

The complete plastid genome of *Ferula communis* (Apiaceae): structural features and phylogenetic insights

Mergi Daba Dinka ^{a*}, Madalina Oana Popa ^b, Paulina Trzeciak ^c, Krzysztof Spalik ^a, John Soghigian ^d and Marcin Piwczynski ^e

Table S1. Voucher information for *Ferula* sample used in this study

Species	Locality	Coordinates	Collection date	Collector(s)	Voucher code	Herbarium / Repository
<i>Ferula communis</i>	Menorca, Spain	39°54'25.9"N, 4°15'30.0"E	16 May 2016	Piwczynski & Corral	KEIB_AP_00184	Herbarium, Dept. of Ecology and Biogeography, Faculty of Biological and Veterinary Sciences, Nicolaus Copernicus University, Toruń

Table S2. The corresponding GenBank reference numbers for the plastid genomes used in the comparative analysis and for outgroups were utilized to determine the position of *Ferula communis* in the phylogenetic tree.

Ferula akitschkensis B.Fedtsch. ex Koso-Pol., NC_071166; *Ferula bungeana* Kitag., NC_067786; *Ferula canescens* (Ledeb.) Ledeb., NC_071171; *Ferula caspica* M.Bieb., NC_071172; *Ferula communis* L., PP693041; *Ferula conocaula* Korovin, NC_071176; *Ferula dissecta* (Ledeb.) Ledeb., NC_071174; *Ferula equisetacea* Koso-Pol., NC_067556; *Ferula fedtschenkoana* Koso-Pol., NC_067563; *Ferula feruloides* (Steud.) Korov., NC_071167; *Ferula gigantea* B.Fedtsch., NC_067562; *Ferula gracilis* (Ledeb.) Ledeb., NC_071169; *Ferula hexiensis* K.M.Shen, NC_071175; *Ferula karelinii* Bunge, NC_067559; *Ferula kelifi* Korovin, NC_067564; *Ferula kingdon-wardii* H.Wolff, NC_071179; *Ferula kirialovii* Pimenov, NC_071178; *Ferula lehmannii* Boiss., NC_071170; *Ferula leiophylla* Korovin, NC_071164; *Ferula litwinowiana* Koso-Pol., NC_067565; *Ferula olivacea* (Diels) H.Wolff, NC_067566; *Ferula oopoda* (Boiss. & Buhse) Boiss., NC_067567; *Ferula ovina* Boiss., NC_067558; *Ferula renardii* (Regel & Schmalh.) Pimenov, NC_067568; *Ferula sibirica* Willd., NC_067560; *Ferula sinkiangensis* K.M.Shen, OP324722; *Ferula songarica* Pall. ex Willd., NC_071168; *Ferula syreitschikowii* Koso-Pol., NC_071177; *Ferula teterrima* Kar. & Kir., NC_071165; *Ferula transiliensis* (Regel & Herder) Pimenov, NC_067561.

Outgroup:

Daucus carota L., NC_008325; *Anthriscus sylvestris* L., MT561042, used for comparative analysis on IR boundaries.

Daucus carota L., NC_008325; *Torilis scabra*, MT561029; used for Phylogenetic analysis.

Table S3. Gene composition of the plastid genome of *Ferula communis*, categorized by function, based on the complete genome assembly.

Gene category	Functional groups	Gene Names
Photosynthesis	Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	Subunit of NADH-dehydrogenase	<i>ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Cytochrome b/f complex	<i>petA, petB, petD, petG, petL, petN</i>
	RUBISCO	<i>rbcL</i>
Self-replication	Large subunit of ribosome	<i>rpl14, rpl16, rpl2, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	Small subunit of ribosome	<i>rps11, rps12, rps12, rps14, rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps7, rps8</i>
	rRNA Genes	<i>rrn4.5S, rrn5S, rrn16S, rrn23S</i>
	tRNA Genes	<i>trnC-GCA, trnD-GTC, trnE-TTC, trnF-GAA, trnFM-CAT, trnG-GCC, trnH-GTG, trnM-CAT, trnP-TGG, trnQ-TTG, trnR-TCT, trnS-GCT, trnS-GGA, trnS-TGA, trnT-GGT, trnT-TGT, trnW-CCA, trnY-GTA, trnL-TAG, trnI-CAT, trnI-GAT, trnL-CAA, trnN-GTT, trnR-ACG, trnV-GAC</i>
Other functions	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Envelop membrane protein	<i>cemA</i>
	Protease	<i>clpP</i>
	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
Unknown functions	Conserved open reading frames	<i>yef1, yef2, yef3, yef4, yef15</i>

Table S4. Distribution of intron and exon lengths in genes of the *Ferula communis* plastid genome with group classifications. For trans-spliced genes such as *rps12*, the 5' exon (Exon I) is physically separated from the 3' exons (Exon II and Exon III), and splicing occurs post-transcriptionally; introns are reported only between contiguous exons, so Intron II is not present in *rps12*.

Group	Gene	Strand	Start	End	Exon I	Intron I	Exon II	Intron II	Exon III
Group I:	<i>trnL-UAA</i>	+	48521	49117	35	512	50		
Group II:									
A1	<i>trnA-UGC</i>	+	109521	110405	37	812	36		
	<i>trnV-UAC</i>	-	52676	53316	39	567	35		
	<i>trnK-UUU</i>	+	108435	109456	32	950	40		
	<i>clpP</i>	-	71103	73165	71	846	291	626	229
A2	<i>atpF</i>	-	11992	13270	145	733	401		
	<i>rpl2</i>	+	165080	166555	391	651	434		
B1	<i>PetB</i>	+	76133	77529	6	737	654		
	<i>rps16</i>	-	4856	5948	40	856	197		
	<i>rps12</i>	-	70815	99039	114	536	26	-	232
	<i>rpoC1</i>	-	21394	24167	432	737	1605		
	<i>yef3</i>	-	43589	45601	124	727	230	779	153
B2	<i>rps16</i>	-	4856	5948	40	856	197		
	<i>ndhB</i>	-	95209	97423	775	682	758		
	<i>ndhA</i>	-	127246	129313	479	1048	541		
	<i>ndhB</i>	+	154623	156837	775	682	758		
	<i>trnG-UCC</i>	+	9310	10083	32	682	60		
	<i>trnE-UUC</i>	+	108435	109456	32	950	40		
	<i>trnA-UGC</i>	+	109521	110405	37	812	36		
	<i>trnE-UUC</i>	-	142590	143611	32	950	40		
	<i>trnA-UGC</i>	-	141641	142525	37	812	36		
	<i>yef2</i>	+	87594	93857	4426	36	1802		
	<i>yef2</i>	-	158189	164452	4426	36	1802		

Table S5. REPuter analysis results of forward and palindromic repeats in the plastid genome of *Ferula communis*: detailed insights on repeat length, location, and genomic regions.

Repeat length	Start pos. 1 st repeat unit	Repeat type	Repeat length	Start pos. 2 nd repeat unit	Strand	E-Value	Region	Occurrence
88	64993	F	88	65018	-1	2.15E-41	IGS/LSC	<i>petA-psbJ</i>
79	70905	P	79	104217	0	2.14E-38	IGS/LSC	<i>rpl20-rps12</i>
79	70905	F	79	147749	0	2.14E-38	IGS/LSC	<i>rpl20-rps12</i>
77	91970	F	77	91988	-1	7.91E-35		<i>ycf2</i>
77	91970	P	77	159980	-1	7.91E-35		<i>ycf2</i>
77	91988	P	77	159998	-1	7.91E-35		<i>ycf2</i>
77	159980	F	77	159998	-1	7.91E-35		<i>ycf2</i>
63	64993	F	63	65043	-1	1.74E-26	IGS/LSC	<i>petA-psbJ</i>
60	160004	F	60	160022	-1	1.06E-24		<i>ycf2</i>
53	91970	F	53	92006	-1	1.53E-20		<i>ycf2</i>
53	91970	P	53	159986	-1	1.53E-20		<i>ycf2</i>
53	92006	P	53	160022	-1	1.53E-20		<i>ycf2</i>
53	159986	F	53	160022	-1	1.53E-20		<i>ycf2</i>
42	160004	F	42	160040	-1	5.09E-14		<i>ycf2</i>
41	99076	F	41	127822	-1	1.99E-13	IGS/LSC	<i>rps12-ycf15</i>
41	127822	P	41	152928	-1	1.99E-13	Intron	<i>ndhA</i>
39	44786	F	39	127824	-1	3.03E-12	Intron	<i>ycf3</i>
38	64993	F	38	65068	-1	1.18E-11	IGS/LSC	<i>petA-psbJ</i>
35	91970	F	35	92024	-1	6.95E-10		<i>ycf2</i>
35	91970	P	35	159986	-1	6.95E-10		<i>ycf2</i>
35	92024	P	35	160040	-1	6.95E-10		<i>ycf2</i>
35	159986	F	35	160040	-1	6.95E-10		<i>ycf2</i>
34	113594	F	34	113626	-1	2.70E-09	IGS/SSC	<i>rrn4.5-rrn5</i>
34	113594	P	34	138385	-1	2.70E-09	IGS/SSC	<i>rrn4.5-rrn5</i>
34	113626	P	34	138417	-1	2.70E-09	IGS/SSC	<i>rrn4.5-rrn5</i>
34	138385	F	34	138417	-1	2.70E-09	IGS/SSC	<i>trnR-ACG-rrn5</i>
30	8666	P	30	46553	0	6.78E-09	IGS/LSC	<i>psbI-trnS-GCU</i>
30	30373	P	30	30373	0	6.78E-09	IGS/LSC	<i>petN-psbM</i>
33	44792	F	33	99084	-1	1.05E-08	Intron	<i>ycf3</i>
33	44792	P	33	152928	-1	1.05E-08	Intron	<i>ycf2</i>

Table S6: Summary of tandem repeats identified in the *Ferula communis* plastid genome, including key metrics: genomic positions (Indices, start–end in bp), period size, copy number, consensus length, match percentage, indel percentage, Score, base composition (absolute counts of A, C, G, T in the consensus sequence), and entropy scores as indicators of sequence complexity.

Indices	Period Size	Copy Number	Consensus Size	Match %	Indels %	Score	A	C	G	T	Entropy (0-2)
10169-10194	13	2	13	100	0	52	30	15	7	46	1.74
15963-16005	20	2	20	82	0	50	44	11	18	25	1.84
21168-21264	49	2	49	87	0	140	30	16	20	31	1.95
47507-47533	13	2	13	100	0	54	22	7	0	70	1.12
62963-62989	14	2	14	100	0	54	40	14	7	37	1.74
64981-65106	25	5	25	98	0	234	38	21	7	31	1.82
69018-69063	24	2	24	90	0	74	52	4	6	36	1.47
69154-69194	21	2	21	95	0	73	29	26	9	34	1.88
75430-75466	19	2	19	100	0	74	54	16	8	21	1.68
75985-76015	15	2	15	93	0	53	45	6	9	38	1.63
89529-89596	21	3	21	77	8	75	11	22	8	57	1.61
90704-90747	15	3	15	93	0	70	27	34	13	25	1.93
91964-92065	18	6	18	97	0	177	28	6	30	34	1.83
99969-100003	17	2	17	100	0	70	28	11	0	60	1.32
113595-113660	32	2	32	97	0	123	40	24	9	25	1.84
120492-120518	12	2	12	100	0	54	40	7	0	51	1.3
121104-121133	14	2	14	93	0	51	50	6	10	33	1.62
125183-125207	12	2	12	100	0	50	48	0	8	44	1.32
138386-138451	32	2	32	97	0	123	25	9	24	40	1.84
152043-152077	17	2	17	100	0	70	60	2	11	28	1.32
159981-160082	18	6	18	97	0	177	34	30	6	28	1.83
161299-161342	15	3	15	93	0	70	25	13	34	27	1.93
162468-162521	21	3	21	82	11	74	59	11	20	9	1.59
162451-162520	21	3	19	72	16	59	57	8	22	11	1.61

Table S7. Results of model comparisons for selection pressure analysis (dN/dS) on plastid genes with positively selected sites in compared *Ferula* species, including $\Delta\ln L$ values, degrees of freedom (df), and p-values for various evolutionary models.

Gene	Compared models	$\Delta\ln L$	df	P-value (95%)
<i>ccsA</i>	M1a vs M2a	48.33	2	3.200913e-11
	M7 vs M8	48.51	2	2.925414e-11
<i>ndhD</i>	M1a vs M2a	10.84	2	0.00443
	M7 vs M8	11.99	2	0.00249
<i>ndhF</i>	M1a vs M2a	63.72	2	1.456726e-14
	M7 vs M8	64.70	2	8.924287e-15
<i>ndhH</i>	M1a vs M2a	5.33	2	0.044
	M7 vs M8	5.71	2	0.0376
<i>petD</i>	M1a vs M2a	82.13	2	1.464526e-18
	M7 vs M8	82.71	2	1.095851e-18
<i>rpl32</i>	M1a vs M2a	56.14	2	6.446944e-13
	M7 vs M8	59.84	2	1.013699e-13

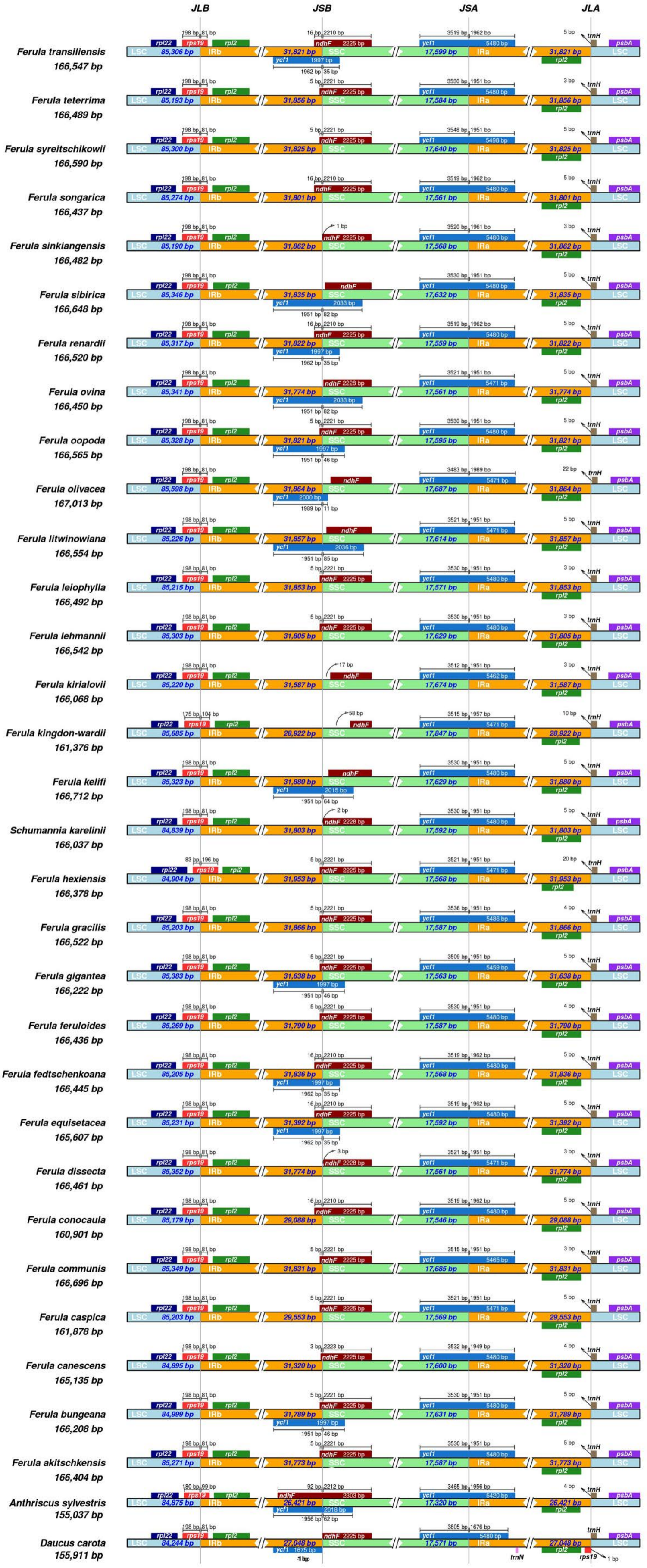
Table S8. A comprehensive evaluation of selection pressure (dN/dS) on plastid genes within compared *Ferula* species, detailing likelihood values (*lnL*), parameter estimates, and specific sites under selection as determined by models favors positive selection sites (M2a & M8).

Gene	Model	<i>LnL</i>	Estimates of parameters	Positively selected sites
<i>ccsA</i>	M2a	-1845.619	P: 0.61344 0.37745 0.00910 w: 0.16343 1.00000 26.97234	178 F 1.000**, 180 F 1.000**
	M8	-1845.599	P0 = 0.99086 p = 0.31261 q = 0.34088 (p1 = 0.00914) w = 26.82902	178 F 1.000**, 180 F 1.000**
<i>ndhD</i>	M2a	-2526.986	p: 0.98244 0.00000 0.01756 w: 0.09981 1.00000 6.40147	29 I 0.962*, 33 L 0.996**, 418 M 0.957*
	M8	-2526.997	P0 = 0.98282 p = 11.41714 q = 99.00000 (p1 = 0.01718) w = 6.50970	29 I 0.964*, 33 L 0.998**, 418M 0.961*
<i>ndhF</i>	M2a	-4123.008	P: 0.81069 0.18654 0.00278 w: 0.02199 1.00000 63.03113	571 V 0.952*, 734 L 0.970*, 735 I 1.000**, 736 L 1.000**
	M8	-4123.068	p0 = 0.99721 p = 0.01043 q = 0.03424 (p1 = 0.00279) w = 63.55268	571 V 0.987*, 734 L 0.995**, 735 I 1.000**, 736 L 1.000**
<i>ndhH</i>	M2a	-1908.999	p: 0.99558 0.00000 0.00442 w: 0.07318 1.00000 8.57190	
	M8	-1909.011	p0 = 0.99560 p = 7.94750 q = 99.00000 (p1 = 0.00440) w = 8.59831	298 V 0.966*
<i>petD</i>	M2a	-826.393	p: 0.97891 0.00000 0.02109 w: 0.05593 1.00000 35.71222	1 M 1.000**, 2 G 1.000**, 3 V 1.000**
	M8	-826.395	p0 = 0.97890 p = 5.91104 q = 99.00000 (p1 = 0.02110) w = 35.71868	1 M 1.000**, 2 G 1.000**, 3 V 1.000**
<i>rpl32</i>	M2a	-371.593	P: 0.77783 0.18494 0.03723 w: 0.17477 1.00000 61.86072	53 K 1.000**, 54 S 1.000**
	M8	-371.618	p0 = 0.96277 p = 0.42305 q = 0.82853 (p1 = 0.03723) w = 62.67485	53 K 1.000**, 54 S 1.000**

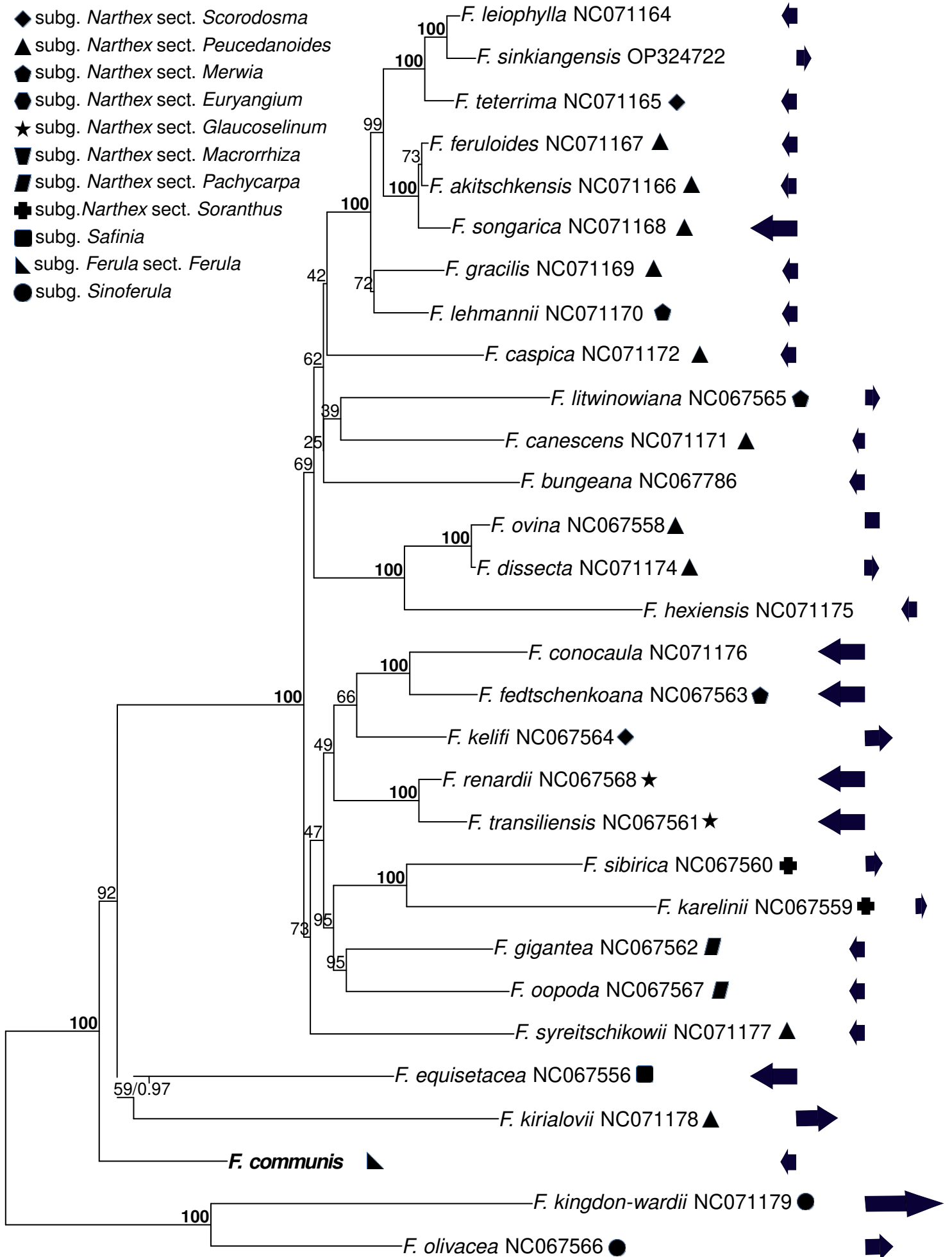
Supplementary Figure 1 The structural variation in the boundaries of the Large Single Copy (LSC), Small Single Copy (SSC), and Inverted Repeat (IR) regions across the plastid genomes of 30 *Ferula* and two outgroup species. It highlights differences in gene positioning, boundary shifts, and the movement of genes such as *ndhF* at the IR/SSC junctions, providing insights in the dynamics of plastid genome organization within the genus.

Supplementary Figure 2 The phylogenetic tree based on plastid genes illustrates the organization of IR boundaries, specifically the IRb/SSC region. The movement of the *ndhF* gene into or away from the IRb region does not display any clear phylogenetic structure among the compared *Ferula* species. A left-pointing arrow indicates the movement of the *ndhF* gene into the IRb region, while a right-pointing arrow represents its movement away from the IRb region. The size of the arrow corresponds to the number of base pairs (smaller arrows indicate fewer base pairs). A square symbol indicates that the *ndhF* gene is located precisely on the IRb/SSC boundary.

Inverted Repeats



- ◆ subg. *Narthex* sect. *Scorodosma*
- ▲ subg. *Narthex* sect. *Peucedanoides*
- ◆ subg. *Narthex* sect. *Merwia*
- subg. *Narthex* sect. *Euryangium*
- ★ subg. *Narthex* sect. *Glaucoselinum*
- ▼ subg. *Narthex* sect. *Macrorrhiza*
- ▀ subg. *Narthex* sect. *Pachycarpa*
- ⊕ subg. *Narthex* sect. *Soranthus*
- subg. *Safinia*
- ▲ subg. *Ferula* sect. *Ferula*
- subg. *Sinoferula*



5e-04