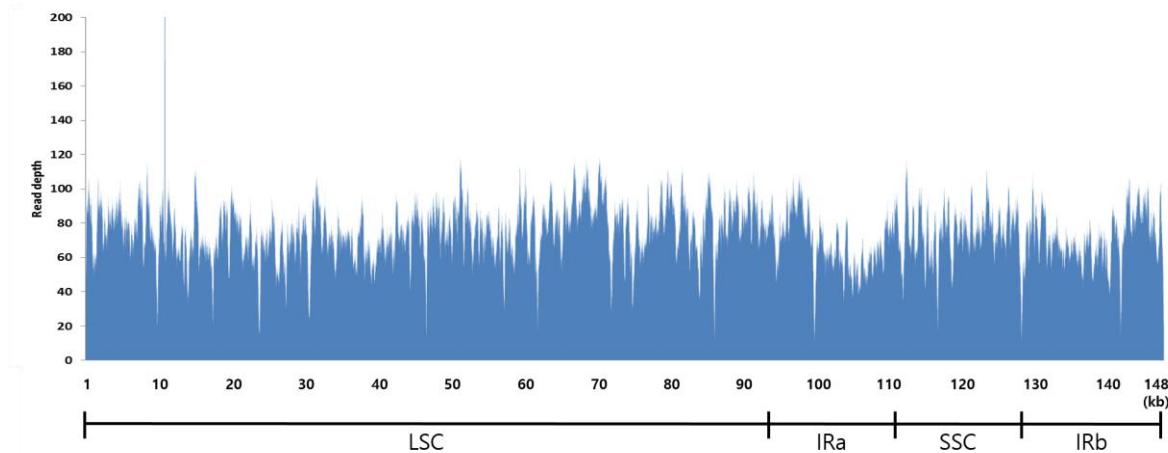
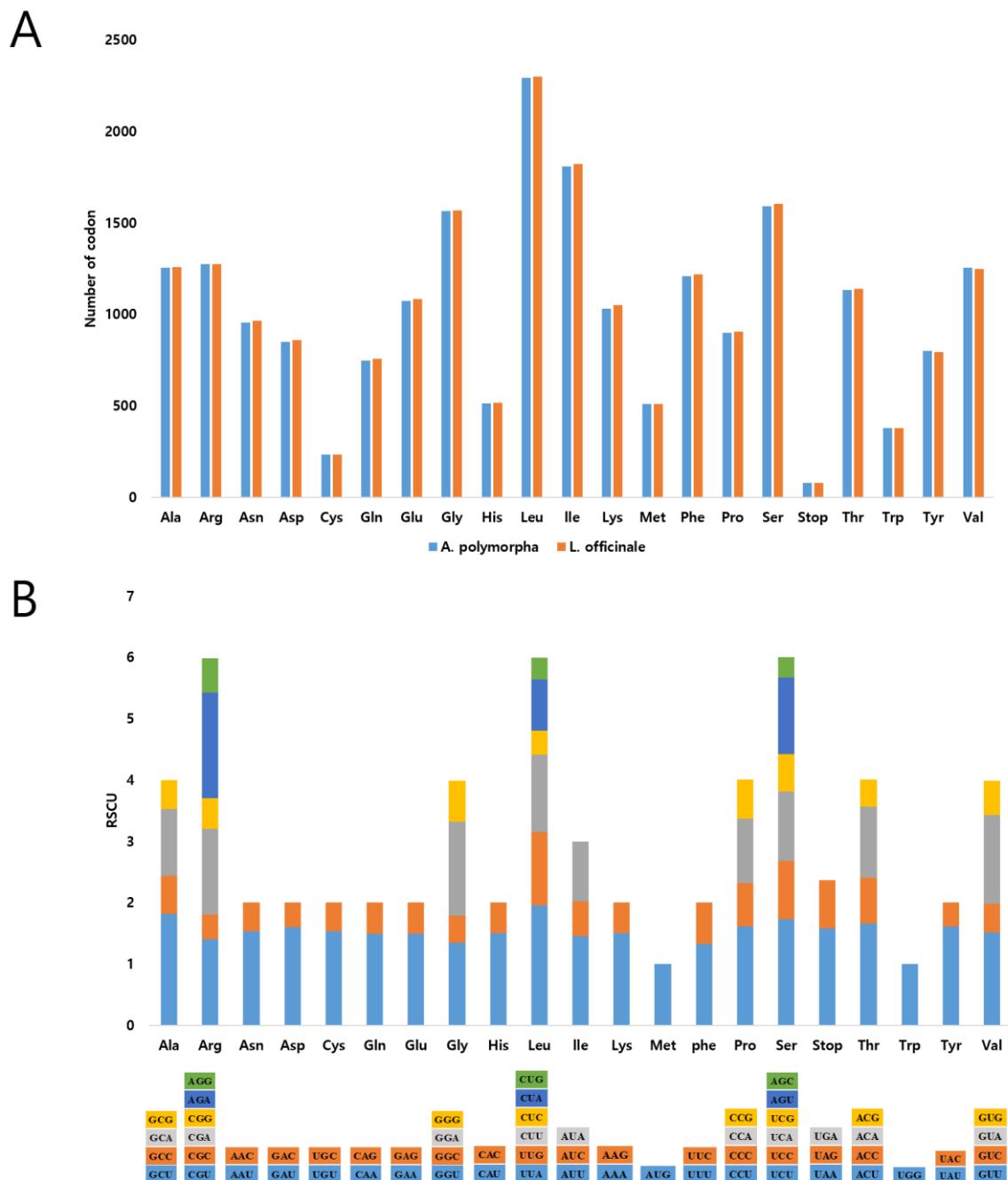


1 Supplementary Materials: Sequencing and
2 Comparative Analysis of the Chloroplast Genome of
3 *Angelica polymorpha* and the Development of a Novel
4 Indel Marker for Species Identification

5 Inkyu Park, Sungyu Yang, Wook Jin Kim, Jun-Ho Song, Hyun-Sook Lee, Hyun Oh Lee, Jung-
6 Hyun Lee, Sang-Nag Ahn, and Byeong Cheol Moon*



7
8 **Figure S1:** Distribution of paired-end reads mapped onto the complete chloroplast (CP) genome sequence of *A.*
9 *polymorpha*.



10

11 **Figure S2:** Codon frequencies and relative synonymous codon usage (RSCU) values of the CP genomes of *A.*
 12 *polymorpha* and *L. officinale*. (A) Amino acid frequencies in protein-coding genes. (B) Codon usage for 20 amino
 13 acids and stop codons in 78 protein-coding genes.

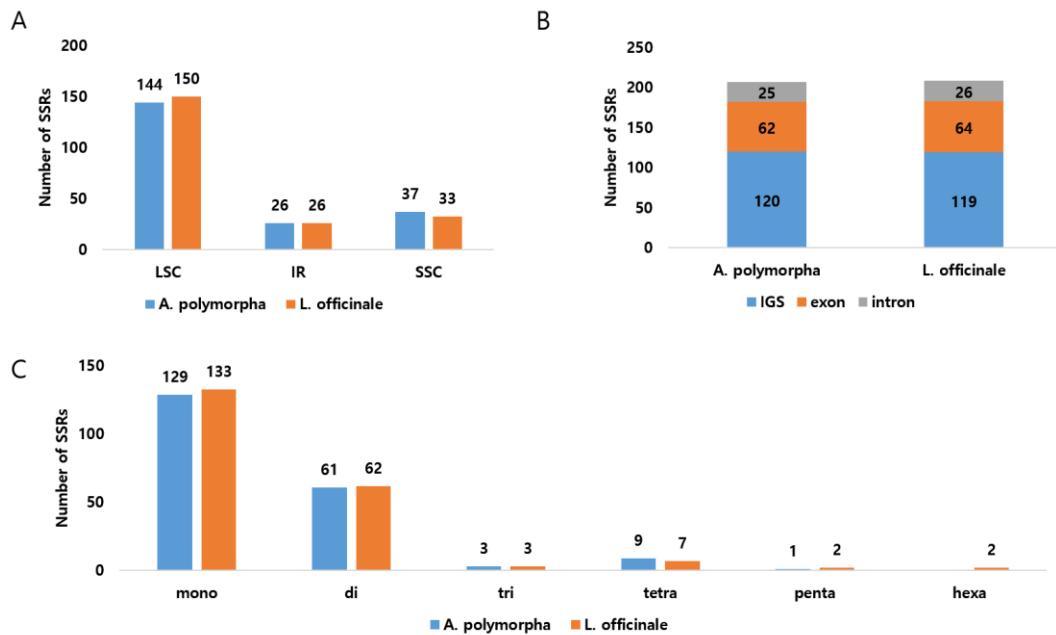
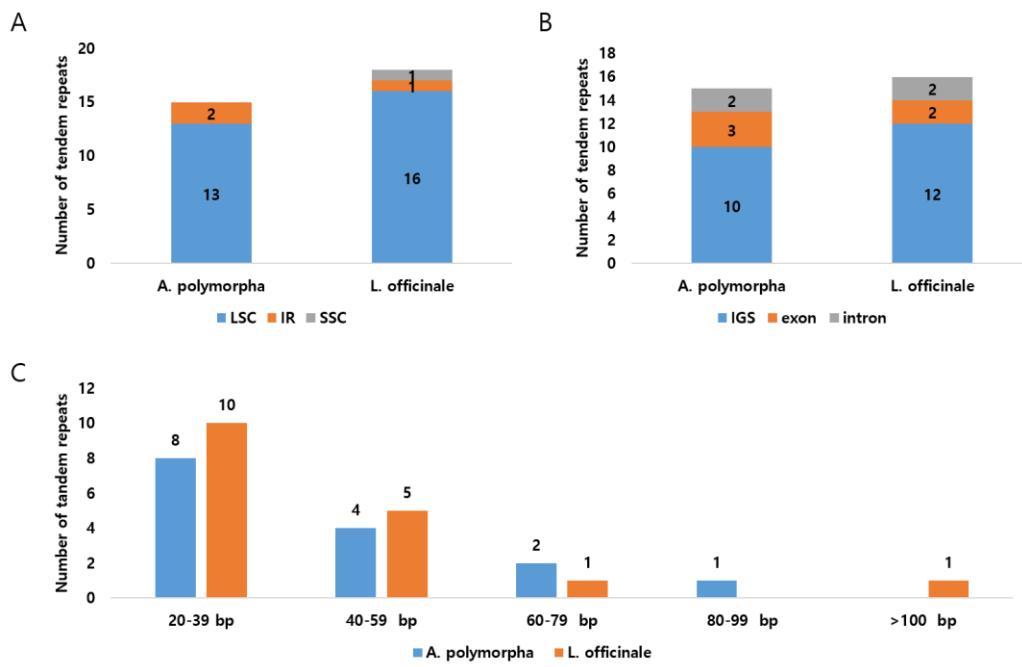


Figure S3: Distribution of simple sequence repeats (SSRs) in the CP genomes of *A. polymorpha* and *L. officinale*.
 (A) Number of SSRs in CP genomes. (B) Number of SSRs in exons, introns and intergenic spacer (IGS) regions.
 (C) Number of different types of SSRs in CP genomes.

14
15
16
17



18

19 **Figure S4:** Analysis of tandem repeats in the CP genomes of *A. polymorpha* and *L. officinale*. (A) Distribution of
 20 tandem repeats in different regions of CP genomes. (B) Number of tandem repeats in IGS regions, exons, and
 21 introns. (C) Distribution of tandem repeats of variable lengths in CP genomes.

22

23

Table S1: Details of raw sequence reads and CP genome assembly of *A. polymorpha*.

Parameter	Number of reads
Input reads	5,000,000
Trimmed reads	4,307,210
Total raw bases (bp)	1,255,000,000
Trimmed bases (bp)	940,906,215
Aligned reads	51,266
Coverage (X)	75
Chloroplast genome size (bp)	147,127

24

Table S2: PCR-based sequence validation of junctions between the large single copy (LSC), small single copy (SSC), and inverted repeat (IRa and IRb) regions in the CP genome of *A. polymorpha*.

Location	PCR-based sequence (bp)	Start position (bp)	End position (bp)	Identity (%)
LSC/IRa	376	93,336	93,711	100
IRa/SSC	175	111,453	111,572	100
SSC/IRb	508	128,855	129,362	100
IRb/LSC	732	146,942	546	100

27

28
29**Table S3:** List of genes and encoded proteins identified in the CP genomes of *A. polymorpha* and *L. officinale*.

Gene group	Gene name
Photosystem I	<i>psaA, B, C, I, J, ycf3²⁾, ycf4</i>
Photosystem II	<i>psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z</i>
Cytochrome b6/f	<i>petA, B¹⁾, D¹⁾, G, L, N</i>
ATP synthase	<i>atpA, B, E, F¹⁾, H, I</i>
Rubisco	<i>rbcL</i>
NADH oxidoreductase	<i>ndhA¹⁾, B¹⁽³⁾, C, D, E, F, G, H, I, J, K</i>
Large subunit ribosomal proteins	<i>rpl2¹⁽³⁾, 14, 16¹⁾, 20, 22, 23³⁾, 32, 33, 36</i>
Small subunit ribosomal proteins	<i>rps2, 3, 4, 7³⁾, 8, 11, 12²⁽³⁾⁴⁾, 14, 15, 16, 18, 19</i>
RNA polymerase	<i>rpoA, B, C¹⁾, C2</i>
Unknown function protein-coding gene	<i>ycf1³⁾, 2, 15³⁾</i>
Other genes	<i>accD, ccsA, cemA, clpP²⁾, matK</i>
Ribosomal RNAs	<i>rrn16³⁾, 23³⁾, 4.5³⁾, 5³⁾</i>
Transfer RNAs	<i>trnA-UGC¹⁽³⁾, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC¹⁾, trnG-GCC, trnH-GUG, trnI-CAU, trnI-GAU¹⁽³⁾, trnK-UUU¹⁾, trnL-UAA¹⁾, trnL-UAG, trnL-CAA³⁾, trnM-CAU, trnM-CAU, trnN-GUU³⁾, trnP-UGG, trnQ-UUG, trnR-ACG³⁾, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-UAC¹⁾, trnV-GAC³⁾, trnW-CCA, trnY-GUA</i>

30 ¹Genes containing a single intron.31 ²Genes containing two introns.32 ³Genes present as two copies in the IR regions.33 ⁴Trans-spliced gene.

34

Table S4: List of intron-containing genes in the CP genomes of *A. polymorpha* and *L. officinale*.

No.	Gene	Region	Exon I ¹	Intron I ¹	Exon II ¹	Intron II ¹	Exon III ¹
1	<i>trnK-UUU</i>	LSC	37	2523*	35		
2	<i>rps16</i>	LSC	40	855 (857)	197		
3	<i>trnG-UCC</i>	LSC	23	704	48		
4	<i>atpF</i>	LSC	145	722	401		
5	<i>rpoC1</i>	LSC	435	754	1605		
6	<i>ycf3</i>	LSC	126	720	226	778	155
7	<i>trnL-UAA</i>	LSC	35	504	50		
8	<i>trnV-UAC</i>	LSC	39	591	35		
9	<i>rps12*</i>	LSC	114		232		26
10	<i>clpP</i>	LSC	71	840	292	638	231
11	<i>petB</i>	LSC	6	756	642		
12	<i>petD</i>	LSC	8	743	475		
13	<i>rpl16</i>	LSC	9	969	399		
14	<i>ndhB</i>	IR	777	684	756		
15	<i>trnI-GAU</i>	IR	37	951	35		
16	<i>trnA-UGC</i>	IR	38	817	35		
17	<i>ndhA</i>	SSC	553	1073	540		

35
36

¹Numbers represent the gene size (bp) in *A. polymorpha*; numbers in parentheses represent the gene size in *L. officinale*. *The intron length of *trnK-UUU* includes *matK*.

Table S5: Codon-anticodon recognition patterns and codon usage in the CP genomes of *A. polymorpha* and *L. officinale*.

Amino acid	Codon	<i>A. polymorpha</i>		tRNA
		Count	RSCU ¹	
Phe	UUU	798	1.32	
Phe	UUC	411	0.68	trnF-GAA
Leu	UUA	749	1.96	trnL-UAA
Leu	UUG	450	1.18	trnL-CAA
Leu	CUU	483	1.26	
Leu	CUC	154	0.4	
Leu	CUA	316	0.83	trnL-UAG
Leu	CUG	141	0.37	
Ile	AUU	875	1.45	
Ile	AUC	340	0.56	trnI-GAU
Ile	AUA	595	0.99	trnI-CAU
				trn(f)M-CAU
Met	AUG	512	1	trnM-CAU
Val	GUU	474	1.51	
Val	GUC	147	0.47	trnV-GAC
Val	GUA	458	1.46	trnV-UAC
Val	GUG	176	0.56	
Ser	UCU	463	1.75	
Ser	UCC	251	0.95	trnS-GGA
Ser	UCA	294	1.11	trnS-UGA
Ser	UCG	167	0.63	
Pro	CCU	354	1.57	
Pro	CCC	163	0.72	trnS-GCU
Pro	CCA	237	1.05	
Pro	CCG	146	0.65	
Thr	ACU	467	1.65	trnP-UGG
Thr	ACC	210	0.74	
Thr	ACA	333	1.18	
Thr	ACG	123	0.43	trnT-GGU
Ala	GCU	569	1.81	trnT-UGU
Ala	GCC	198	0.63	
Ala	GCA	341	1.09	
Ala	GCG	147	0.47	trnG-UCC
Tyr	UAU	644	1.61	trnA-UGC

Tyr	UAC	157	0.39	
Stop	UAA	43	1.61	
Stop	UAG	21	0.79	trnY-GUA
His	CAU	389	1.51	
His	CAC	126	0.49	
Gln	CAA	558	1.49	
Gln	CAG	191	0.51	
Asn	AAU	722	1.51	trnH-GUG
Asn	AAC	233	0.49	trnQ-UUG
Lys	AAA	774	1.5	
Lys	AAG	256	0.5	
Asp	GAU	677	1.59	trnN-GUU
Asp	GAC	172	0.41	trnK-UUU
Glu	GAA	801	1.49	
Glu	GAG	273	0.51	
Cys	UGU	175	1.51	trnD-GUC
Cys	UGC	57	0.49	trnE-UUC
Stop	UGA	16	0.6	
Trp	UGG	377	1	
Arg	CGU	298	1.4	trnC-GCA
Arg	CGC	87	0.41	trnW-CCA
Arg	CGA	299	1.41	trnR-ACG
Arg	CGG	109	0.51	
Ser	AGU	327	1.23	
Ser	AGC	89	0.34	
Arg	AGA	359	1.69	trnR-UCU
Arg	AGG	124	0.58	
Gly	GGU	530	1.35	
Gly	GGC	176	0.45	trnG-GCC
Gly	GGA	591	1.51	trnG-UCC
Gly	GGG	268	0.68	

39 ¹RSCU, relative synonymous codon usage.

Table S6: Details of palindromic repeats present in the LSC region of the CP genomes of *A. polymorpha* and *L. officinale*.

Species	Position	Loop (bp)	Position	Repeat unit length (bp)	Repeat unit sequence
<i>A. polymorpha</i>	IGS ¹ (<i>trnH-GUG, psbA</i>)	6	IGS (<i>rnH-GUG, psbA</i>)	31	TGAAATATAAAAAGAAGAAATACCGCCCTTTG
	IGS (<i>psbM, trnD-GUC</i>)	490	IGS (<i>trnE-UUC, trnT-GGU</i>)	39	AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGTCG
	IGS (<i>trnT-GGU, psbD</i>)	62	IGS (<i>trnT-GGU, psbD</i>)	26	TTATCATCCTACTAAATTAGAATTAA
	IGS (<i>psbE, petL</i>)	0	IGS (<i>psbE, petL</i>)	32	TCTTTCTTACTTTACTTGTGCTTGTCACTGT
	IGS (<i>psbT, psbN</i>)	3	IGS (<i>psbT, psbN</i>)	21	TTGAAGTAATGACTCCCCCAA
<i>L. officinale</i>	IGS (<i>trnH-GUG, psbA</i>)	13	IGS (<i>rnH-GUG, psbA</i>)	24	AAAAAAAGCAATACCGCCCTTTG
	IGS (<i>psbM, trnD-GUC</i>)	490	IGS (<i>trnE-UUC, trnT-GGU</i>)	39	AAAAGGGAAAGATGATTGATGTACTTATTGAATCTGTCG
	IGS (<i>trnT-GGU, psbD</i>)	62	IGS (<i>trnT-GGU, psbD</i>)	26	TTATCATCCTACTAAATTAGAATTAA
	IGS (<i>psbE, petL</i>)	10	IGS (<i>psbE, petL</i>)	27	TCTTTCTTACTTTACTTGTGCTTGTCA
	IGS (<i>psbT, psbN</i>)	3	IGS (<i>psbT, psbN</i>)	21	TTGAAGTAATGAGTCCCCCAA
	Intron (<i>petB, petB</i>)	9	Intron (<i>petB, petB</i>)	20	TAAATCGAAAGAAAGGTTG

Table S7: List of *A. polymorpha* and *L. officinale* accessions used in this study.

No.	Species	Collection information	Voucher number	Coordinates	GenBank accession number	Marker test
1	<i>L. officinale</i>	Ilcheongungjaebaeji, Bonghwa-gun, Gyeongsangbuk-do, Korea	KIOM201501014665	36°48'01.9"N 128°57'47.2"E		o
2		Subi-myeon, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201701018797	36°47'39.8"N 129°12'19.6"E		o
3		Subi-myeon, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201701018796	36°47'39.8"N 129°12'19.6"E		o
4		Yeongyang-eup, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201501014665	36°40'32.0"N 129°06'46.2"E		o
5		Yeongyang-eup, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM200601000207	36°40'32.0"N 129°06'46.2"E		o
6		Yeongyang-eup, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM200601000208	36°40'32.0"N 129°06'46.2"E		o
7		Subi-myeon, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201701018769	36°47'39.8"N 129°12'19.6"E	NC039760	o
8		Subi-myeon, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201701018770	36°47'39.8"N 129°12'19.6"E		o
9		Subi-myeon, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201701018771	36°47'39.8"N 129°12'19.6"E		o
10		Subi-myeon, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201701018795	36°47'39.8"N 129°12'19.6"E		o
11		Subi-myeon, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201701018617	36°47'39.8"N 129°12'19.6"E		o
12		Subi-myeon, Yeongyang-gun, Gyeongsangbuk-do, Korea	KIOM201301005868	36°47'39.8"N 129°12'19.6"E		o
13	<i>A. polymorpha</i>	Wonju-si, Gangwon-do, Korea	KIOM201501014664	37°20'17.6"N 128°02'27.5"E	MH260705	o
14		Seoha-myeon, Hamyang-gun, Gyeongsangnam-do, Korea	KIOM201501015521	35°38'39.7"N 127°44'12.2"E		o

15	Yeonpung-myeon, Goesan-gun, Chungcheongbuk-do, Korea	KIOM201701020170	36°42'55.1"N 128°03'07.5"E	o
16	Changjuk-dong, Taebaek-si, Gangwon-do, Korea	KIOM201701020255	37°13'50.7"N 128°56'17.1"E	o
17	Wicheon-myeon, Geochang-gun, Gyeongsangnam-do, Korea	KIOM201501015376	35°42'53.8"N 127°47'45.7"E	o
18	Gohan-eup, Jeongseon-gun, Gangwon-do, Korea	KIOM201501015172	37°12'14.3"N 128°54'31.1"E	o
19	Dunnae-myeon, Hoengseong-gun, Gangwon-do, Korea	KIOM201001003015	37°29'28.6"N 128°14'52.1"E	o
20	Jinbu-myeon, Pyeongchang-gun, Gangwon-do, Korea	KIOM200901002083	37°47'58.0"N 128°32'35.2"E	o
21	Jinbu-myeon, Pyeongchang-gun, Gangwon-do, Korea	KIOM200901002084	37°47'58.0"N 128°32'35.2"E	o

44

Table S8: List of primers used for the validation of the CP genome sequence of *A. polymorpha*.

Primer name	Primer sequence (5'→3')	Junction
LIGLF	GGATCCTCGCGGACAGAAAA	LSC/IRa
LIGLR	CGTGTCTGGTACTGCATGGT	
LIGISF	ACAATTCTAACTAGCCCTAATGGTC	IRa/SSC
LIGISR	TCAAATCTGTAGGTAGTGGCG	
LIGSIF	ACGCCTTGAAACACGCATAG	SSC/IRb
LIGSIR	AGAGGTTGAACAGAAAATAGACCG	
LIGILF	GATCTGCAGGGTCCCAAATGA	IRb/LSC
LIGILR	ACTTGATCCAACGGAGCCC	

45

Table S9: List of CP genomes downloaded from NCBI for phylogenetic analysis.

Order	Family	No.	Taxon	GenBank accession number
Apiales	Apiaceae	1	<i>Angelica acutiloba</i>	NC029391.1
		2	<i>Angelica dahurica</i>	NC029392.1
		3	<i>Angelica decursiva</i>	KT781591.1
		4	<i>Angelica gigas</i>	NC029393.1
		5	<i>Angelica nitida</i>	MF594405.1
		6	<i>Arracacia xanthorrhiza</i>	KY117235.1
		7	<i>Glehnia littoralis</i>	KT153022.1
		8	<i>Ledebouriella seseloides</i>	KT153021.1
		9	<i>Peucedanum insolens</i>	NC033344.1
		10	<i>Seseli montanum</i>	KM035851.1
		11	<i>Ligusticum tenuissimum</i>	NC029394.1
		12	<i>Ligusticum officinale</i>	NC039760.1
		13	<i>Daucus carota</i>	NC008325.1
		14	<i>Anthriscus cerefolium</i>	NC015113.1
	Araliaceae	15	<i>Tiedemannia filiformis</i> subsp. <i>greenmannii</i>	HM596071.1
		16	<i>Coriandrum sativum</i>	NC029850.1
		17	<i>Bupleurum latissimum</i>	NC033346.1
		18	<i>Bupleurum falcatum</i>	NC027834.1
		20	<i>Anethum graveolens</i>	NC029470.1
		21	<i>Petroselinum crispum</i>	NC015821.1
		22	<i>Foeniculum vulgare</i>	NC029469.1
		23	<i>Eleutherococcus senticosus</i>	JN637765.1
	Dipsacales	24	<i>Fatsia japonica</i>	NC027685.1
		25	<i>Kalopanax septemlobus</i>	NC022814.1
		26	<i>Metapanax delavayi</i>	NC022812.1
		27	<i>Schefflera delavayi</i>	NC022813.1
		28	<i>Aralia undulata</i>	NC022810.1
		29	<i>Panax ginseng</i>	AY582139.1
		30	<i>Panax notoginseng</i>	NC026447.1
		31	<i>Adoxa moschatellina</i>	NC034792.1
	Adoxaceae	32	<i>Tetraodoxa omeiensis</i>	NC034793.1