

Article title: Dynamic changes in the plastid and mitochondrial genomes of the angiosperm *Corydalis pauciovulata* (Papaveraceae)

Authors: Seongjun Park, Boram An, SeonJoo Park

Additional material

Figure S1. Sequencing coverage of the *Corydalis pauciovulata* plastome and mitogenome.

Figure S2. Structural alignments of the *Corydalis pauciovulata* plastome with outgroups.

Figure S3. Illustration of corrected ONT reads mapped against the assembled *Corydalis* mitogenome.

Figure S4. Mitochondrial maps surrounding the *orf146*, *orf244*, and *orf457ab*.

Figure S5. BUSCO assessment of the three transcriptomes.

Figure S6. Characterization of the nuclear-encoded acetyl-CoA carboxylase beta subunit (*ACCD*) genes of *Corydalis pauciovulata*.

Figure S7. Characterization of the nuclear-encoded *ACC*, *RPL20*, *RPL23*, and *RPS16* genes of *Corydalis pauciovulata*.

Figure S8. Genome completeness size and estimation.

Figure S9. Variation in sequence divergence among in *Corydalis pauciovulata* and *Nelumbo nucifera* plastid and mitochondrial protein-coding genes.

Table S1. List of annotated genes in the *Corydalis pauciovulata* plastome.

Table S2. Blast result of plastid-derived DNA segments in the *Corydalis pauciovulata* mitogenome.

Table S3. Predicted RNA editing in 38 protein-coding genes for the *Corydalis pauciovulata* mitogenome.

Table S4. RNA editing validation of 21 genes using transcriptome data.

Table S5. Blast results of ORFs (>150 bp) in the *Corydalis pauciovulata* mitogenome.

Table S6. Potential chimeric ORFs.

Table S7. Information on the nuclear-encoded NDH complex genes.

Table S8. Information on the nuclear-encoded DNA-RRR genes.

Figure S2. Structural alignments of the *Corydalis pauciovulata* plastome with outgroups. The colored blocks represent collinear sequence blocks shared by all plastomes. Blocks drawn below the horizontal line indicate sequences found in an inverted orientation. Individual genes and strandedness are represented below the *Liriodendron* genome block. Only one copy of the inverted repeat (IR) is shown for each plastome and pink boxes below each plastome block indicate its IR. Red boxes correspond to the purple box (second inversion) on Figure 3.

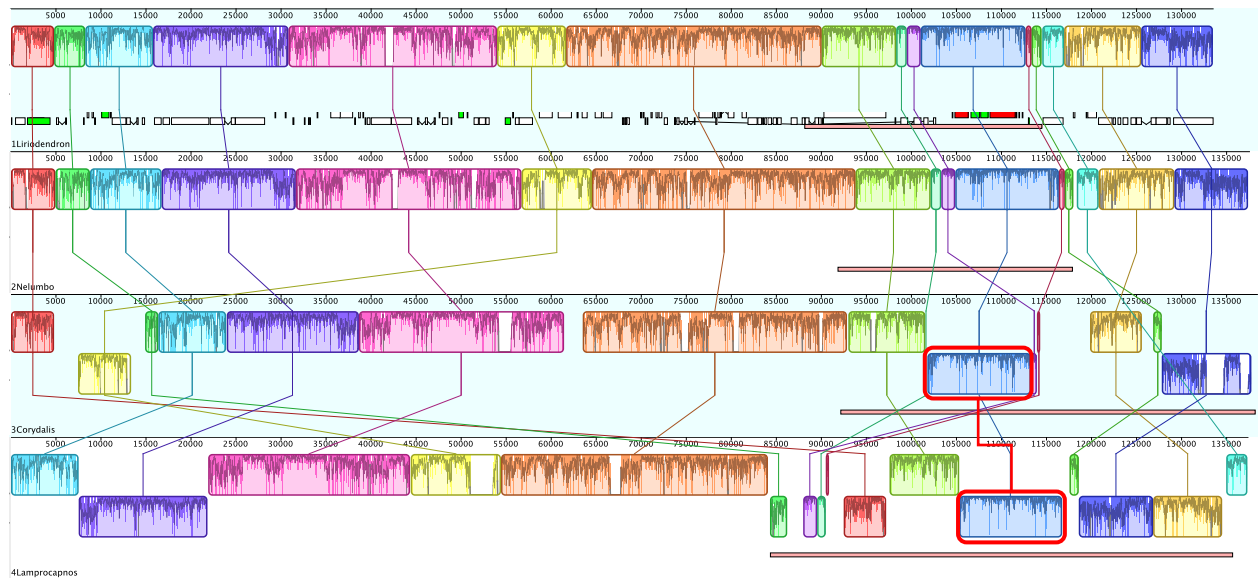


Figure S3. Illustration of corrected ONT reads mapped against the assembled *Corydalis* mitogenome. Blue lines indicate that those regions are identical to the parts of the mitogenome. Red lines indicate regions of ONT reads not aligned to the mitogenome, which are alternative genomic regions.

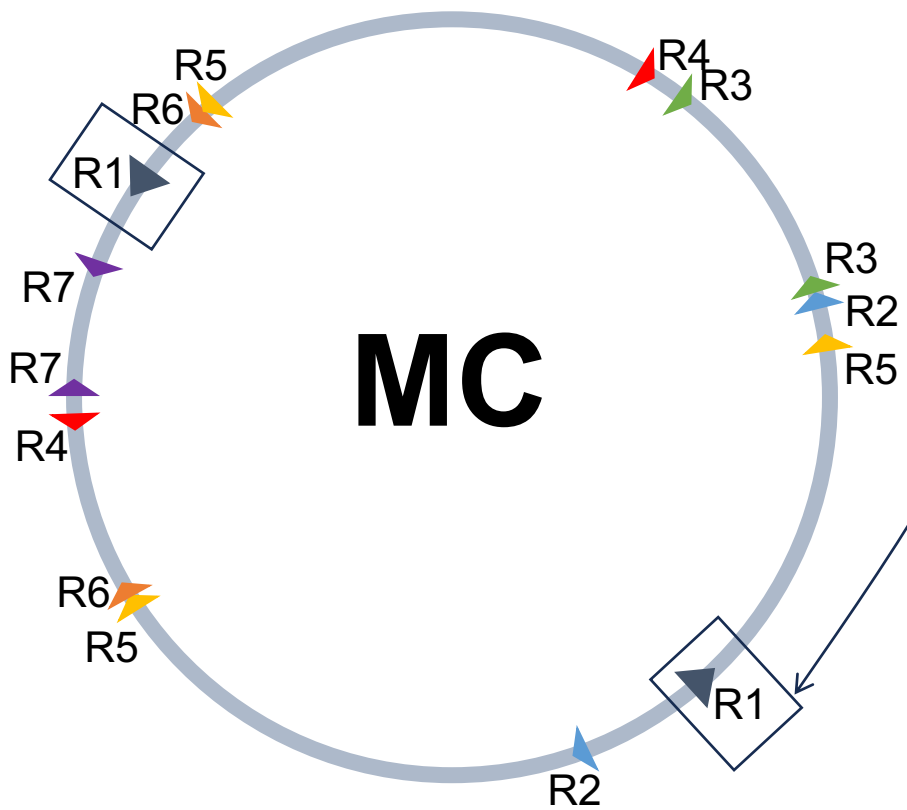
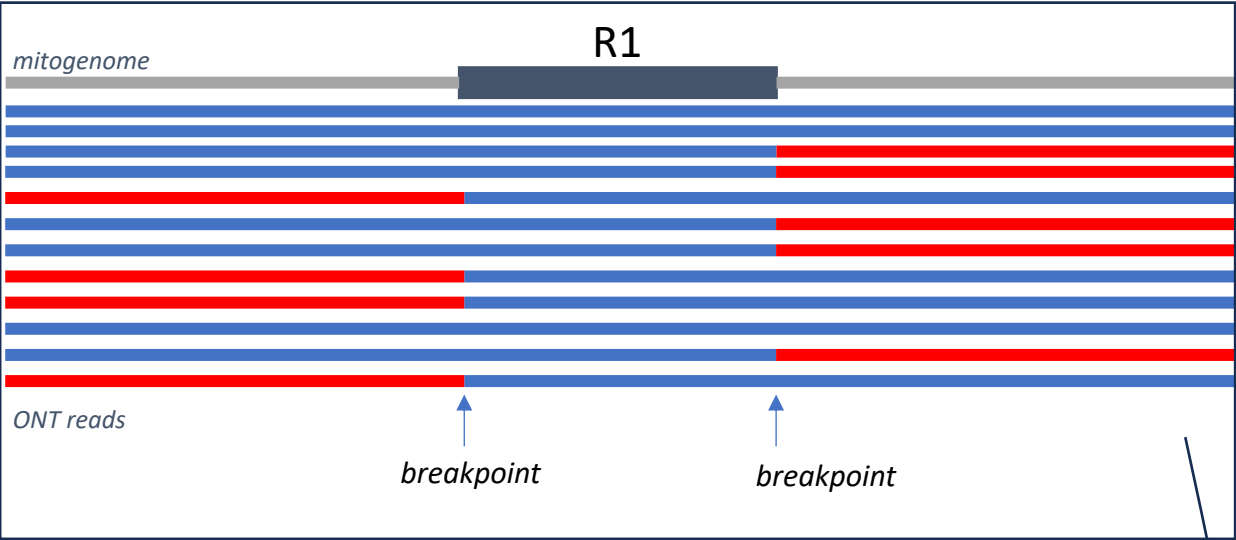


Figure S4. Mitochondrial maps surrounding the *orf146*, *orf244*, and *orf457ab*. Black annotation indicates exons and yellow annotation indicates coding sequence. Green annotation indicates gene. Gray annotation indicates each fragment of the genes. Ocher annotation indicates the repeats (R1, R5, and R6).

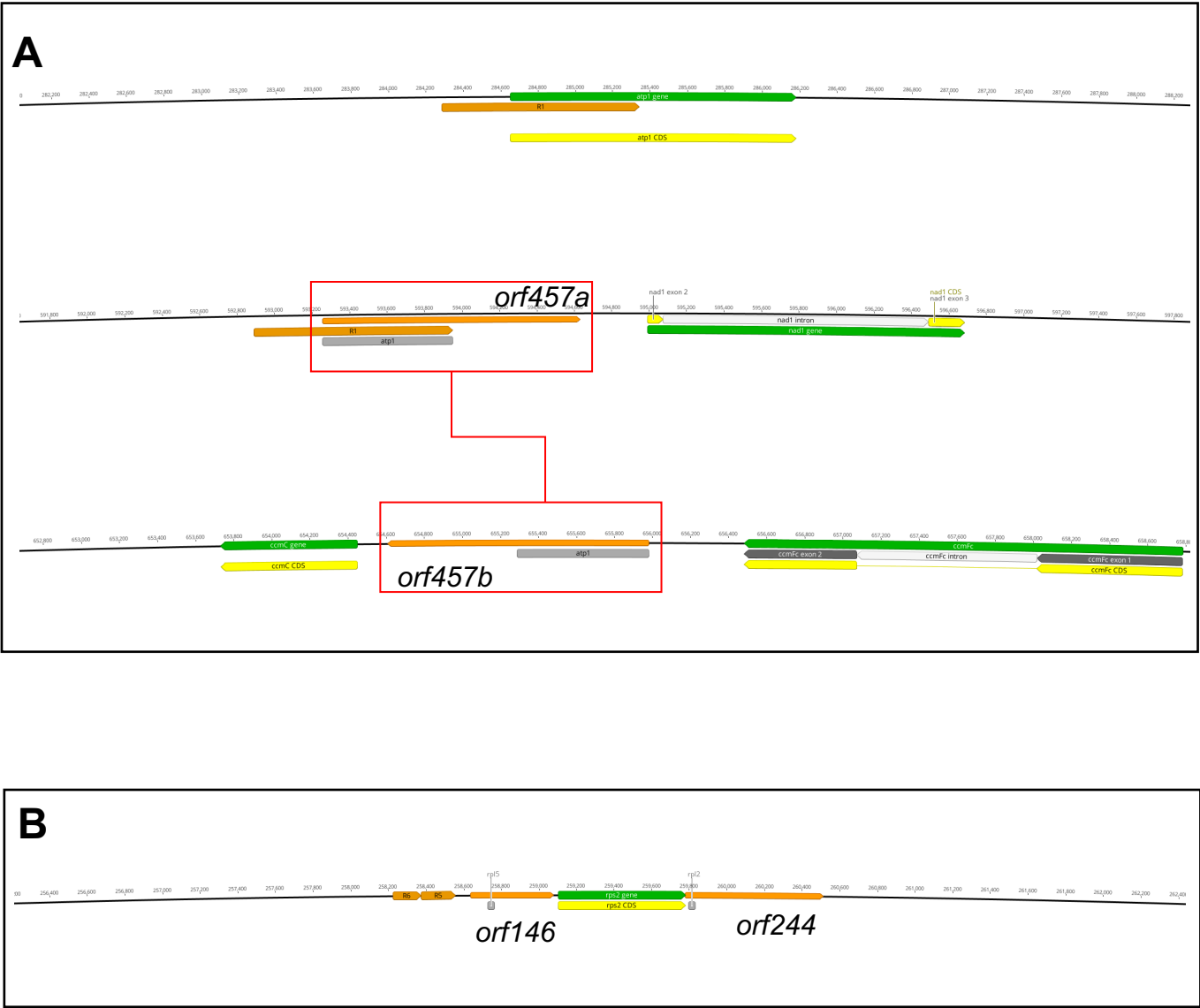
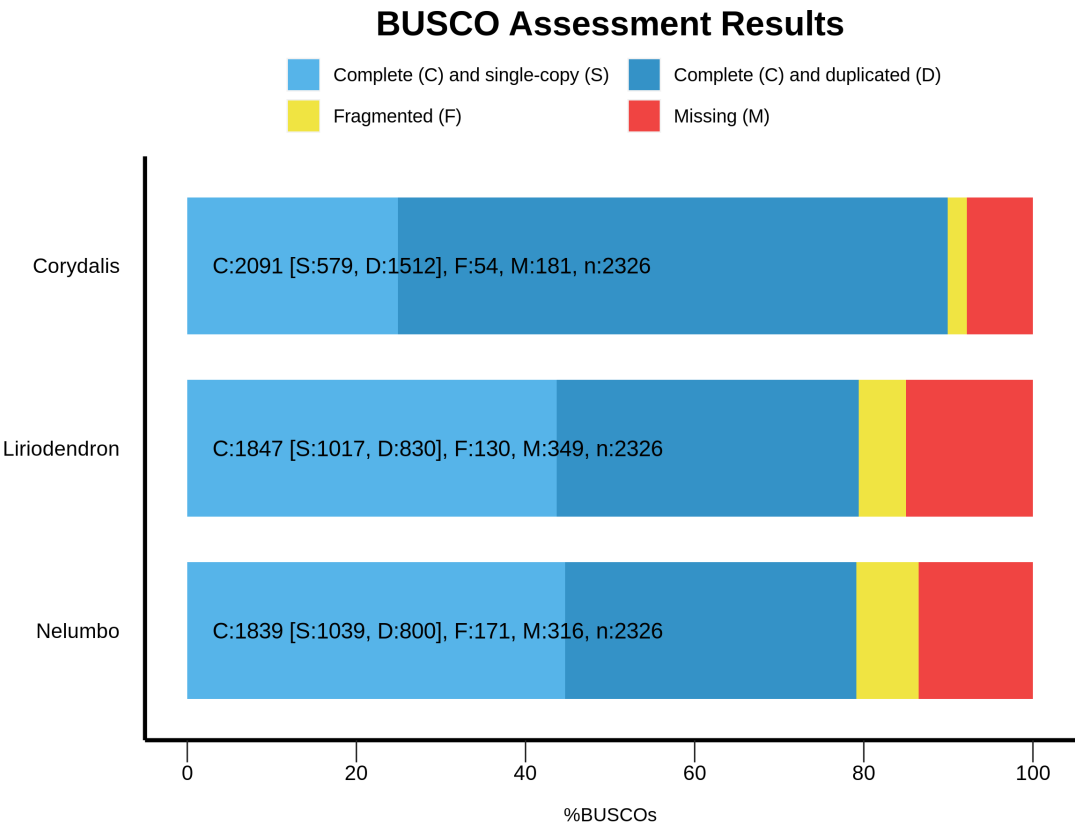
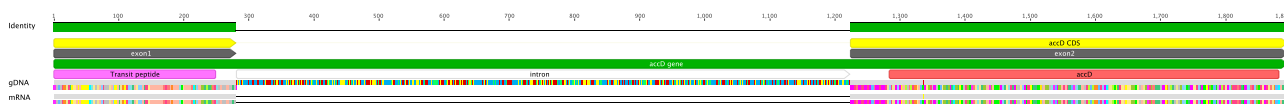


Figure S5. BUSCO assessment of the three transcriptomes. Eudicots database; 2,326 genes.



A



B

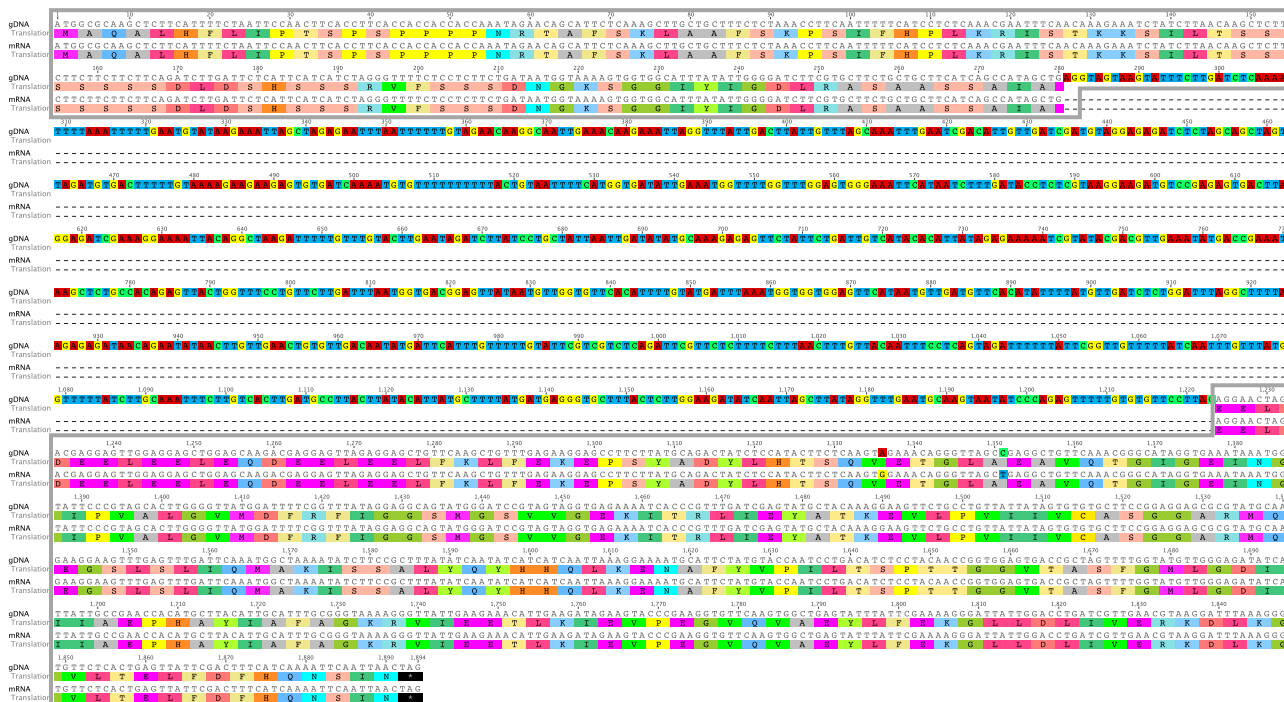


Figure S7. Characterization of the nuclear-encoded *ACC*, *RPL20*, *RPL23*, and *RPS16* genes of *Corydalis pauciovulata*. The nuclear-encoded *RPS16* (a), *ACC* (b) , *RPL20* (c), and *RPL23* (d) genes **A.** Comparison of the genomic DNA and mRNA sequences from *C. pauciovulata*. Pink annotation indicates transit peptide and red annotation indicates a conserved domain of each gene. Black annotation indicates exons and yellow annotation indicates coding sequence. Green annotation indicates gene. **B.** Nucleotide sequence alignment of the nuclear and transcript copies of each gene.

a) *RPS16*

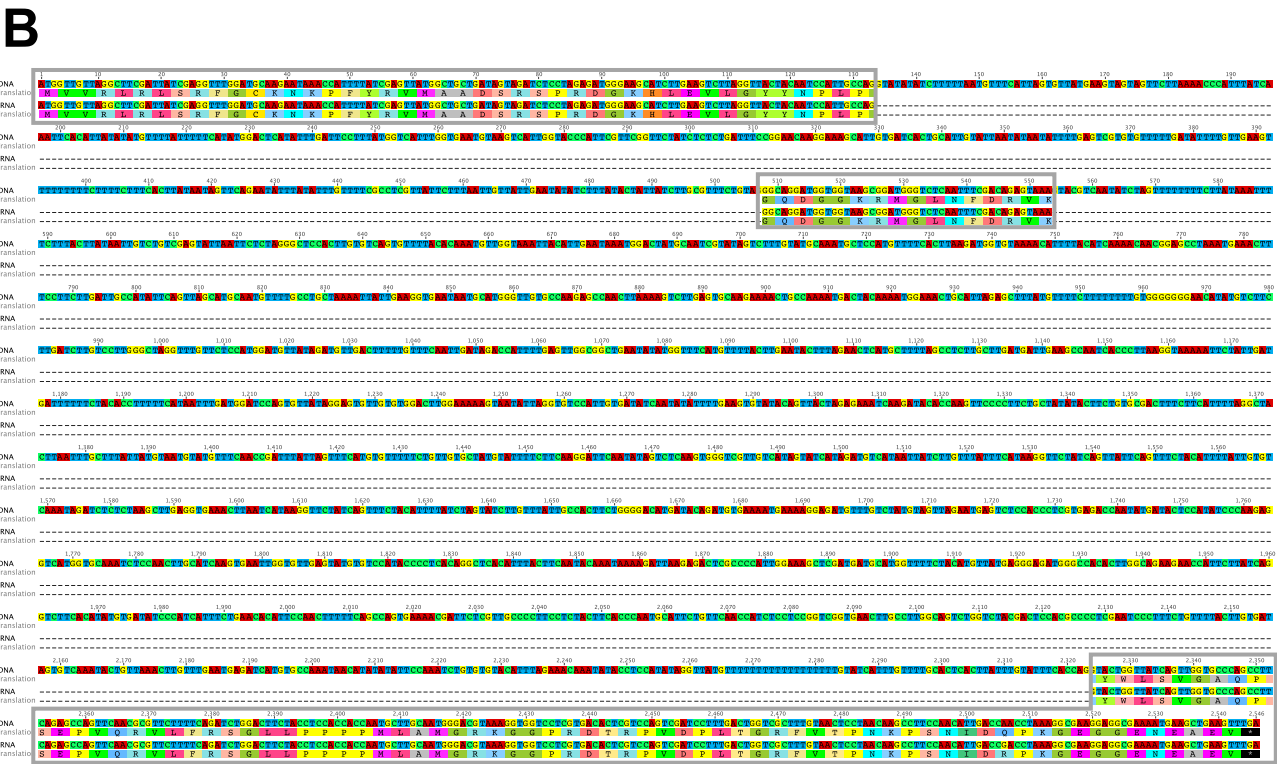
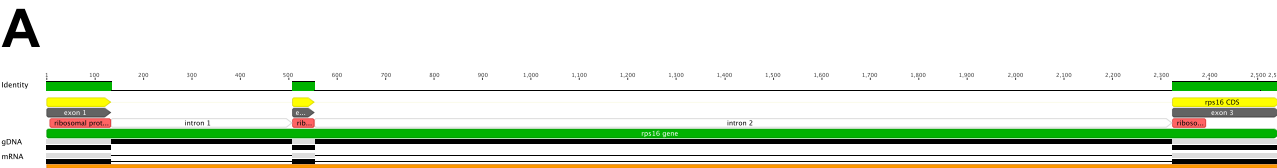
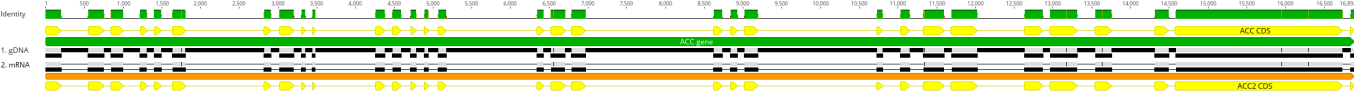


Figure S7. (continued).

b) ACC

A



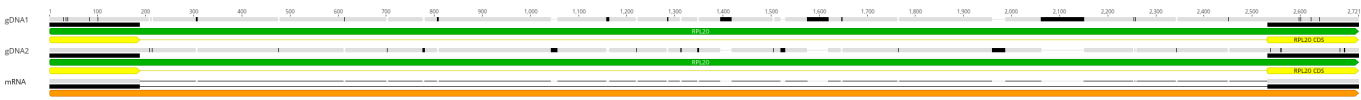
B



Figure S7. (continued).

c) *RPL20*

A



B

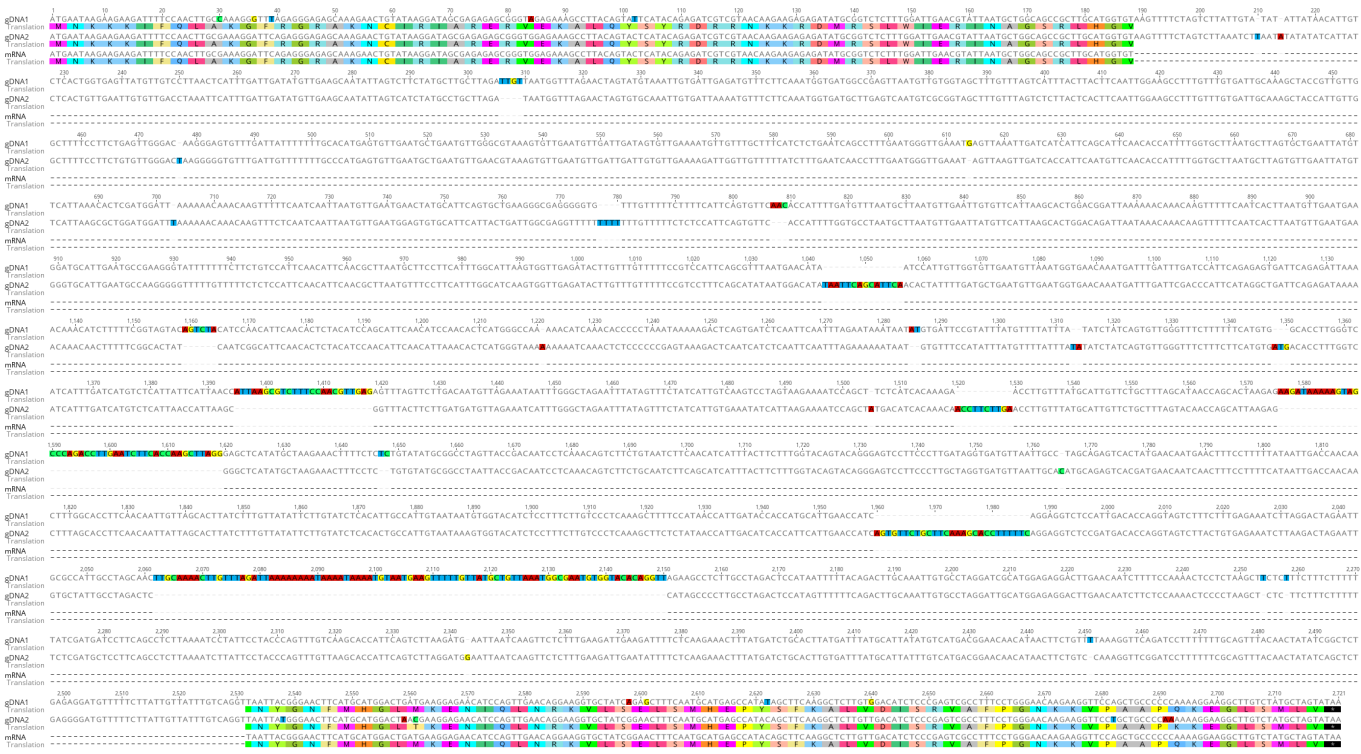
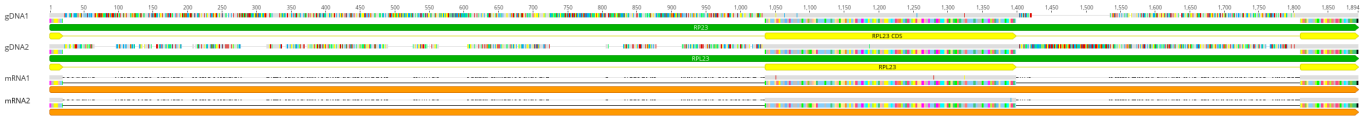


Figure S7. (continued).

d) *RPL23*

A



B

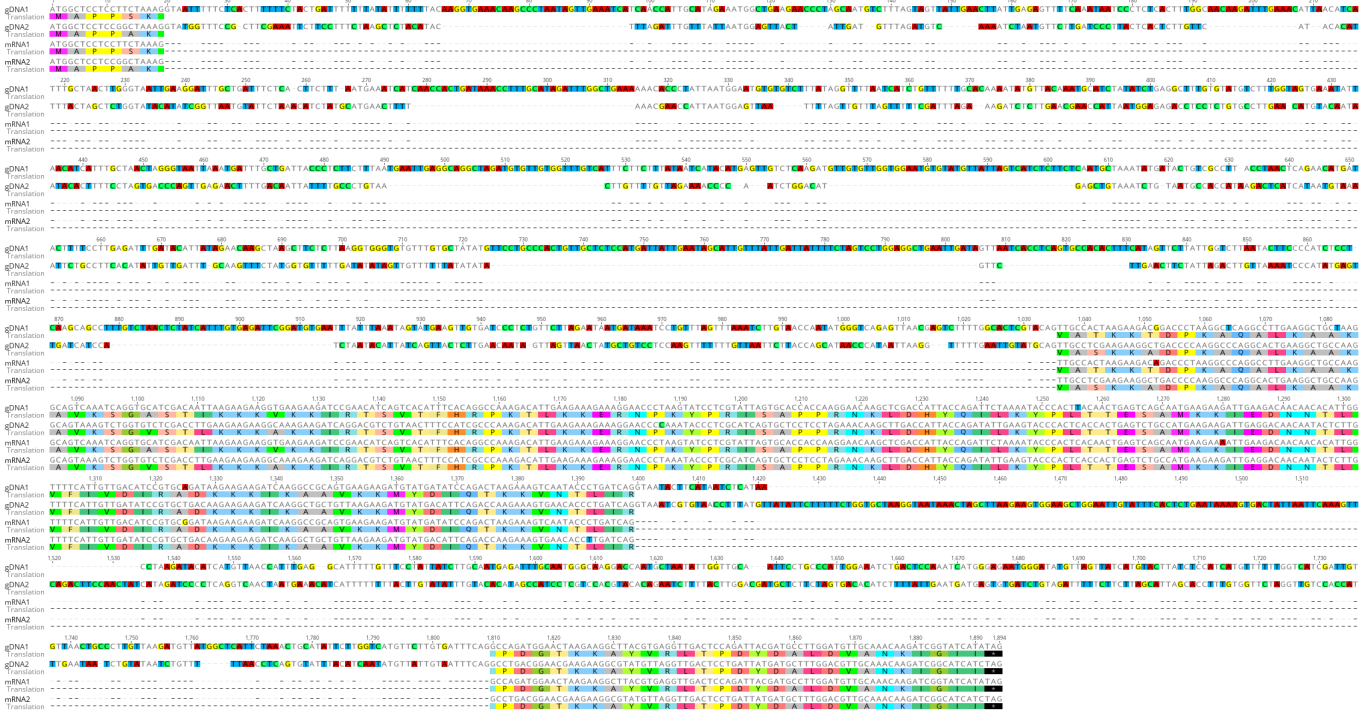


Figure S8. Genome completeness size and estimation. A. GenomeScope plot for *Corydalis pauciovulata*. The 21 *k*-mer distribution used for the estimation of *Corydalis* genome size. **B.** BUSCO assessment of the *Corydalis* genome (eudicots database; 2,326 genes).

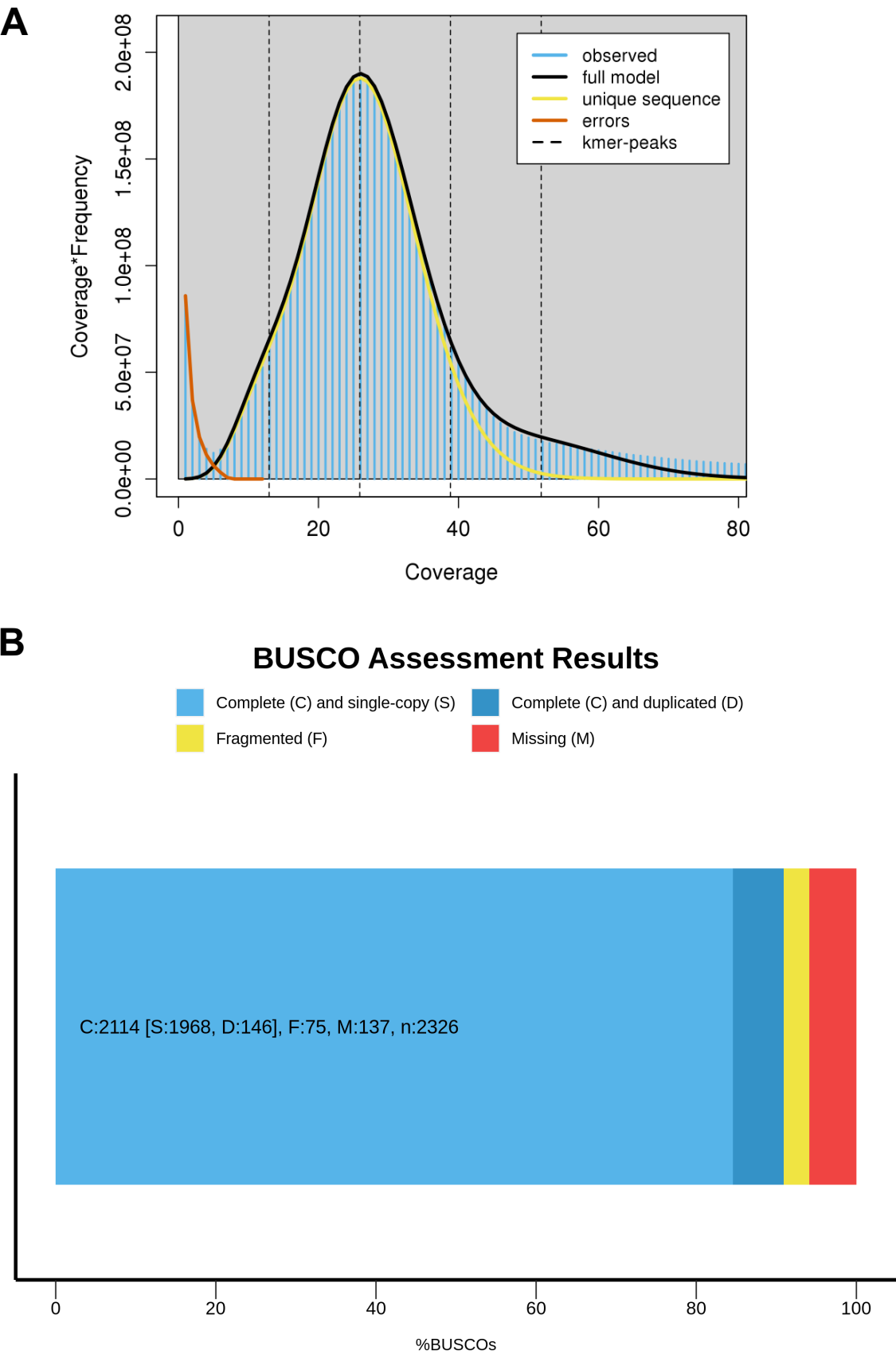


Figure S9. Variation in sequence divergence among in *Corydalis pauciovulata* and *Nelumbo nucifera* plastid and mitochondrial protein-coding genes.

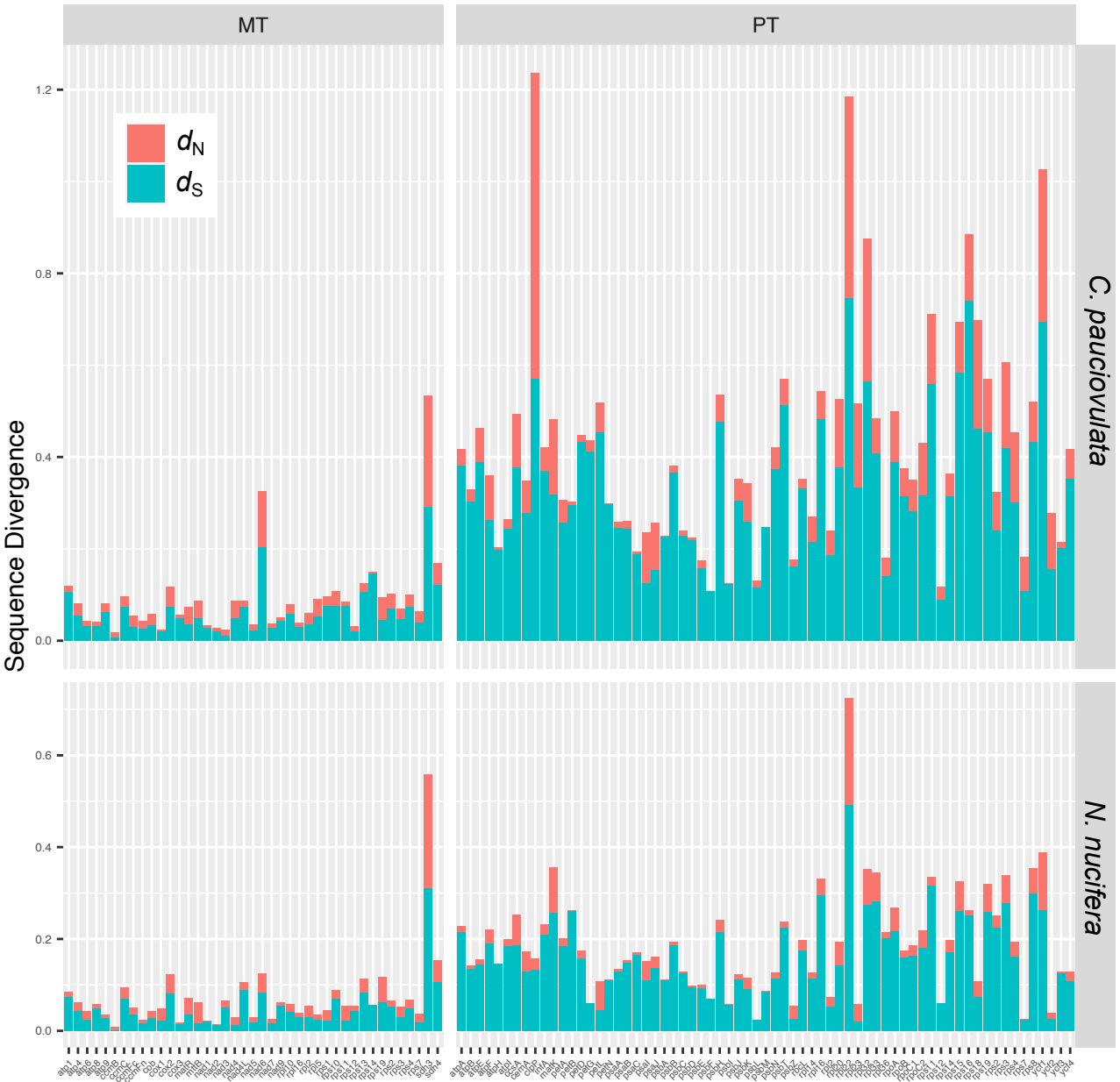


Table S1. List of annotated genes in the *Corydalis pauciovulata* plastome.

Catergory for genes	Group of gene	Name of gene
Genes for photosynthetic apparatus	photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>
	photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	cytochrome b6f	<i>petA, petB, petD, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	RuBisCo	<i>rbcl</i>
	NAD(P)H dehydrogenase genes	<i>ψndhA, ψndhB, ψndhC, ψndhD, ψndhE, ndhF, ψndhG, ψndhH, ndhI, ψndhJ, ψndhK</i>
RNA genes and genes for the genetic apparatus	transfer RNA	<i>trnH-GUG, trnK-UUU, trnQ-UUG, trnS-GCU, trnG-UCC, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnG-GCC, trnfM-CAU, trnS-GGA, trnT-UGU, trnL-UAA, trnF-GAA, trnV-UAC, trnM-CAU, trnW-CCA, trnP-UGG</i>
	ribosomal RNA	<i>rrn4.5, rrn5, rrn16, rrn23</i>
	RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	Large ribosomal subunit genes	<i>rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
	Small ribosomal subunit genes	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps16, rps18, rps19,</i>
Potential protein-coding genes	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
	Protease	<i>clpP</i>
	Envelop membrane protein	<i>cemA</i>
	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>
	c-type cytochrom synthesis gene	<i>ccsA</i>
Conserved ORFs (ycfs)		<i>ycf1, ycf2, ycf3, ycf4,</i>

ψ indicate pseudogene, Strikethrough gene indicate gene loss.

Table S2. Blast result of plastid-derived DNA segments in the *Corydalis pauciovulata* mitogenome.

Mt-length	Pt-length	Identigy	Aligned length	Query start	Query end	Hit start	Hit end	E value	Bit-score	Annotated plastid genes
6500	6485	94.502	6530	29267	35751	232844	226345	0	10163	<i>rpoB*</i> , <i>rpoC1</i> , <i>rpoC2*</i>
2620	2646	95.342	2662	107667	110312	392857	395476	0	4241	<i>16S rRNA*</i> , <i>trnI</i> , <i>trnA*</i>
2419	2363	93.054	2433	103266	105628	43488	41070	0	3635	<i>23S rRNA*</i> , <i>4.5S rRNA</i> , <i>5S rRNA</i> , <i>trnR-ACG</i>
1668	1624	83.118	1700	69152	70775	92878	91211	0	1780	<i>psbE</i> , <i>ψpsbF</i> , <i>psbL</i> , <i>psbJ</i> , <i>psaB*</i>
1585	1610	79.514	1645	25804	27413	497649	499233	0	1443	<i>atpI*</i> , <i>ψrps2</i> , <i>rpoC2*</i>
1211	1185	75.922	1275	70893	72077	579649	580859	0	956	<i>petL</i> , <i>petG</i> , <i>trnW-CCA</i> , <i>ψtrnP-UGC</i>
1221	1232	95.536	1232	44287	45518	2293	1073	0	1975	<i>psbD*</i> , <i>psbC*</i>
1194	1194	95.645	1194	48216	49409	92892	94085	0	1920	<i>psaB*</i>
1120	1130	83.319	1151	125649	126778	410490	409371	0	1190	<i>ndhE</i> , <i>ndhG*</i>
1083	1078	90.511	1096	45781	46858	1084	2	0	1505	<i>psbC*</i> , <i>trnS-UGA</i> , <i>psbZ</i>
908	1031	76.731	1040	129840	130870	323447	324354	0	840	<i>ycf1*</i>
878	912	63.938	965	38979	39890	652571	651694	2.78E-36	154	<i>petN</i>
890	919	70.947	950	29449	30367	148993	148104	3.84E-123	444	<i>rpoC2*</i>
872	863	85.36	888	3222	4084	97181	98052	0	1017	<i>matK*</i> , <i>trnK-UUU intron*</i>
332	371	66.142	381	42757	43127	649827	649496	5.36E-20	101	<i>trnT-GGU/psbD IGS*</i>
315	324	88.58	324	48143	48466	48922	49236	1.53E-115	417	<i>psaB*</i>
288	312	78.275	313	44029	44340	431467	431180	9.06E-68	260	<i>psbD*</i>
301	290	74.434	309	39920	40209	650954	650654	2.00E-44	181	<i>psbM</i>
291	278	81.849	292	28440	28717	667193	667483	2.28E-75	284	<i>rpoC2*</i>
231	240	76.763	241	125046	125285	56112	55882	2.96E-42	175	<i>ndhD*</i>
222	229	80.426	235	87992	88220	384110	384331	2.28E-56	221	<i>rpl14*</i>
201	215	77.626	219	62013	62227	191676	191476	1.03E-41	172	<i>ndhC*</i>
194	195	92.821	195	48864	49058	191870	191677	6.52E-76	286	<i>psaB*</i>
119	151	72.185	151	35885	36035	579295	579177	1.87E-19	99.6	<i>rpoB*</i>
121	121	90.083	121	41186	41306	650156	650036	1.54E-39	165	<i>trnD-GUC</i>
94	95	77.778	99	4153	4247	213271	213178	2.14E-12	75.2	<i>trnK-UUU intron*</i>
83	83	92.771	83	115276	115358	351098	351180	4.70E-27	123	<i>trnN-GUU</i>
82	81	96.341	82	37	117	636210	636291	3.17E-29	132	<i>trnH-GUG</i>
80	80	88.75	80	47268	47347	26135	26056	4.40E-21	104	<i>trnfM-CAU</i>
68	75	81.333	75	94280	94354	653639	653706	2.14E-12	75.2	<i>ψtrnI-CAU</i>
73	73	94.521	73	13303	13375	163267	163195	2.44E-24	114	<i>trnM-CAU</i>
64	68	88.235	68	1622	1689	407658	407721	1.80E-17	86.9	<i>ndhA*</i>
439	453	88.742	453	1694	2146	407784	408222	1.78E-169	591	<i>ndhA*</i>
531	528	90.467	535	1	528	408301	408831	0	732	<i>ψndhI</i>
527	531	90.112	536	1	531	409204	409730	0	723	<i>ψndhG</i>
303	303	91.531	307	1	303	409977	410279	1.16E-123	435	<i>ndhE</i>

Bold fonts indicate intact genes, * indicates fragments, ψ indicate pseudogene

Table S3. Predicted RNA editing in 41 protein-coding genes for the *Corydalis pauciovulata* mitogenome.

Gene	<i>C. pauciovulata</i>		<i>N. nucifera</i>		<i>L. tulipifera</i>	
	Length (bp)	RNA editing	Length (bp)	RNA editing	Length (bp)	RNA editing
<i>atp1</i>	1,530	12	1,530	12	1,530	13
<i>atp4</i>	597	14	597	14	582	15
<i>atp6</i>	798	26	774	25	807	28
<i>atp8</i>	480	5	480	5	480	5
<i>atp9</i>	282	9	225	12	226	11
<i>ccmB</i>	621	40	621	41	621	41
<i>ccmC</i>	720	36	759	34	723	39
<i>ccmFc</i>	1,356	21	1356	21	1359	21
<i>ccmFn</i>	1,734	39	1734	41	1806	42
<i>cob</i>	1,212	22	1182	21	1182	28
<i>cox1</i>	1,584	34	1584	10	1584	36
<i>cox2</i>	786	17	783	18	759	17
<i>cox3</i>	798	19	798	20	798	21
<i>matR</i>	1,968	19	1968	18	1953	19
<i>mttB</i>	828	36	828	37	768	35
<i>nad1</i>	978	34	978	33	978	32
<i>nad2</i>	1,467	37	1467	37	1467	40
<i>nad3</i>	357	18	357	16	357	20
<i>nad4</i>	1,461	57	1488	14	1488	15
<i>nad4L</i>	303	15	303	63	303	61
<i>nad5</i>	2,001	39	3013	40	2013	44
<i>nad6</i>	738	20	633	17	723	24
<i>nad7</i>	1,188	40	1185	39	1185	39
<i>nad9</i>	573	14	573	15	573	14
<i>rpl2</i>	1,536	8	999	3	1665	6
<i>rpl5</i>	567	10	561	9	561	10
<i>rpl10</i>	489	6	489	7	480	8
<i>rpl16</i>	435	7	516	9	435	7
<i>rps1</i>	609	4	606	5	606	5
<i>rps2</i>	681	8	663	7	621	7
<i>rps3</i>	1,686	14	1692	13	1584	15
<i>rps4</i>	1,065	21	1059	21	1071	21
<i>rps7</i>	447	2	447	2	450	2
<i>rps10</i>	387	3	363	4	360	4
<i>rps11</i>	579	3	444	3	516	3
<i>rps12</i>	378	10	378	10	378	11
<i>rps13</i>	351	6	351	6	351	6
<i>rps14</i>	303	3	303	3	303	3
<i>rps19</i>	285	3	285	3	282	3
<i>sdh3</i>	336	2	375	2	330	8
<i>sdh4</i>	687	5	450	5	405	5
		738		715		784

Table S4. RNA editing validation of 21 genes using transcriptome data.

Gene	PRE P-MT	> 0.5	TC	Gene	PREP- MT	> 0.5	TC	Gene	PREP- MT	> 0.5	TC	Gene	PRE P-MT	> 0.5	TC		
atp1	57		O	ccmC	458	0.78	O	cox2	676	1.00	O	nad9	92	0.75	O		
	971	0.80	O			715	0.79		O		113		0.92	O			
	1039	1.00	O			736	1.00		O		167		0.92	O			
	1064	1.00	O			473	1.00	O		769	0.50			190	1.00	O	
	1115	1.00	O			497	1.00	O	cox3	112	1.00	O		223	1.00	O	
	1168	1.00	O			499	1.00	O			245	0.92	O		298	0.83	O
	1178	0.90	O			521	1.00	O			257	0.67	O		311	0.75	O
	1216	1.00	O			548	1.00	O		263	1.00	O		328	1.00	O	
	1262	1.00	O			568	1.00	O		289	0.92	O		368	1.00	O	
	1292	0.80	O			572	1.00	O		298	1.00	O		398	1.00	O	
	1415	1.00	O			575	1.00	O		304	1.00	O		406	1.00	O	
	1490	0.90	O			605	1.00	O		311	0.92	O		439	1.00	O	
	1499	0.90	O			608	0.89	O		314	0.92	O		539	1.00	O	
atp4	56	0.86	O		614	0.78	O		388	1.00	O	rpl10	77	1.00			
	59	0.57	O		619	0.78	O		413	0.83	O			83		O	
	71	1.00	O		650	0.78			419	1.00	O			101	0.83	O	
	89	1.00	O		656	0.89			422	0.92	O		134	0.83	O		
	118	0.71	O		665	0.78			512	0.75	O		155	0.83	O		
	121	1.00	O		673	0.78			527	0.92	O		314	0.83	O		
	215	1.00	O	cob	118	0.92	O		566	0.92	O		397		O		
	226	0.57				178	1.00	O		653	1.00	O		409	0.83		
	227	0.57	O			286	1.00	O		754	0.92	O	rps1	26		O	
	248	1.00	O			298	1.00	O		764	0.92	O			45		O
			O			325	1.00	O	matR	26	1.00	O			74	1.00	O
			O			358	1.00	O			32	0.62	O		128	0.67	O
	395	1.00	O			407	1.00	O			43		O		161	0.67	O
407	0.71	O			419	1.00	O		193	0.75	O		212		O		
416	0.86	O			564		O		235	0.62	O		485	0.67			
490	0.57				568	0.92	O		236	0.62	O		590		O		
atp6	37	0.75	O			580	1.00	O		239	0.75		rps3	58	1.00	O	
	47	1.00	O			680	1.00	O		249		O			64	1.00	O
	116	1.00	O			715	1.00	O		326	1.00	O			69		O
	167	1.00	O		725	1.00	O		382	0.75	O		92	0.71	O		
	173	1.00	O		808	1.00	O		413		O		126		O		
	224	1.00	O		853	1.00	O		923	1.00			512	0.71	O		
	229	0.75	O		908	1.00	O		1009	0.88			599	1.00	O		
	236	0.67	O		914	1.00	O		1064	1.00	O		699		O		
	254	1.00	O		982	0.85			1065		O		713	0.86	O		
	262	1.00	O		1015	1.00	O		1675	1.0							

	205		O		1402	0.67	O		418	0.62	O	rps12	71	0.94	O
	212	1.00	O		1405	0.78	O		445		O		100	1	O
	223	1.00	O		1433	1.00	O		446	0.75	O		104	1	O
ccmC	5		O		1446		O		453		O		112	0.88	O
	103	1.00	O		1489	1.00	O		476	0.57	O		143		O
	115	0.78	O	cox2	27		O		511	0.88	O		146	1	
	128	0.56			33		O		536	0.88	O		148		O
	133	0.67	O		38		O		544	1.00	O		196	0.94	O
	179	0.78	O		71	1.00	O		580	0.88	O		221	0.88	O
	184	1.00	O		161	0.95	O		587	0.62	O		232	1	O
	227	0.78	O		253	1.00	O		593	0.50	O		269	0.94	O
	281	0.56	O		278	1.00	O		617	0.75	O		284	0.76	O
	299	1.00	O		379	1.00	O		649	1.00	O	rps13	5	0.6	O
	331	1.00	O		443	1.00	O		699		O		26	0.9	O
	358	1.00	O		460		O		706	0.62	O		56	0.9	O
	395	1.00	O		461	1.00	O		711		O		100	0.9	O
	399		O		476	1.00	O		712		O		256	1	O
	400	0.89	O		544	1.00	O		713	0.50			287	1	
	421	0.78	O		557	1.00	O		722	0.50	O	sdh3	54		O
	436	0.89	O		581	1.00	O		743	1.00	O		67	1	O
	446	0.78	O		614	1.00	O		752	0.75	O		74	1	O
	451	1.00	O		632	0.84	O	nad9	14	0.73	O				

Table S5. Blast results of ORFs (>150 bp) in the *Corydalis pauciovulata* mitogenome.

Name	aa	Length	Min	Max	E-Value	Bitscore	Accession	Short name
ORF115	115	348	17,249	16,902	0.00030787	37.241	cl21549	rve superfamily
ORF283	283	852	34,730	35,581	0.00373583	37.6768	cl24149	MCPV1 superfamily
ORF94	94	285	44,127	44,411	4.44E-06	42.2746	cl37957	RVT_1 superfamily
ORF126	126	381	61,235	60,855	3.36E-18	76.4708	cl02808	RT_like superfamily
ORF71	71	216	63,098	62,883	0.00514233	33.0558	cl04955	LanC_like superfamily
ORF68	68	207	124,460	124,666	6.96E-03	32.6041	cl00467	Ntn_hydrolase superfamily
ORF192	192	579	147,778	147,200	5.06E-34	116.823	cd09274	RNase_HI_RT_Ty3
ORF53a	53	162	166,866	166,705	1.95E-03	32.6155	cl16506	zf-RVT superfamily
ORF102	102	309	192,760	192,452	1.10E-08	47.2757	pfam17921	Integrase_H2C2
ORF97a	97	294	219,704	219,997	1.71E-04	38.3459	cl28114	MatK_N superfamily
ORF62	62	189	221,790	221,602	0.00681989	31.6238	cl00075	HATPase superfamily
ORF67	67	204	224,063	224,266	1.17E-20	81.0079	cl06662	RVT_2 superfamily
ORF117a	117	354	224,227	224,580	4.85E-25	94.4899	cl06662	RVT_2 superfamily
ORF235a	235	708	225,328	226,035	2.49E-06	47.3813	cl05469	Mitovir_RNA_pol superfamily
ORF252	252	759	254,757	255,515	7.77E-13	65.718	pfam03372	Exo_endo_phos
ORF235b	235	708	255,825	256,532	6.50E-16	73.4794	cl02808	RT_like superfamily
ORF201	201	606	261,261	261,866	1.41E-04	41.6033	cl05469	Mitovir_RNA_pol superfamily
ORF154	154	465	264,904	265,368	1.62E-21	84.185	cl16623	DUF4283 superfamily
ORF208	208	627	265,788	266,414	0.00647206	36.1753	cl00490	EEP superfamily
ORF121	121	366	272,489	272,124	1.86E-03	33.304	cl24970	zf-C2H2_10 superfamily
ORF100a	100	303	281,917	281,615	3.61E-11	56.9158	cl02808	RT_like superfamily
ORF111	111	336	293,729	293,394	1.44E-21	84.56	cl02808	RT_like superfamily
ORF117b	117	354	294,271	293,918	3.73E-17	75.5008	cl05469	Mitovir_RNA_pol superfamily
ORF178	178	537	322,474	321,938	9.75E-27	104.733	cl28058	DNA_pol_B_2 superfamily
ORF110a	110	333	335,020	334,688	1.90E-06	43.0479	cl07848	ELF superfamily
ORF117c	117	354	336,406	336,053	5.23E-03	33.0234	pfam17921	Integrase_H2C2
ORF87	87	264	356,456	356,719	7.39E-31	106.526	cd01647	RT_LTR
ORF63	63	192	363,019	362,828	4.63E-03	33.0334	cl02808	RT_like superfamily
ORF234	234	705	368,537	369,241	6.89E-05	40.3515	cl00350	Ribosomal_S19 superfamily
ORF128	128	387	379,319	379,705	8.16E-21	83.4044	cl02808	RT_like superfamily
ORF54	54	165	451,349	451,513	0.0030072	31.1294	cl18687	zf-CCHC_4 superfamily
ORF123	123	372	461,763	461,392	3.72E-41	132.137	cl19078	REC superfamily
ORF75	75	228	461,962	461,735	8.99E-08	45.527	cl00075	HATPase superfamily
ORF57	57	174	462,552	462,379	7.41E-07	41.0154	pfam00512	HisKA
ORF79	79	240	463,305	463,066	2.44E-13	60.1745	cl14782	RNase_H_like superfamily
ORF100b	100	303	473,907	473,605	0.00668953	34.2415	cl17011	Arginase_HDAC superfamily
ORF55	55	168	476,720	476,553	8.19E-03	31.877	cl21496	2OG-FelI_Oxy superfamily
ORF155	155	468	478,020	478,487	2.90E-18	77.6264	cl02808	RT_like superfamily
ORF93	93	282	492,067	492,348	3.78E-19	75.2217	cl14782	RNase_H_like superfamily
ORF97b	97	294	500,786	500,493	2.32E-03	35.2798	cl19167	Bac_export_2 superfamily
ORF106	106	321	506,331	506,011	0.00640333	34.1905	cl06214	FYTT superfamily
ORF119	119	360	509,910	510,269	6.91E-28	99.2078	pfam14111	DUF4283
ORF333	333	1,002	528,991	529,992	6.08E-71	218.234	cd01647	RT_LTR
ORF127	127	384	530,032	530,415	1.89E-14	63.6471	pfam17919	RT_RNaseH_2
ORF167	167	504	530,351	530,854	4.25E-14	64.8213	cl14782	RNase_H_like superfamily
ORF226	226	681	530,863	531,543	0.00298028	35.9113	cl14782	RNase_H_like superfamily
					4.37E-17	74.4312	cl14782	RNase_H_like superfamily
ORF148	148	447	531,583	532,029	6.38E-04	36.8753	pfam17921	Integrase_H2C2
					1.12E-14	65.3605	pfam00665	rve
ORF539	539	1,620	534,606	532,987	4.12E-04	41.899	cl37731	BASP1 superfamily
ORF145	145	438	534,992	534,555	4.81E-31	113.564	cl39183	PMD superfamily
ORF110b	110	333	545,805	546,137	5.59E-08	46.4036	cl21610	PQ-loop superfamily
ORF125	125	378	575,016	574,639	1.08E-07	45.0354	cl39048	Retrotran_gag_3 superfamily
ORF107	107	324	576,313	575,990	1.24E-44	141.066	cd09272	RNase_HI_RT_Ty1
ORF88	88	267	576,612	576,346	1.58E-04	37.4477	cl14782	RNase_H_like superfamily
ORF244	244	735	592,626	591,892	1.80E-12	61.5792	pfam03732	Retrotrans_gag
ORF53b	53	162	663,285	663,124	7.44E-03	32.0115	cl34545	PilF superfamily

Table S6. Potential chimeric ORFs.

	ORF start	ORF end	ORF length	Identity	ORF hit start	ORF hit end	Chimera length	Gene hit start	Gene hit end	E-value	Gene	No of Transmembrane helices / probabilities	Transcript
ORF457a	593255	594628	1374	99.7	1	701	701	1	701	0.00E+00	<i>atp1</i>	1 / 0.72285	○
ORF457b	655980	654607											
ORF244	259778	260512	735	89.5	19	56	38	1	38	6.38E-09	<i>rpl2</i>	1 / 0.99742	○
ORF234	368537	369241	705	84.1	1	79	82	1	82	8.50E-20	<i>rps19</i>	1 / 0.97547	○
ORF146	258635	259075	441	94.3	94	128	35	307	341	3.10E-10	<i>rpl5</i>	3 / 0.13557	○
ORF112	528026	527688	339	96.8	138	168	31	303	283	2.87E-09	<i>ccmFc</i>	0 / 0.48217	
ORF60a	246603	246785	183	100	31	93	63	447	385	1.49E-28	<i>rps7</i>	0 / 0.34003	
ORF60b	633132	632950		82.1	140	177	39	718	756	1.15E-04	<i>cob</i>	0 / 0.21732	
				96.8	136	166	31	274	304	1.49E-09	<i>nad5</i>		
ORF56	239013	239183	171	95.4	84	148	65	681	745	2.50E-25	<i>cob</i>	1 / 0.15049	○
ORF54	154633	154469	165	97.1	1	34	34	1	34	3.13E-11	<i>sdh3</i>	1 / 0.04103	
ORF49a	45477	45328	150	90.3	27	57	31	241	271	6.20E-07	<i>cob</i>	0 / 0.42089	
ORF49b	353025	351876	150	100	74	103	30	610	639	9.82E-11	<i>sdh4</i>	0 / 0.51849	○

Table S7. Information on the nuclear-encoded NDH complex genes.

Group	Gene	Accession number		
		<i>Arabidopsis thaliana</i>	<i>Aquilegia coerulea</i>	<i>Corydalis ochotensis</i>
Linkers	LHCA5	AT1G45474	Aqcoe6G141900.1	○
	LHCA6	AT1G19150	Aqcoe7G431400.1	○
subcomplex A	NdhL	AT1G70760	Aqcoe3G005000.1	-
	NdhM	AT4G37925	Aqcoe6G031500.1	-
	NdhN	AT5G58260	Aqcoe5G227700.1	-
	NdhO	AT1G74880	Aqcoe5G236000.1	-
subcomplex EDB	NdhS	AT4G23890	Aqcoe2G163300.1	-
	NdhT	AT4G09350	Aqcoe6G015200.1	○
	NdhU	AT5G21430	Aqcoe7G114900.1	-
	NdhV	AT2G04039	Aqcoe1G375600.1	-
Subcomplex B	PnsB1	AT1G15980	Aqcoe7G199900.1	-
	PnsB2	AT1G64770	Aqcoe2G064400.1	-
	PnsB3	AT3G16250	Aqcoe7G162000.1	○
	PnsB4	AT1G18730	Aqcoe3G060000.1	-
	PnsB5	AT5G43750	Aqcoe5G042700.1	-
subcomplex L	PnsL1	AT2G39470	Aqcoe1G473000.1	○
	PnsL2	AT1G14150	Aqcoe3G060200.1	○
	PnsL3	AT3G01440	Aqcoe1G367800.1	○
	PnsL4	AT4G39710	Aqcoe6G058500.1	○
	PnsL5	AT5G13120	Aqcoe3G164000.1	○

○: present, -: absent

Table S8. Information on the nuclear-encoded DNA-RRR genes.

Category	gene	TAIR ID	Targeting	<i>Liriodendron tulipifera</i>	<i>Nelumbo nucifera</i>	<i>Corydalis pauciovulata</i>			
						GenBank accession numbers	CDS	length	exon
DNA polymerase	POLIA	AT1G50840	plastid/mitochondria	O	O	PP437511	3339	8,383	13
	POLIB	AT3G20540	plastid/mitochondria	-	-	-	-	-	-
Helicase/primase	Twinkle	AT1G30680	plastid/mitochondria	O	O	PP437512	2130	28,267	21
Topoisomerase	GYRA	AT3G10690	plastid/mitochondria	O	O	PP437513	2910	26,767	27
	GYRBC	AT3G10270	plastid	O	O	PP437514	2253	2,253	1
	GYRBM	AT5G04130	mitochondria	-	-	-	-	-	-
	TopoI	AT4G31210	plastid/mitochondria	O	partial/XM_010259587	PP437515	3618	14,933	22
Recombinase	RECA1	AT1G79050	plastid	O	O	PP437516	1257	18,399	13
	RECA2	AT2G19490	plastid	O	partial/XM_010252152	PP437517	1278	3,097	8
	RECA3	AT3G10140	mitochondria	O	O	PP437518	1218	4,875	8
	RECG	AT2G01440	plastid/mitochondria	O	partial/XM_010276366	PP437519	3189	33,633	17
	RECX	AT3G13226	plastid/mitochondria	O	O	PP437520	999	2,494	6
ssDNA-binding/recombination mediator	SSB1	AT4G11060	mitochondria	O	O	PP437521	684	8,003	4
	SSB2	AT3G11060	mitochondria	-	-	-	-	-	-
	OSB1	AT1G47720	mitochondria	O	O	PP437522	900	1,846	4
	OSB2	AT4G20010	plastid	O	O	PP437523	1245	13,149	8
	OSB3	AT5G44785	plastid/mitochondria	-	-	-	-	-	-
	OSB4	AT1G31010	mitochondria	-	-	-	-	-	-
	WHY1	AT1G14410	plastid	O	O	PP437524	804	3,261	7
	WHY2	AT1G71260	mitochondria	O	O	PP437525	738	4,787	8
	WHY3	AT2G02740	plastid	-	O	-	-	-	-
	ODB1	AT1G71310	mitochondria	O	O	PP437526	612	3,383	4
	ODB2	AT5G47870	plastid	O	O	PP437527	687	10,426	4
MutS	MSH1	AT3G24320	plastid/mitochondria	partial	partial/XM_019198676	PP437528	3378	46,688	22
DNA glycosylase	UNG	AT3G18630	mitochondria	partial	O	PP437529	966	4,031	7
	OGG1	AT1G21710	nucleus/plastid	O	O	PP437530	1206	2,824	3
	NTH1	AT2G31450	plastid	O	O	PP437531	1053	3,963	11
	NTH2	AT1G05900	Plastid/Mitochondria	-	-	-	-	-	-
AP endonuclease	ARP	AT2G41460	plastid	O	partial/XM_010275015	PP437532	1668	5,471	13
	APE1L	AT3G48425	nuclear/plastid	O	O	PP437533	854	5,187	9
	APE2	AT4G36050	nuclear/mitochondria	O	partial/XM_010270362	PP437534	1947	3,671	10
DNA ligase	LIG1	AT1G08130	nuclear/mitochondria	O	O	PP437535	2451	5,591	17

○: present, -: absent