

Table 1 Suppl. Genes used in the phylogenetic analyses.

Groups of genes	Name of genes
Proteins of small ribosomal subunit	<i>rps11, rps14, rps18, rps4, rps8</i>
Proteins of large ribosomal subunit	<i>rpl32, rpl33</i>
Subunits of NADH dehydrogenase	<i>ndhC, ndhE, ndhJ, ndhD</i>
Subunits of photosystem I	<i>psaA, psaB</i>
Subunits of photosystem II	<i>psbA, psbC, psbD, psbH, psbK</i>
Subunits of cytochrome <i>b/f</i> complex	<i>petA</i>
Subunits of ATP synthase	<i>atpE, atpH, atpI, atpB</i>
Maturase	<i>matK</i>
Proteins of large ribosomal subunit	<i>rpl14</i>

Table 2 Suppl. List of genes obtained from the *Arctium lappa* and *A. tomentosum* chloroplast genome sequences (^a - genes containing two introns, ^b - genes containing a single intron, ^c - genes containing three introns, ^d - two gene copies in the inverted repeat (IRs), ^e - four gene copies in the IRs).

Groups of genes	<i>Arctium lappa</i>	<i>Arctium tomentosum</i>
Ribosomal RNAs	<i>rrn23^d, 16^d, 5^d, 4.5^d</i>	<i>rrn23^d, 16^d, 5^d, 4.5^d</i>
Transfer RNAs	<i>trnA-UGC^d, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC, trnH-GUG, trnI-CAT^d, trnI-CAU^d, trnI-GAU^d, trnK-UUU, trnL-CAA^d, trnL-CAA^d, trnM-CAU, trnM-CAT, trnN-GTT^d, trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU^d, trnP-TGG, trnQ-TTG, trnR-ACG^d, trnR-TCG, trnP-UGG, trnQ-UUG, trnR-ACG^d, trnR-UCU, trnS-GCU, trnS-GCU, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC^d, trnV-UAC, trnW-CCA, trn-GUA</i>	<i>trnL-TAG, trnC-GCA, trnD-GTC, trnE-TTC, trnF-GAA, trnG-GCC, trnH-GTG, trnI-CAT^d, trnI-CAU^d, trnI-GAU^d, trnK-UUU, trnL-CAA^d, trnL-CAA^d, trnM-CAU, trnM-CAT, trnN-GTT^d, trnP-TGG, trnQ-TTG, trnR-ACG^d, trnR-TCG, trnP-UGG, trnQ-UUG, trnR-ACG^d, trnR-UCU, trnS-GCU, trnS-GCU, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC^d, trnV-UAC, trnW-CCA, trnY-GTA</i>
Proteins of small ribosomal subunit	<i>rps2,3,4,7^d, 8, 11, 12^d, 14, 15, 16^a, 18, 19</i>	<i>rps2, 3,4,7^d, 8, 12, 11, 14, 15, 16, 18, 19</i>
Proteins of large ribosomal subunit	<i>rpl2^d, 14, 16, 20, 22, 23^d, 32, 33, 36</i>	<i>rpl2^{db}, 14, 16, 20, 22, 23^d, 32, 33, 36</i>
Subunits of RNA polymerase	<i>rpoA, B, C1^a, C2</i>	<i>rpoA, B, C1, C2</i>
Subunits of NADH dehydrogenase	<i>ndhA^a, B^{ad}, C, D, E, F, G, H, I, J, K</i>	<i>ndhA^b, B^d, C, D, E, F, G, H, I, J, K</i>
Subunits of photosystem I	<i>psaA, B, C, I, J</i>	<i>psaA, B, C, I, J</i>
Subunits of photosystem II	<i>psbA,B, C, E, D, F, H, I, J, K, L, M, N, T, Z</i>	<i>psbA,B, C, E, D, F, H, I, J, K, M, N, T, Z</i>
Large subunit of Rubisco	<i>rbcL</i>	<i>rbcL</i>
Subunits of cytochrome <i>b/f</i> complex	<i>petA, B, D, G, L, N</i>	<i>petA, B, D, G, L, N</i>
Subunits of ATP synthase	<i>atpA, B, E, F, H, I</i>	<i>atpA, B, E, F, H, I</i>
Acetyl-CoA carboxylase	<i>accD</i>	<i>accD</i>
C-type cytochrom synthesis gene	<i>ccsA</i>	<i>ccsA</i>
Maturase	<i>matK</i>	<i>matK</i>
Protease	<i>clpP^c</i>	<i>clpP^b</i>
Envelope membrane protein	<i>cemA</i>	<i>cemA</i>
Conserved hypothetical chloroplast reading frames	<i>ycf1^d, 2^d, 3^a, 4, 15^d</i>	<i>ycf1^{dc}, 2^{db}, 3^a, 4, 15^e, 15^d</i>
Translational initiation factor	<i>infA</i>	<i>infA</i>
Hypothetical protein	<i>orf42-D2, orf42</i>	

Table 3 Suppl. Characteristics of coding genes, including introns and exons, in the chloroplast genomes of *Arctium lappa* and *A. tomentosum*.

Species	Gene	Exon I [bp]	Intron I [bp]	Exon II [bp]	Intron II [bp]	Exon III [bp]	Intron III [bp]	Exon IV [bp]
<i>A. lappa</i>	<i>rps16</i>	225	841	42				
	<i>rpoC1</i>	434	719	1651				
	<i>clpP</i>	228	1	26	609	292	795	60
	<i>rpl2</i>	446	653	391				
	<i>ndhA</i>	518	1070	562				
	<i>ndhB</i>	782	643	778				
	<i>rpl2-D2</i>	391	653	446				
	<i>ndhB-D2</i>	778	643	782				
	<i>ycf3</i>	141	752	221	699	124		
	<i>atpF</i>	154	683	425				
<i>A. tomentosum</i>	<i>rpoC1</i>	431	735	1637				
	<i>ycf3</i>	126	1043	229	1122	152		
	<i>clpP</i>	307	1146	225				
	<i>rpl2</i>	393	1489	469				
	<i>ycf2</i>	4974	52	1846				
	<i>ycf1</i>	1284	46	1637	34	1268	52	874
	<i>ndhA</i>	552	1059	538				
	<i>ycf2-D2</i>	4977	6872	1846				
	<i>rpl2-D2</i>	390	630	469				
	<i>rps12</i>	231	778	10				
	<i>rps12-D2</i>	10	537	231				

Table 4 Suppl. The single sequence repeat (SSR) position in the chloroplast genomes of *Arctium lappa* and *A. tomentosum*.

	No.	SSR type	SSR	Size	Start	End
<i>A. lappa</i>	1	p1	(T)11	11	1976	1986
	2	p1	(C)11	11	5398	5408
	3	p2	(AT)7	14	26375	26388
	4	p1	(T)12	12	34477	34488
	5	p1	(T)12	12	49996	50007
	6	p1	(T)14	14	62993	63006
	7	p1	(A)11	11	65256	65266
	8	p2	(TA)7	14	67582	67595
	9	p1	(A)11	11	77558	77568
	10	p1	(T)14	14	80237	80250
	11	p1	(A)13	13	107301	107313
	12	p1	(T)12	12	107764	107775
	13	p1	(T)14	14	111731	111744
	14	p1	(A)12	12	128814	128825
	15	p1	(T)13	13	129276	129288
<i>A. tomentosum</i>	1	p1	(T)11	11	1976	1986
	2	p1	(C)11	11	5398	5408
	3	p1	(A)10	10	13125	13134
	4	p1	(A)10	10	18174	18183
	5	p1	(T)10	10	22978	22987
	6	p2	(AT)6	12	26373	26384
	7	p1	(T)12	12	34472	34483
	8	p1	(A)10	10	43735	43744
	9	p1	(T)10	10	46171	46180
	10	p1	(T)12	12	49991	50002
	11	p1	(T)10	10	59010	59019
	12	p1	(T)14	14	62912	62925
	13	p1	(A)11	11	65175	65185
	14	p2	(TA)7	14	67501	67514
	15	p1	(A)10	10	70773	70782
	16	p1	(T)10	10	76662	76671
	17	p1	(A)10	10	77477	77486
	18	p1	(T)10	10	77723	77732
	19	p1	(T)14	14	80155	80168
	20	p1	(A)13	13	107224	107236
	21	p1	(T)12	12	107687	107698
	22	p1	(A)13	13	124767	124779
	23	p1	(A)12	12	128735	128746
	24	p1	(T)13	13	129197	129209

Table 5 Suppl. Long repeat sequences in *Arctium lappa* chloroplast genome (F - forward; P - palindromic; IGS - intergenic space).

ID	Repeat start 1	Type	Size [bp]	Repeat start 2	Mismatch [bp]	E-value	Gene	Region
1	7338	F	32	7363	3	1.55E-03	IGS	LSC
2	8476	F	32	34630	3	1.55E-03	<i>trnS</i> -GCU; <i>trnS</i> -GCU	LSC
3	8478	P	30	44604	0	5.69E-09	<i>trnS</i> -GCU; <i>trnS</i> -UGA	LSC
4	11731	F	33	30440	3	4.27E-04	IGS	LSC
5	23725	F	31	23756	0	1.42E-09	IGS	LSC
6	34631	P	30	44604	3	2.03E-02	<i>trnS</i> -UGA; <i>trnS</i> -GCU	LSC
7	42850	F	42	119415	0	3.39E-16	IGS	LSC; SSC
8	42851	F	41	97858	1	6.63E-13	IGS	LSC; IRa
9	42851	P	41	138691	1	6.63E-13	IGS	LSC; IRb
10	42854	P	30	75314	3	2.03E-02	<i>petB</i> ; IGS	LSC
11	42856	F	35	94815	3	3.20E-05	IGS	LSC; IRa
12	42856	P	35	141740	3	3.20E-05	IGS	LSC; IRb
13	46175	P	37	46178	2	2.67E-08	IGS	LSC
14	58527	R	31	58530	1	5.25E-07	IGS	LSC
15	58551	F	116	58570	0	9.51E-61	IGS	LSC
16	58846	F	30	58866	2	2.03E-05	IGS	LSC
17	58888	R	31	58891	3	5.63E-03	IGS	LSC
18	67224	F	31	99035	3	5.63E-03	IGS	LSC; IRa
19	67224	P	31	137525	3	5.63E-03	IGS	LSC
20	73968	P	48	73968	0	8.28E-20	<i>psbN</i> ; <i>psbN</i>	LSC; IRb
21	75314	P	30	119419	3	2.03E-02	<i>petB</i> ; IGS	LSC; SSC
22	87477	F	30	149084	3	2.03E-02	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
23	91063	F	30	91081	3	2.03E-02	<i>ycf2</i> ; <i>ycf2</i>	IRa
24	91063	P	30	145479	3	2.03E-02	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
25	91078	F	46	91096	3	1.77E-11	<i>ycf2</i> ; <i>ycf2</i>	IRa
26	91078	P	46	145448	3	1.77E-11	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
27	91081	P	30	145497	3	2.03E-02	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
28	91094	F	30	91112	1	2.03E-06	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
29	91094	P	30	145448	1	2.03E-06	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
30	91096	P	46	145466	3	1.77E-11	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
31	91112	P	30	145466	1	2.03E-06	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
32	94815	F	35	119421	3	3.20E-05	IGS	IRa; SSC
33	97858	F	41	119416	1	6.63E-13	IGS	IRa; SSC
34	98214	F	31	111979	3	5.63E-03	IGS	IRa; SSC
35	106941	F	34	106973	3	1.17E-04	IGS	IRa
36	106941	P	34	129584	3	1.17E-04	IGS	IRa; IRb
37	106973	P	30	129619	2	2.85E-04	IGS	IRa; IRb
38	111979	P	31	138345	3	5.63E-03	IGS	SSC; IRb
39	113421	F	34	113455	3	3.20E-05	IGS	SSC
40	119416	P	41	138691	1	6.63E-13	IGS	SSC; IRb
41	119421	P	35	141740	3	3.20E-05	IGS	SSC; IRb
42	122892	P	31	125256	3	5.63E-03	IGS	SSC
43	126787	F	46	126832	3	1.77E-11	IGS	SSC
44	129584	F	33	129615	3	1.17E-04	IGS	IRb
45	145448	F	46	145466	3	1.77E-11	<i>ycf2</i> ; <i>ycf2</i>	IRb
46	145479	F	30	145497	3	2.03E-02	<i>ycf2</i> ; <i>ycf2</i>	IRb

Table 6 Suppl. Long repeat sequences in *Arctium tomentosum* chloroplast genome (F - forward; P - palindromic; IGS - intergenic space).

ID	Repeat start 1	Type	Size [bp]	Repeat start 2	Mismatch [bp]	E-value	Gene	Region
1	8475	F	32	34624	-3	1.55E-03	<i>trnS</i> -GCT, IGS; <i>trnS</i> -TGA, IGS	LSC
2	11730	F	39	30434	-3	1.75E-07	IGS	LSC
3	23724	F	31	23755	0	1.42E-09	<i>rps2</i> , IGS; IGS	LSC
4	42845	F	41	97780	-1	6.62E-13	<i>ycf3</i> ; IGS	LSC; IRa
5	42850	F	35	94737	-3	3.20E-05	<i>ycf3</i> ; IGS	LSC; IRa
6	67142	F	31	98957	-3	5.63E-03	IGS	LSC; IRa
7	75232	F	30	117063	-3	2.03E-02	<i>petB</i> , IGS; <i>ndhA</i>	LSC; SSC
8	87399	F	30	149004	-3	2.03E-02	<i>ycf2</i> (CDS)	IRa; IRb
9	106863	F	34	106895	-3	1.17E-04	<i>rrn4.5S</i>	IRa
10	109634	F	46	109679	-3	1.77E-11	<i>ycf1</i>	SSC
11	117055	F	41	138611	-1	6.62E-13	<i>ndhA</i> ; IGS	SSC; IRb
12	117056	F	35	141660	-3	3.20E-05	<i>ndhA</i> ; IGS	SSC; IRb
13	117071	F	31	138627	-3	5.63E-03	<i>ndhA</i> ; IGS	SSC; IRb
14	123021	F	35	123056	-3	3.20E-05	IGS	SSC
15	124501	F	31	138265	-3	5.63E-03	IGS	SSC; IRb
16	129504	F	33	129535	-3	1.17E-04	<i>rrn4.5S</i>	IRb
17	129507	F	30	129539	-2	2.85E-04	<i>rrn4.5S</i>	IRb
18	8477	P	30	44598	0	5.69E-09	<i>trnS</i> -GCT, IGS; <i>trnS</i> -GGA, IGS	LSC
19	34625	P	30	44598	-3	2.03E-02	<i>trnS</i> -TGA, IGS; <i>trnS</i> -GGA, IGS	LSC
20	42844	P	42	117055	0	3.39E-16	<i>ycf3</i> ; <i>ndhA</i>	LSC; SSC
21	42845	P	41	138611	-1	6.62E-13	<i>ycf3</i> ; IGS	LSC; IRb
22	42848	P	30	75232	-3	2.03E-02	<i>ycf3</i> ; <i>petB</i> (CDS), IGS	LSC
23	42850	P	35	141660	-3	3.20E-05	<i>ycf3</i> ; IGS	LSC; IRb
24	67142	P	31	137445	-3	5.63E-03	IGS	LSC; IRb
25	90985	P	30	145399	-3	2.03E-02	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
26	91000	P	46	145368	-3	1.77E-11	<i>ycf2</i>	IRa; IRb
27	91003	P	30	145417	-3	2.03E-02	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
28	91016	P	30	145368	-1	2.03E-06	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
29	91018	P	46	145386	-3	1.77E-11	<i>ycf2</i>	IRa; IRb
30	91034	P	30	145386	-1	2.03E-06	<i>ycf2</i> ; <i>ycf2</i>	IRa; IRb
31	94737	P	35	117056	-3	3.20E-05	IGS; <i>ndhA</i>	IRa; SSC
32	97780	P	41	117055	-1	6.62E-13	IGS; <i>ndhA</i>	IRa; SSC
33	98136	P	31	124501	-3	5.63E-03	IGS	IRa; SSC
34	106863	P	34	129504	-3	1.17E-04	<i>rrn4.5S</i>	IRa; IRb
35	106895	P	30	129539	-2	2.85E-04	<i>rrn4.5S</i>	IRa; IRb
36	111225	P	31	113589	-3	5.63E-03	<i>ycf1</i> ; IGS	SSC

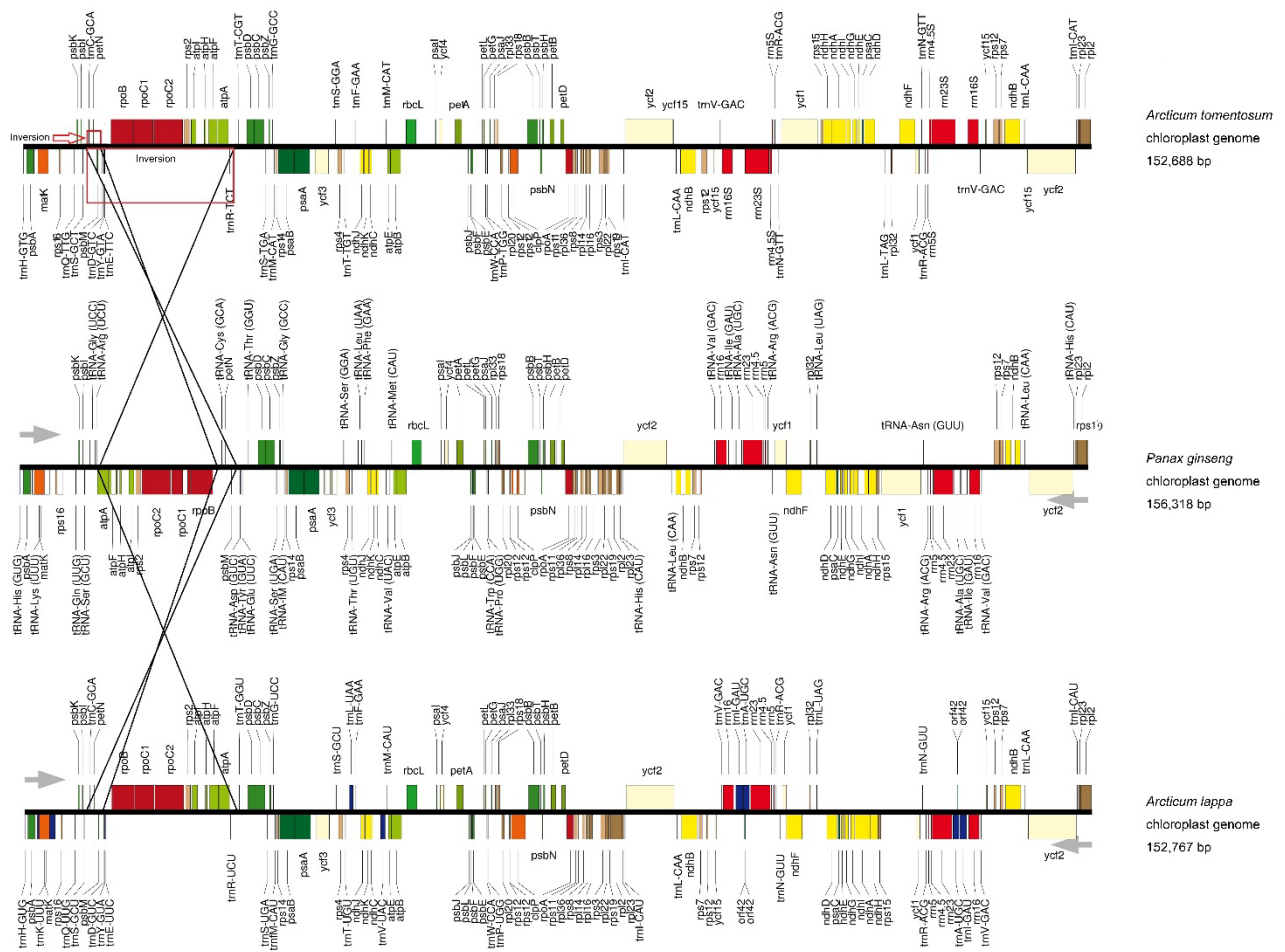


Fig. 1 Suppl. Two inversions in plastome maps.

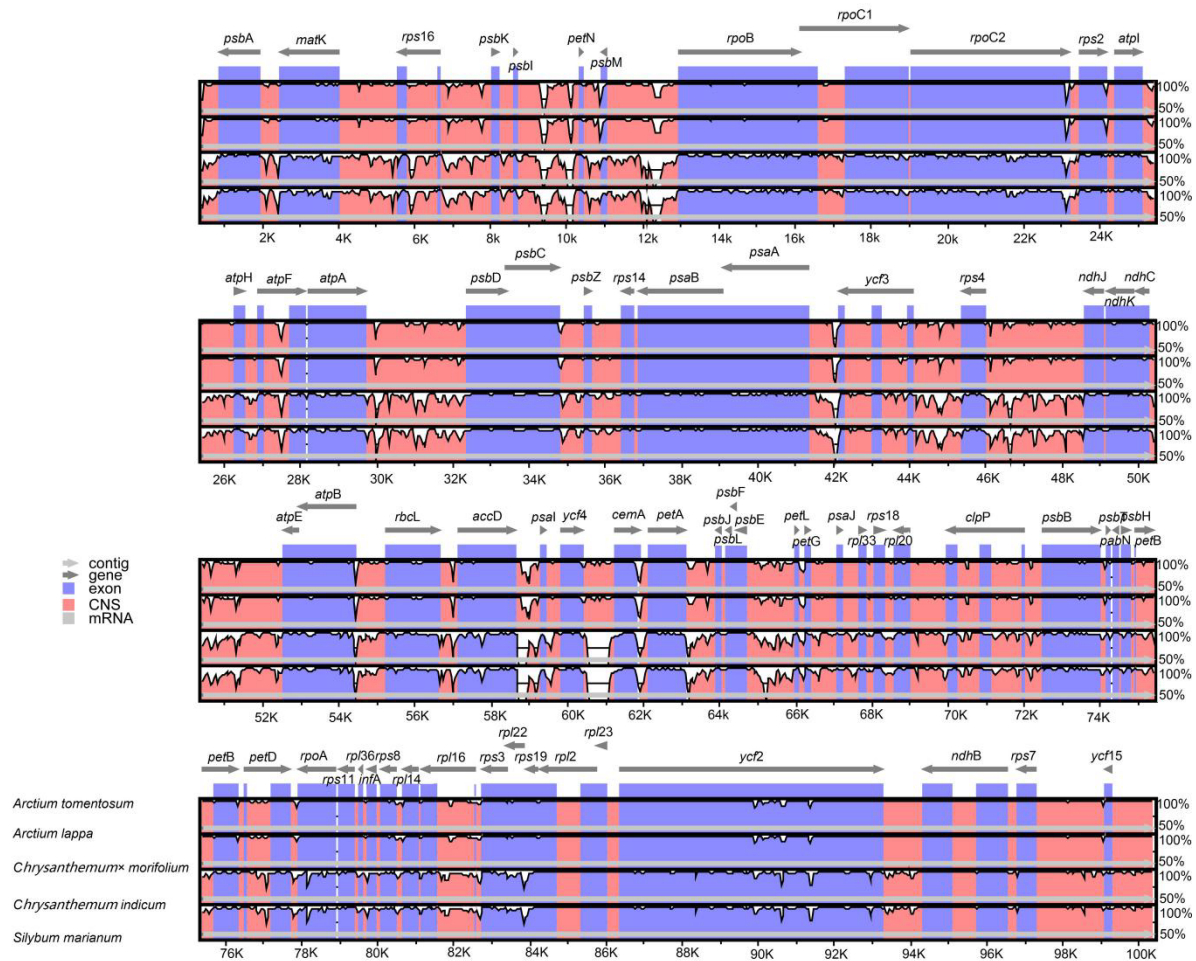


Fig. 2 Suppl. Comparison of chloroplast genomes from five Asteraceae species using mVISTA.