2*. Fragments of sequences similar to *ccmFn1* and *ccmFn2* were identified and examined for intact copies of *ccmFn1* and *ccmFn2*. Fragments of sequences similar to *ccmFn1* and *T. subterraneum* and *T. pratense* but no intact copies were detected. However, intact copies both of *ccmFn1* and *ccmFn2* were identified in *T. subterraneum* and *T. pratense* but no intact copies were detected. However, intact copies both of *ccmFn1* and *ccmFn2* from *T. pallescens* (chromosome 4) and *T. repens* (chromosomes 4 and 9) were identified and were adjacent as in mitogenomes of *Trifolium*. Eleven *ccmFn* sequences (eight mitochondrial and three nuclear copies) were detected in *Trifolium* (Figure S4a). All nuclear copies were identical to *ccmFn* sequences (eight mitochondrial and three nuclear copies) were detected in *Trifolium* (Figure S4a). All nuclear copies were identical to their corresponding mitochondrial copy. Among mitochondrial copies, only three *Trifolium* species (*T. aureum*, *T. grandiflorum* and *T. pallescens*) showed unique sequence and the remaining sequences in the other five species were identical to each other in the coding region (Figure S4b). Fission of *ccmFc* was analyzed in three species of *Marchantia* and two other genera of the Marchantiales. Sequence alignment revealed that a single nucleotide deletion caused *ccmFc* fission in one species of *Marchantia*, *M. paleacea* (Figure S5). Examination of *ccmFn* fission in Brassicaceae included 17 taxa (Table S2). The *ccmF* genes were assembled from *Cleome violacea*, the sister family of Brassicaceae and two early diverging Brassicaceae genera (*Aethionema* and *Odontarrhena*). The fission of *ccmFn* was shared by all Brassicaceae except *Aethionema* and in all cases *ccmFn1* and *ccmFn2* were found in different loci. *Odontarrhena argentea* was the only member of the Brassicaceae that lost the *ccmFc* intron. The phylogenetic position of *ccmFn* fission and separation in Fabaceae and Brassicaceae (Table S2), were plotted on cladograms of each of family (Figure 6a,b). The location of the breakpoint of *ccmFn* fission was also compared among the three families Fabaceae, Brassicaceae and Amaryllidaceae (Figure 6c). The fission occurred in different locations in the gene within each family and occurred

and occurred in a more basal position in Brassicaceae than Fabaceae. The separation of *ccmFn1* and *ccmFn2* only occurred in Amaryllidaceae and Brassicaceae.

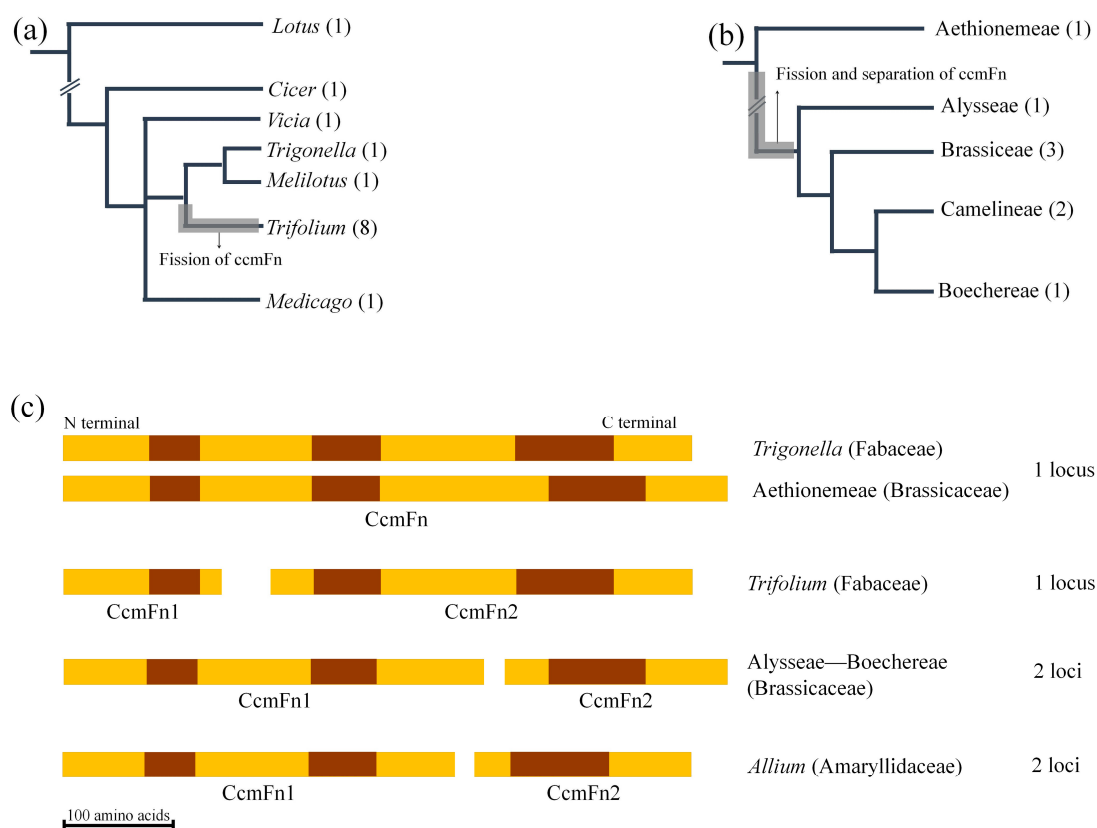


Figure 6. Fission events of *ccmFn* in angiosperms. (a) Examined genera of Fabaceae are plotted on cladogram from Ellison et al. [42]. Number of examined species is indicated in the parenthesis. The branch where the fission occurred is marked with grey bar. (b) Examined tribes of Brassicaceae are plotted on cladogram from Huang et al. [47]. Number of examined genera is indicated in the parenthesis. The branch where the fission and separation occurred is marked with grey bar. (c) *CcmFn* gene organization compared among various taxa of angiosperms. Dark brown box indicates conserved domains.

3. Discussion

3.1. Contrasting Evolutionary Trajectories of *Trifolium* Organelle Genomes

Trifolium mitogenomes (294,911 to 348,724 bp) (Table 1) are similar in size to the other Trifolieae genus *Medicago* (271,618 bp), which has the smallest currently sequenced papilionoid mitogenome [8]. Mitogenomes of *Trifolium* have relatively little repetitive DNA (6.6–8.6%) (Table 2) compared to mitogenomes of other Papilionoideae species (2.9–60.6%) [8]. This low repeat content in the mitogenome is in contrast to the plastome of some *Trifolium* species. The acquisition of numerous, novel repeat sequences and drastic rearrangement in the plastome of *T. subterraneum* and related species has been reported [31,33,35]. Increased taxon sampling by Sveinsson and Cronk [34] revealed that that plastome expansion is shared by five sections, referred to as the “refractory clade” in subgenus *Trifolium* (*Lupinaster*, *Trichocephalum*, *Trifolium*, *Vesicastrum* and *Trifolium*). The distinct evolutionary trajectory of organelle genomes in the genus is particularly evident in *T. pratense*, which has the lowest percentage of repetitive DNA in the mitogenome and the highest in the plastome as well as the most highly rearranged structure (Table 2 and Figure 3). In plant mitogenomes, genome expansions and rearrangements may be a consequence of error-prone DNA repair mechanisms such as nonhomologous end-joining or

such as nonhomologous end-joining or break-induced-replication [48–50]. In Geraniaceae, a correlation between nonsynonymous substitution rates for DNA replication, recombination and repair (DNA-RRR) genes and plastome complexity was reported [51]. The plastome-specific increase in repeat complexity in the *Trifolium* refractory clade may be the result of disruption of ‘plastid specific’ DNA-RRR-protein genes, some of which are targeted to both mitochondria and plastids [7]. More comprehensive taxon sampling that includes data from all three plant genomic compartments of *Trifolium* is required to test this hypothesis.

3.2. Multiple Functional Transfers of the Mitochondrial *rps1* Gene to the Nucleus in Papilionoideae

An earlier investigation reported the functional transfer of mitochondrial *rps1* to the nucleus in three genera of Trifolieae (*Trigonella*, *Melilotus* and *Medicago*) [39]. In the current study, the complete deletion of *rps1* gene from mitogenomes of four *Trifolium* species was detected (Figure S1), which is shared by the distantly related genus *Lotus*, a member of the tribe Loteae (Figure S2). There are two possible explanations for the phylogenetic distribution of the loss/transfer. The loss of mitochondrial *rps1* could be due to a single IGT in a common ancestor with differential resolution in descendant lineages, that is, acquisition of functional signals (or not) to stabilize transfer. Alternatively, there may have been independent functional transfers from an ancestor in each of the two unrelated lineages. To examine these alternatives, a maximum likelihood (ML) analysis was conducted using expanded taxon sampling of nuclear and mitochondrial *rps1* sequences. The resulting tree (Figure 5) included some long branches, which may be affected by the well-known phenomenon of long-branch attraction [52]. Nuclear *rps1* from *Lotus* and Trifolieae species were split into two independent clades, with intact and pseudogenized mitochondrial *rps1* placed between them. This pattern supports the explanation that functional transfers of *rps1* occurred at least two times in Papilionoideae, once in *Lotus* and a separate event in the ancestor of the Trifolieae clade that includes *Trigonella*, *Melilotus*, *Medicago* and *Trifolium*. The timing of the functional transfer of *rps1* in Trifolieae would likely be after the divergence of *Ononis* (Figure 5), which only has a mitochondrial copy [39].

Despite the putative functional replacement by nuclear *rps1*, the mitochondrial *rps1* in three genera (*Trigonella*, *Melilotus* and *Medicago*) was retained with limited sequence divergence (Figure 5), whereas it is completely and precisely deleted in *Trifolium* (Figure S1). Coding regions of plant mitogenomes are conserved by an accurate long homology-based repair mechanism, while non-coding regions are not conserved and are repaired by error-prone mechanisms [50]. Differential selection on mitogenomic molecules, which reduces harmful mutations on coding regions after double strand breaks (DSBs), was proposed to explain this [48,49]. Pseudogenized copies of mitochondrial *rps1* in the three genera *Trigonella*, *Melilotus* and *Medicago* are located adjacent to *nad5* exon1 (ca. 200 bp apart) [39]. Mutations in 5' region of *nad5* exon1 that do not disturb transcription or translation of the functional gene and only affect pseudogenized *rps1* can be inherited by selection after DSBs. So, the adjacent location of mitochondrial *rps1* to *nad5* exon1 may enable retention of high sequence identity after functional replacement by sharing the benefit of accurate repair. a similar situation is known for the *rps14* pseudogene that is adjacent to *rpl5* in grasses [53]. Conservation of non-coding regions adjacent to coding regions is also present in mitogenome-wide sequence divergence comparisons across Fabaceae [8].

3.3. Shared DNA Among Genomes of *Trifolium*

Comparative analyses of the three genomic compartments (nuclear, mitochondrial and plastid) in *T. pratense* revealed a substantial amount of shared DNA between nuclear and organelle genomes, most of which was short fragments (Figure 4, Table 3). The shared DNAs between nuclear and mitochondrial genome was 135.4 kb (Figure 4) and had GC content more similar to those of mitogenomes (Tables 1 and 3) suggesting that most IGT was unidirectional (i.e., mitochondrion to nucleus) and the nuclear genome of *T. pratense* includes numerous NUMTs. These NUMTs may integrate into the

nuclear genome of *T. pratense* as short fragments. Alternatively, these short fragments may be the consequence of post-IGT mutational decay and rearrangement of longer NUMT sequences [54].

The discovery of a long stretch of NUMTs (spanning 348.5 kb; GC: 44.3%) in chromosome 4 of *T. repens* (Figure S3) supports a recent genomic scale IGT event. This type of large IGT was identified in *Arabidopsis thaliana* (Brassicaceae) in which ~270 kb of 367 kb mitogenome transferred to the nucleus [55] and covers an ~620 kb region of the nuclear genome [56]. To estimate the amount of NUMTs in *T. repens*, a mitogenome sequence from the same DNA source (white clover cv 'Crau' derivative) [46,57] is necessary. Large NUMTs were reported for animal nuclear genomes (little brown bat and fugu), however, these were later shown represent artifacts of genome assembly [58,59]. The nuclear genomes of *Trifolium* species are drafts with many gaps [43–46]. Verification of long putative NUMTs in *Trifolium* is needed to confirm genomic scale IGT events from the mitochondrial to nuclear genome.

3.4. Multiple Fissions of *ccmF* in Land Plants and a Novel Event in *Trifolium*

The first fission of mitochondrial *ccmF* dates back to the early evolution of land plants and split the gene into N-terminal (*ccmFn*) and C-terminal (*ccmFc*) coding regions [60]. In Marchantiales, the ORFs are closely adjacent (Figure S5). The mitogenome study of *Marchantia paleacea* (misidentified as *M. polymorpha* [61]) from the early 1990s [22] reported a fission of *ccmFc* (i.e., *ccmFc1* and *ccmFc2*) due to a single nucleotide deletion. This fission event was accepted in several subsequent papers [3,21,60], however, mitogenome sequences of two other *Marchantia* species (*M. inflexa* and *M. polymorpha* subsp. *ruderalis*) did not show the single nucleotide deletion, consistent with the other two available mitogenomes of Marchantiales (Figure S5). The initial report of a *ccmFc* fission in *Marchantia* should be re-examined to determine if it is specific to *M. paleacea* or the result of sequencing error.

In angiosperms, two independent fissions of *ccmFn* have been reported in *Allium* (Amaryllidaceae) [25] and Brassicaceae [24,62]. In both cases, *ccmFn1* and *ccmFn2* are distant from each other in the mitogenome and they share a similar breakpoint for the fission (Figure 6). The phylogenetic distribution of the fission in Amaryllidaceae was investigated by polymerase chain reaction using four genera in the family (*Narcissus*, *Tulbaghia*, *Ipheion* and *Allium*) and revealed that the separation of the two sequences is restricted to *Allium* [25]. However, the status of the other three genera without separation of *ccmFn* sequences does not necessarily guarantee that the gene is not split because there are cases of gene fission where the two new genes occupy a single locus, for example, fission of *ccmF* (into *ccmFn* and *ccmFc*) in Marchantiales (Figure S5) and *ccmFn* (into *ccmFn1* and *ccmFn2*) in *Trifolium* (Figure 2). The distribution and status of *ccmFn* fission in Amaryllidaceae needs further investigation including broad taxon sampling as well as confirmation with additional sequencing.

In Brassicaceae, it was argued that the fission is shared by all members of the family because it is present in five complete or draft mitochondrial genomes covering the earliest diverging genus (*Aethionema*) and other core genera (*Arabidopsis*, *Brassica*, *Raphanus*), whereas the mitogenome of the sister family Cleomaceae does not have the fission [62]. Further investigation, including additional published mitogenomes and assembled mitochondrial contigs for *ccmF* genes (Table S2), indicates that three species of *Aethionema* do not have the fission of *ccmFn* (Figure 6b). This discrepancy could be due an assembly error since the *Aethionema* data in the previous study was a draft mitogenome [62]. Whatever was the cause of discrepancy, it is clear that the fission of *ccmFn* is shared by many but not all Brassicaceae. The fission occurred after the divergence of *Aethionema* (Figure 6b); however, it is unknown if there was an intermediate stage that had experienced the fission but not physical separation of the *ccmFn1* and *ccmFn2*.

The independent fission of *ccmFn* in *Trifolium* represents a novel event. The fission was caused by a deletion of 59 bp resulting in a frame shift and premature stop codon (Figure 2). An alternative outcome of this deletion may be pseudogenization of the *ccmFn*. Mutational decay and deletion of pseudogenized mitochondrial genes can be delayed by proximity to functional genes (e.g., *rps1* in some Trifolieae genera and *rps14* in grasses, see Section 3.2). However, the gene that is consistently adjacent to *ccmFn* (*ccmFn1* and *ccmFn2*) is *ccmC*, which is ca. 8kb away from *ccmFn* in the four *Trifolium* species

(Figure 1). Moreover, the expanded *ccmFn* sequence sampling confirms that the two ORFs (*ccmFn1* and *ccmFn2*) are conserved in eight *Trifolium* species with only a limited amount of sequence variation in coding regions (Figure S4). The fission break point in *Trifolium* is different from other angiosperms that express cytochrome c maturation protein from two ORFs, yet the conserved domains of the product remain intact (Figure 6c). Hence, the two ORFs of *ccmFn* are regarded as functional. The fission occurred after the divergence of genera *Trigonella* and *Melilotus* in the Trifolieae. The conserved adjacency of the two ORFs (*ccmFn1* and *ccmFn2*) may represent an early stage of the fission as in *ccmFn* and *ccmFc* in Marchantiales (Figure S5).

The fission of *ccmFn* in *Trifolium* leads to another question: is this event related to “intercompartmental piecewise gene transfer” [21]? To explore this question, we searched for ORFs of *ccmFn* in draft nuclear genomes of four *Trifolium* species (*T. subterraneum*, *T. pratense*, *T. pallescens* and *T. repens*). Both *T. pallescens* and *T. repens* (Figure S4) contained the *ccmFn* NUMTs however these were not restricted to a single ORF but included a locus covering both ORFs (*ccmFn1* and *ccmFn2*) and their flanking regions. The NUMTs were identical to their counterpart in mitogenome suggesting that the transfer was a recent event (or artifact in nuclear genome assembly, see discussion Section 3.3). Furthermore, there was no post-IGT sequence modification to suggest a functional transfer. Evidence did not support a relationship between fission of the mitochondrial gene *ccmFn* and piecewise or functional transfer in *Trifolium* species.

4. Materials and Methods

4.1. Assembly of *Trifolium* Mitogenomes

Four species of *Trifolium* from the two subgenera *Chronosemium* (*T. aureum* and *T. grandiflorum*) and *Trifolium* (*T. meduseum* and *T. pratense*) were selected for mitogenome assembly. The 100 bp paired-end raw Illumina (San Diego, CA, US) reads (Table 1) for mitogenome assembly were from Sabir et al. [31]. Assembly and mapping were conducted in Geneious Prime (<https://www.geneious.com>) using Geneious assembler and mapper, respectively. To assemble mitogenomes, the methods in Choi et al. [8] were followed. First, raw reads from the plastome were excluded by mapping total raw reads to corresponding plastomes [*T. aureum* (NC_024035.1), *T. grandiflorum* (NC_024034.1), *T. meduseum* (NC_024166.1) and *T. pratense* (MT039393)]. De novo assembly was subsequently conducted for each with ~30 million plastome-filtered reads. Among the assembled contigs, mitochondrial contigs were selected by BLAST searches against reference Fabaceae mitogenome sequences at National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/genome/organelle/>) using BLASTN 2.8.0+ [63] with default options. Mitochondrial contigs were manually assembled as single chromosomes in Geneious. Finally, draft mitogenomes were refined by mapping total plastome-filtered reads.

4.2. Annotation and Genome Content Comparison of Mitogenomes

To compare gene and intron content of *Trifolium* mitogenomes with related taxa, five previously published mitogenomes were acquired—two from IRLC [*Vicia faba* (KC189947) and *Medicago truncatula* (NC_029641)], one from the robinoid clade [*Lotus japonicus* (NC_016743)], which is sister to IRLC; and two from millettoid *sensu lato* clade [*Millettia pinnata* (NC_016742)] and *Glycine max* (NC_020455)], which is sister to the hologalegina clade (robinoid + IRLC). Annotation of rRNAs, protein coding genes and introns was conducted based on a reference mitogenome of *Liriodendron tulipifera* (NC_021152) with a set of 41 conserved mitochondrial genes in Geseq [64]. Annotation for protein coding genes was manually corrected in Geneious to fit ORFs. The annotation for tRNAs was cross-checked by tRNAscan-SE v2.0 [65].

4.3. Completion of the *Trifolium Pratense* Plastome

Plastome drafts of *Trifolium pratense* were reported in two different studies [31,34] but these sequences contained a complex repeat structure. Since these previous assemblies were based on short insert size data only (400–800 bp), the *T. pratense* plastome was redone using sequences generated from one of the previous studies [31] as well as mapping data from mate-pair Illumina sequences (ERX946087) with long insert sizes (7 kb) [43]. The newly assembled plastome was annotated as described above but with MPI-MP chloroplast references in GeSeq [64].

4.4. Repeat Estimation in Organelle Genomes

Repeat content was estimated in four mitogenomes and 13 plastomes (Table 2). Tandem repeats were identified using Tandem Repeats Finder version 4.09 [66] with default options. Other repeats (larger than 30 bp) were analyzed by BLASTN [63] searches using each genome as both subject and query with a word size of 7 and an e-value of $1e^{-6}$ as described in Guo et al. [67]. All BLAST hits were retained. Sequence coordinate information for BLAST hits was transferred to each genome as an annotation in Geneious and overlapping regions between hits were excluded from the estimations for repetitive DNA content. The distribution of dispersed repeat sequences across the genomes was visualized by Circoletto [68].

4.5. Shared DNA among Different Genomic Compartments

Shared DNA was evaluated in *Trifolium pratense* because this is the only species examined with completed sequences from all three genomic compartments. The mitogenome (MT039389) and plastome (MT039393) in this study were utilized and the nuclear genome was available as a chromosome-scale reference draft (LT990601–LT990607) [43]. Shared DNA among the genomes was evaluated in MegaBLAST with a word size of 28 and an e-value of $1e^{-6}$. For nuclear and organelle genome comparisons, each organelle sequence was used as the query against a subject database comprising the nuclear genome. For the comparison of organelle genomes, the plastome was used as the query and the mitogenome was the subject. BLAST hits with sequence identity higher than 90% were retained. Overlapping regions between hits were excluded from the estimations of shared DNA.

To search for putative large-scale IGT (> 100 kb) events, shared DNA analysis was conducted as described above but in this case the largest mitogenome (*T. meduseum*) and other published nuclear genomes of *Trifolium* (Table S1) were utilized. BLAST hits between the mitogenome and a long stretch of the nuclear region of *T. repens* were visualized by Circoletto [68].

4.6. Investigation on Status of *rps1* in Nuclear and Mitochondrial Genome

Nuclear and mitochondrial sequences of *rps1* generated for a previous study [39] were acquired from NCBI. Nuclear *rps1* sequences for other species were searched by MegaBLAST using the options described above. Mitochondrial *rps1* of *Vicia faba* was used to query nuclear genomes of *Lotus japonicus*, *Medicago truncatula*, *Trifolium subterraneum*, *T. pratense*, *T. pallescens* and *T. repens* (Table S1). Mitochondrial *rps1* sequences were also extracted from mitogenomes of *Glycine max*, *Milletia pinnata*, *Vicia faba* and *Medicago truncatula*. All *rps1* sequences were aligned with MAFFT v.7.017 [69] using default options. Nucleotide substitution models were evaluated in jModelTest v.2.1.6 [70] by Akaike information criterion. ML analysis (GTR + G with 1000 bootstrap replications) was conducted using *G. max* and *M. pinnata* as outgroups in RAxML v.8 [71] in the CIPRES Science Gateway [72].

The status of mitochondrial *rps1* in *Trifolium* was tested by sequence alignment of the mitochondrial locus containing *rps1* and *nad5* exon1 in *M. truncatula* and the corresponding regions in four mitogenomes of *Trifolium*. Sequences were aligned in MAFFT [69] using default options followed by manual adjustments to minimize gaps and maximize apparent homologous regions.

4.7. Investigation of *ccmF* Fissions in Selected Land Plants

To investigate previously reported fission events of *ccmF* genes in land plants [22–25], all available sequences from published mitogenome sequences related to *Marchantia*, Brassicaceae, *Allium* (Amaryllidaceae) and Fabaceae were acquired. For *Marchantia*, published mitogenomes [22,73,74] of three species (*M. inflexa*, *M. polymorpha* subsp. *ruderalis* and *M. paleacea*) were examined: two from NCBI [*M. polymorpha* subsp. *ruderalis* (NC_037508.1) and *M. paleacea* (NC_001660.1)] and *M. inflexa*, which was downloaded from FigShare (https://figshare.com/articles/Marchantia_inflexa_mitochondrion_and_chloroplast_genomes/6639209/1). Two mitogenomes [*Dumortiera hirsuta* (NC_042873) and *Riccia fluitans* (NC_043906)], which are closely related to *Marchantia* in Marchantiales [75,76], were also included. For Amaryllidaceae, a single mitogenome [*Allium cepa* (NC_030100)] was available.

In addition to previously published and newly assembled mitogenomes, mitochondrial contigs were generated from available NGS reads for Brassicales and Fabaceae (Table S2). Raw sequences were mapped to reference *ccmF* sequences and the mapped reads were assembled in Geneious. The *ccmF* sequences of *Medicago truncatula* and *Batis maritima* were used as references for Fabaceae and Brassicales, respectively. Read depth of assembled *ccmF* genes (*ccmFn* and *ccmFc*) were compared to confirm that sequences originated from mitogenome rather than from other genomic compartments (i.e., nuclear and plastid genome). To search for nuclear copies of *ccmFn1* and *ccmFn2*, subject databases comprising four *Trifolium* nuclear genomes (Table S1) were queried with the mitochondrial *ccmFn* of *T. aureum* using MegaBLAST with default options. All sequences were aligned with MAFFT as described above. The status of *ccmFn* was plotted on cladograms from published phylogenetic studies of *Trifolium* [42] and Brassicaceae [47]. Conserved domains of *ccmFn* were detected using the Motif Scan of MyHits (http://myhits.isb-sib.ch/cgi-bin/motif_scan) [77,78].

5. Conclusions

The newly sequenced mitogenomes of *Trifolium* allowed comparative analyses of genome evolution for all three cellular compartments—mitochondrion, nucleus and plastid. Unlike many angiosperms, *Trifolium* lacks the highly repetitive genome organization of mitogenome. Some *Trifolium* plastomes has a much more complex organization and has accumulated more repeat contents than the mitogenome. a substantial amount of organellar DNA was detected in nuclear genomes of *Trifolium*, likely resulting from recent and nonfunctional IGT events. In addition, there has been an ancestral, functional transfer of mitochondrial *rps1* to the nuclear genome. a notable finding from the mitogenome of *Trifolium* was a novel gene fission of *ccmFn*. Analyses of *ccmF* genes in selected land plants provided further insights into the fission events. Although the current study is based on limited sampling of the three genomic compartments, our findings expand the understanding of how these genomes evolved in *Trifolium*. The underlying evolutionary and molecular mechanisms should be examined in future comparisons that incorporate broader taxonomic sampling for all three genomic compartments.

Supplementary Materials: Supplementary Materials can be found at <http://www.mdpi.com/1422-0067/21/6/1959/s1>. Figure S1. Nucleotide alignment showing deletion of mitochondrial *rps1* in *Trifolium* species. Figure S2. Gene and cis-spliced intron content across six Papilionoideae genera. Figure S3. Circoletto map showing similar sequences between mitogenome of *T. meduseum* (left arc) and a continuous region of nuclear genome (right arc) of *T. repens* (chromosome 4; NCBI accession: VCDJ01010667; position: 72,476,623–72,825,180). Figure S4. Sequence variation of *ccmFn* in *Trifolium* species. Figure S5. Alignment of the mitochondrial region containing *ccmFn* and *ccmFc* genes from five species of Marchantiales. Table S1. Information on nuclear genomes, used for comparative study. Table S2. List of taxa for *ccmF* analysis with information about sequence sources and status of the genes.

Author Contributions: Conceptualization, I.-S.C. and R.K.J.; Validation, R.K.J. and T.A.R.; Formal Analysis, I.-S.C.; Investigation, I.-S.C.; Resources, R.K.J. and T.A.R.; Data Curation, R.K.J. and T.A.R.; Writing—Original Draft Preparation, I.-S.C.; Writing—Review & Editing, I.-S.C., T.A.R. and R.K.J.; Visualization, I.-S.C.; Supervision, R.K.J.; Funding Acquisition, R.K.J. and T.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a National Science Foundation grant (DEB-1853024) to R.K.J. and T.A.R. and the S. F. Blake Centennial Professorship to R.K.J.

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

Abbreviations

IGT	Intracellular gene transfer
NUMT	Nuclear mitochondrial DNA sequences
NUPT	Nuclear plastid DNA sequences
IR	Inverted repeat
IRLC	Inverted repeat lacking clade
ORF	Open reading frame
NGS	Next-generation sequencing
DNA-RRR	DNA replication, recombination and repair
ML	Maximum likelihood
DSB	Double strand break
NCBI	National Center for Biotechnology Information

References

1. Jansen, R.K.; Ruhlman, T.A. Plastid Genomes of Seed Plants. In *Genomics of Chloroplasts and Mitochondria*; Bock, R., Knoop, V., Eds.; Springer: Dordrecht, The Netherlands, 2012; pp. 103–126.
2. Ruhlman, T.A.; Jansen, R.K. The plastid genomes of flowering plants. In *Chloroplast Biotechnology: Methods and Protocols*; Maliga, P., Ed.; Springer: New York, NY, USA, 2014; pp. 3–38.
3. Knoop, V. The mitochondrial DNA of land plants: Peculiarities in phylogenetic perspective. *Curr. Genet.* **2004**, *46*, 123–139. [[CrossRef](#)] [[PubMed](#)]
4. Wynn, E.L.; Christensen, A.C. Repeats of unusual size in plant mitochondrial genomes: Identification, incidence and evolution. *G3 Genes Genomes Genet.* **2019**, *9*, 549–559. [[CrossRef](#)] [[PubMed](#)]
5. 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](#)] [[PubMed](#)]
6. Kubo, T.; Mikami, T. Organization and variation of angiosperm mitochondrial genome. *Physiol. Plant.* **2007**, *129*, 6–13. [[CrossRef](#)]
7. Gualberto, J.M.; Newton, K.J. Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* **2017**, *68*, 225–252. [[CrossRef](#)] [[PubMed](#)]
8. Choi, I.-S.; Schwarz, E.N.; Ruhlman, T.A.; Khiyami, M.A.; Sabir, J.S.; Hajarrah, N.H.; Sabir, M.J.; Rabah, S.O.; Jansen, R.K. Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. *BMC Plant Biol.* **2019**, *19*, 448. [[CrossRef](#)] [[PubMed](#)]
9. Wolfe, K.H.; Li, W.-H.; Sharp, P.M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 9054–9058. [[CrossRef](#)]
10. Drouin, G.; Daoud, H.; Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* **2008**, *49*, 827–831. [[CrossRef](#)]
11. Selosse, M.-A.; Albert, B.; Godelle, B. Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends Ecol. Evol.* **2001**, *16*, 135–141. [[CrossRef](#)]
12. Lopez, J.V.; Yuhki, N.; Masuda, R.; Modi, W.; O'Brien, S.J. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* **1994**, *39*, 174–190.
13. Timmis, J.N.; Ayliffe, M.A.; Huang, C.Y.; Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **2004**, *5*, 123–135. [[CrossRef](#)] [[PubMed](#)]
14. Richardson, A.O.; Palmer, J.D. Horizontal gene transfer in plants. *J. Exp. Bot.* **2006**, *58*, 1–9. [[CrossRef](#)] [[PubMed](#)]
15. Bonen, L. Mitochondrial genes leave home. *New Phytol.* **2006**, *172*, 379–381. [[CrossRef](#)] [[PubMed](#)]
16. Huang, C.Y.; Grünheit, N.; Ahmadinejad, N.; Timmis, J.N.; Martin, W. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. *Plant Physiol.* **2005**, *138*, 1723–1733. [[CrossRef](#)]

17. Adams, K.L.; Song, K.; Roessler, P.G.; Nugent, J.M.; Doyle, J.L.; Doyle, J.J.; Palmer, J.D. Intracellular gene transfer in action: Dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13863–13868. [[CrossRef](#)]
18. Laroche, J.; Li, P.; Maggia, L.; Bousquet, J. Molecular evolution of angiosperm mitochondrial introns and exons. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 5722–5727. [[CrossRef](#)]
19. Adams, K.L.; Daley, D.O.; Qiu, Y.-L.; Whelan, J.; Palmer, J.D. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **2000**, *408*, 354–357. [[CrossRef](#)]
20. Adams, K.L.; Ong, H.C.; Palmer, J.D. Mitochondrial gene transfer in pieces: Fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. *Mol. Biol. Evol.* **2001**, *18*, 2289–2297. [[CrossRef](#)]
21. Szafranski, P. Intercompartmental piecewise gene transfer. *Genes* **2017**, *8*, 260. [[CrossRef](#)]
22. Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Kanegae, T.; Ogura, Y.; Kohchi, T. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. *J. Mol. Biol.* **1992**, *223*, 1–7. [[CrossRef](#)]
23. Handa, H.; Bonnard, G.R.; Grienberger, J.-M. The rapeseed mitochondrial gene encoding a homologue of the bacterial protein *Ccl1* is divided into two independently transcribed reading frames. *Mol. Gen. Genet.* **1996**, *252*, 292–302. [[CrossRef](#)]
24. 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](#)]
25. Kim, B.; Kim, K.; Yang, T.-J.; Kim, S. Completion of the mitochondrial genome sequence of onion (*Allium cepa* L.) containing the CMS-S male-sterile cytoplasm and identification of an independent event of the *ccmFN* gene split. *Curr. Genet.* **2016**, *62*, 873–885. [[CrossRef](#)]
26. Thöny-Meyer, L.; Fischer, F.; Künzler, P.; Ritz, D.; Hennecke, H. *Escherichia coli* genes required for cytochrome c maturation. *J. Bacteriol.* **1995**, *177*, 4321–4326. [[CrossRef](#)]
27. Giegé, P.; Grienberger, J.; Bonnard, G. Cytochrome c biogenesis in mitochondria. *Mitochondrion* **2008**, *8*, 61–73. [[CrossRef](#)]
28. LPWG, Legume Phylogeny Working Group. a new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* **2017**, *66*, 44–77. [[CrossRef](#)]
29. Lewis, G.P.; Schrire, B.; Mackinder, B.; Lock, M. *Legumes of the World*; Royal Botanic Gardens: Kew, UK, 2005.
30. Wojciechowski, M.F.; Sanderson, M.J.; Steele, K.P.; Liston, A. Molecular phylogeny of the “temperate herbaceous tribes” of papilionoid legumes: a supertree approach. In *Advances in Legume Systematics*; Herendeen, P., Bruneau, A., Eds.; Royal Botanic Gardens: Kew, UK, 2000; Volume 9, pp. 277–298.
31. Sabir, J.; Schwarz, E.; Ellison, N.; Zhang, J.; Baeshen, N.A.; Mutwakil, M.; Jansen, R.; Ruhlman, T. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnol. J.* **2014**, *12*, 743–754. [[CrossRef](#)]
32. Magee, A.M.; Aspinall, S.; Rice, D.W.; Cusack, B.P.; Sémon, M.; Perry, A.S.; Stefanović, S.; Milbourne, D.; Barth, S.; Palmer, J.D. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* **2010**, *20*, 1700–1710. [[CrossRef](#)]
33. Cai, Z.; Guisinger, M.; Kim, H.-G.; Ruck, E.; Blazier, J.C.; McMurtry, V.; Kuehl, J.V.; Boore, J.; Jansen, R.K. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J. Mol. Evol.* **2008**, *67*, 696–704. [[CrossRef](#)]
34. Sveinsson, S.; Cronk, Q. Evolutionary origin of highly repetitive plastid genomes within the clover genus (*Trifolium*). *BMC Evol. Biol.* **2014**, *14*, 228. [[CrossRef](#)]
35. Milligan, B.G.; Hampton, J.N.; Palmer, J.D. Dispersed repeats and structural reorganization in subclover chloroplast DNA. *Mol. Biol. Evol.* **1989**, *6*, 355–368.
36. Choi, I.S.; Jansen, R.; Ruhlman, T. Lost and Found: Return of the inverted repeat in the legume clade defined by its absence. *Genome Biol. Evol.* **2019**, *11*, 1321–1333. [[CrossRef](#)]
37. Negruk, V. Mitochondrial genome sequence of the legume *Vicia faba*. *Front. Plant Sci.* **2013**, *4*, 128. [[CrossRef](#)]
38. Bi, C.; Wang, X.; Xu, Y.; Wei, S.; Shi, Y.; Dai, X.; Yin, T.; Ye, N. The complete mitochondrial genome of *Medicago truncatula*. *Mitochondrial DNA B* **2016**, *1*, 122–123. [[CrossRef](#)]
39. Hazle, T.; Bonen, L. Status of genes encoding the mitochondrial S1 ribosomal protein in closely-related legumes. *Gene* **2007**, *405*, 108–116. [[CrossRef](#)]

40. Kazakoff, S.H.; Imelfort, M.; Edwards, D.; Koehorst, J.; Biswas, B.; Batley, J.; Scott, P.T.; Gresshoff, P.M. Capturing the biofuel wellhead and powerhouse: The chloroplast and mitochondrial genomes of the leguminous feedstock tree *Pongamia pinnata*. *PLoS ONE* **2012**, *7*, e51687. [\[CrossRef\]](#)
41. Lock, J. Trifolieae. In *Legumes of the World*; Lewis, G.P., Schrire, B., Mackinder, B., Lock, M., Eds.; Royal Botanic Gardens: Kew, UK, 2005; pp. 499–504.
42. Ellison, N.W.; Liston, A.; Steiner, J.J.; Williams, W.M.; Taylor, N.L. Molecular phylogenetics of the clover genus (*Trifolium*—Leguminosae). *Mol. Phylogenet. Evol.* **2006**, *39*, 688–705. [\[CrossRef\]](#)
43. De Vega, J.J.; Ayling, S.; Hegarty, M.; Kudrna, D.; Goicoechea, J.L.; Ergon, Å.; Rognli, O.A.; Jones, C.; Swain, M.; Geurts, R. Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. *Sci. Rep.* **2015**, *5*, 17394. [\[CrossRef\]](#)
44. Hirakawa, H.; Kaur, P.; Shirasawa, K.; Nichols, P.; Nagano, S.; Appels, R.; Erskine, W.; Isobe, S.N. Draft genome sequence of subterranean clover, a reference for genus *Trifolium*. *Sci. Rep.* **2016**, *6*, 30358. [\[CrossRef\]](#)
45. Dudchenko, O.; Pham, M.; Lui, C.; Batra, S.S.; Hoeger, M.; Nyquist, S.K.; Durand, N.C.; Shamim, M.S.; Machol, I.; Erskine, W. Hi-C yields chromosome-length scaffolds for a legume genome, *Trifolium subterraneum*. *bioRxiv* **2018**, 473553. [\[CrossRef\]](#)
46. Griffiths, A.G.; Moraga, R.; Tausen, M.; Gupta, V.; Bilton, T.P.; Campbell, M.A.; Ashby, R.; Nagy, I.; Khan, A.; Larking, A. Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. *Plant Cell* **2019**, *31*, 1466–1487. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Huang, C.-H.; Sun, R.; Hu, Y.; Zeng, L.; Zhang, N.; Cai, L.; Zhang, Q.; Koch, M.A.; Al-Shehbaz, I.; Edger, P.P. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol. Biol. Evol.* **2015**, *33*, 394–412. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Christensen, A.C. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. *Genome Biol. Evol.* **2013**, *5*, 1079–1086. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Christensen, A.C. Genes and junk in plant mitochondria—Repair mechanisms and selection. *Genome Biol. Evol.* **2014**, *6*, 1448–1453. [\[CrossRef\]](#)
50. Christensen, A.C. Mitochondrial DNA repair and genome evolution. In *Annual Plant Reviews*; Logan, D.C., Ed.; Wiley-Blackwell: New York, NY, USA, 2018; Volume 50, pp. 11–32.
51. Zhang, J.; Ruhlman, T.A.; Sabir, J.S.; Blazier, J.C.; Weng, M.-L.; Park, S.; Jansen, R.K. Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. *Genome Biol. Evol.* **2016**, *8*, 622–634. [\[CrossRef\]](#)
52. Bergsten, J. a review of long-branch attraction. *Cladistics* **2005**, *21*, 163–193. [\[CrossRef\]](#)
53. 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. [\[CrossRef\]](#)
54. Noutsos, C.; Richly, E.; Leister, D. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Res.* **2005**, *15*, 616–628. [\[CrossRef\]](#)
55. Lin, X.; Kaul, S.; Rounsley, S.; Shea, T.P.; Benito, M.-I.; Town, C.D.; Fujii, C.Y.; Mason, T.; Bowman, C.L.; Barnstead, M. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **1999**, *402*, 761–768. [\[CrossRef\]](#)
56. Stupar, R.M.; Lilly, J.W.; Town, C.D.; Cheng, Z.; Kaul, S.; Buell, C.R.; Jiang, J. Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: Implication of potential sequencing errors caused by large-unit repeats. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5099–5103. [\[CrossRef\]](#)
57. Cousins, G.; Woodfield, D. Effect of inbreeding on growth of white clover. In *Proceedings of the 13th Australasian Plant Breeding Conference*, Christchurch, New Zealand, 18–21 April 2006; pp. 568–572.
58. Venkatesh, B.; Dandona, N.; Brenner, S. Fugu genome does not contain mitochondrial pseudogenes. *Genomics* **2006**, *87*, 307–310. [\[CrossRef\]](#)
59. Shi, H.; Xing, Y.; Mao, X. The little brown bat nuclear genome contains an entire mitochondrial genome: Real or artifact? *Gene* **2017**, *629*, 64–67. [\[CrossRef\]](#)
60. Rayapuram, N.; Hagenmuller, J.; Grienenberger, J.M.; Bonnard, G.; Giegé, P. The three mitochondrial encoded CcmF proteins form a complex that interacts with CCMH and c-type apocytochromes in *Arabidopsis*. *J. Biol. Chem.* **2008**, *283*, 25200–25208. [\[CrossRef\]](#)
61. Kisiel, K.; Miwa, H.; Odrzykoski, I.J. Taxonomic identification of chloroplast genome of *Marchantia polymorpha* using DNA barcode sequences. In *Proceedings of the Fourth International Barcode of Life Conference*, Adelaide, Australia, 28 November–3 December 2011.

62. Grewe, F.; Edger, P.P.; Keren, I.; Sultan, L.; Pires, J.C.; Ostersetzer-Biran, O.; Mower, J.P. Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. *Mitochondrion* **2014**, *19*, 135–143. [\[CrossRef\]](#)
63. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [\[CrossRef\]](#)
64. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [\[CrossRef\]](#)
65. Lowe, T.M.; Chan, P.P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [\[CrossRef\]](#)
66. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [\[CrossRef\]](#)
67. Guo, W.; Grewe, F.; Fan, W.; Young, G.J.; Knoop, V.; Palmer, J.D.; Mower, J.P. *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol. Biol. Evol.* **2016**, *33*, 1448–1460. [\[CrossRef\]](#)
68. Darzentas, N. Circoletto: Visualizing sequence similarity with Circos. *Bioinformatics* **2010**, *26*, 2620–2621. [\[CrossRef\]](#)
69. Katoh, K.; Misawa, K.; Kuma, K.i.; Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [\[CrossRef\]](#)
70. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [\[CrossRef\]](#)
71. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [\[CrossRef\]](#)
72. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; pp. 1–8.
73. Bowman, J.L.; Kohchi, T.; Yamato, K.T.; Jenkins, J.; Shu, S.; Ishizaki, K.; Yamaoka, S.; Nishihama, R.; Nakamura, Y.; Berger, F. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **2017**, *171*, 287–304. [\[CrossRef\]](#)
74. Marks, R.A.; Smith, J.J.; Cronk, Q.; Grassa, C.J.; McLetchie, D.N. Genome of the tropical plant *Marchantia inflexa*: Implications for sex chromosome evolution and dehydration tolerance. *Sci. Rep.* **2019**, *9*, 8722. [\[CrossRef\]](#)
75. Kwon, W.; Kim, Y.; Park, J. The complete chloroplast genome sequence of *Dumortiera hirsuta* (Sw.) Nees (Marchantiophyta, Dumortieraceae). *Mitochondrial DNA B* **2019**, *4*, 318–319. [\[CrossRef\]](#)
76. Myszczyński, K.; Ślipiko, M.; Sawicki, J. Potential of transcript editing across mitogenomes of early land plants shows novel and familiar trends. *Int. J. Mol. Sci.* **2019**, *20*, 2963. [\[CrossRef\]](#)
77. Pagni, M.; Ioannidis, V.; Cerutti, L.; Zahn-Zabal, M.; Jongeneel, C.V.; Hau, J.; Martin, O.; Kuznetsov, D.; Falquet, L. MyHits: Improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res.* **2007**, *35*, W433–W437. [\[CrossRef\]](#)
78. Sigrist, C.J.; Cerutti, L.; De Castro, E.; Langendijk-Genevaux, P.S.; Bulliard, V.; Bairoch, A.; Hulo, N. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.* **2009**, *38*, D161–D166. [\[CrossRef\]](#)

