Supplementary Information

Localized hypermutation and associated gene losses in legume chloroplast genomes

Alan M. Magee[1], Sue Aspinall[2], Danny W. Rice[3], Brian P. Cusack[1], Marie Sémon[4], Antoinette S. Perry[1], Sasa Stefanovic[5], Dan Milbourne[6], Susanne Barth[6], Jeffrey D. Palmer[3], John C. Gray[2], Tony A. Kavanagh[1], and Kenneth H. Wolfe[1]

[1] Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland
[2] Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom
[3] Department of Biology, Indiana University, Bloomington, Indiana 47405
[4] Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, CNRS, INRA, UCB Lyon 1, Ecole Normale Supérieure de Lyon, 69364 Lyon Cedex 07, France
[5] Department of Biology, University of Toronto, Mississauga, Ontario L5L 1C6, Canada
[6] Teagasc Crops Research Centre, Oak Park, Carlow, Ireland

Supplementary Figure legends

Figure S1. Synonymous and nonsynonymous divergence in angiosperm chloroplast *matK* (A), *rbcL* (B), and *ycf4* (C) genes. The trees are the same as those in Figure 1, except for the addition of taxon names. Shown are *dN* (left) and *dS* (right) trees resulting from a codon-based likelihood analysis and a constrained topology, rooted using gymnosperm sequences (which are not included in the trees). See Methods for details of tree construction. The species are in the same order from top to bottom in all trees to the greatest extent possible. In (C), numbers after species names indicate the *ycf4* gene length in codons. The *ycf4* pseudogenes in *Desmodium* and *Clitoria* are included in the *ycf4* trees and labeled in the left panel. The *Lathyrus odoratus* and *Pisum sativum ycf4* pseudogenes were not included in the tree because they are truncated and very divergent, making their boundaries hard to define.

Figure S2. Dot matrix plots showing repetitive DNA around the *accD-ycf4* region in legumes. All plots are drawn to the same scale and show (A) self-comparison of *Pisum sativum*, (B) self-comparison of *Lathyrus sativus*, (C) comparison of *P. sativum* to *L. sativus*, and (D) self-comparison of *L. latifolius.* Panels A, B and D used a criterion of 28 matching bases per 30 bp window, and panel C used a criterion of 20 matches per 30 bp window. Black dots show matches on the same DNA strand and red dots show matches on the complementary strand.

Figure S3. (A-C) Gene maps of the chloroplast genomes of *Lathyrus sativus* (grasspea), *Pisum sativum* (pea), and *Nicotiana tabacum* (tobacco), drawn using GenomeVx (Conant and Wolfe 2008). Larger versions of the grasspea and pea maps are presented in Figure S7. Uppercase letters A through R show the 18 segments of conserved gene order among these three genomes, with plus or minus signs indicating orientation. The tobacco map shows an isomer whose small single copy region is inverted relative to the usual presentation of this genome (Shinozaki et al. 1986). The dashed line in region A of the tobacco map represents the copy of the IR that was deleted in legumes. (D) One scenario for rearranging the 18 conserved segments to turn an ancestral angiosperm genome with a tobacco-like gene order (but lacking an IR) into the current *L. sativus* and *P. sativum* gene orders, using only inversions. This scenario was calculated using GRIMM (Tesler 2002). The specific 6- and 8-step paths leading to *L. sativus* and *P. sativum*, respectively, are computationally optimal solutions (most parsimonious), but of course are not necessarily what actually happened. The 8-step path to *P. sativum* proposed by GRIMM involves the same set of inversions, but in a different order, to those proposed by Palmer et al. (1988) whose segment nomenclature is shown at the bottom. (E) Phylogenetic context of inversions and gene losses. The relative order of events within individual branches is not known. We calculated *T. subterraneum* (Cai et al. 2008) to have a total of 15

inversions and 4 protein gene losses, using GRIMM with a modified annotation that included *ycf1*, *ycf4*, *rps18* and *trnG*-UCU and ignored a duplicate copy of *trnI*-CAU.

Figure S4. Dot matrix plots of self-similarity in six chloroplast genome sequences, using a criterion of 28 matches per 30 bp window. To allow tandem and near-tandem repeats to be seen, we do not show the major diagonal which would normally appear as a solid black line running from the bottom-left to the top-right corner of each plot. The IR appears as a pair of long red (inverted orientation) diagonals in the four IR-containing genomes. Arrows show the positions (where present) of *ycf4*, *accD* and *psaI* in the *P. sativum* and *L. sativus* plots.

Figure S5. Nucleotide sequence alignment between *L. latifolius* and *L. cirrhosus* in the *accD-ycf4-cemA* region. Caret symbols show nucleotide substitutions. In *L. latifolius*, blue triangles show the 57-bp repeat (lowercase letters indicate differences relative to the consensus), and orange triangles show the 67-bp repeat (the two copies are identical). Complete copies of the repeats are labeled A1-A6 and B1-B2, and incomplete copies are show by triangles without labels. The two largest repeat sequences in the *L. cirrhosus accD-ycf4* intergenic region are shown by magenta and purple underlining (2 x 15 bp each). Sequences were aligned using Muscle and were constrained (after inspection of dot-matrix plots) to align only the B1 and A1 copies of the *L. latifolius* repeats to *L. cirrhosus*.

Figure S6. The *LPD2-accD* fusion in *Trifolium repens*.
(A) Comparative organization of *LPD* genes and cDNAs in *Medicago truncatula* and *T. repens*, not drawn to scale. cDNA sequences for the four genes were inferred from EST data. Intron/exon organization is known only for the *M. truncatula* genes, and is conserved between *LPD1* and *LPD2* except in the transit peptide region. The chloroplast-derived *accD* sequence appears to have replaced the last two exons of *T. repens LPD2*.
(B) Evidence that the *T. repens* fusion gene is orthologous to *M. truncatula LPD2.* Synonymous (*dS*) and nonsynonymous (*dN*) nucleotide divergences among the four genes were calculated using yn00 for the region corresponding to exons 2-13 of *M. truncatula LPD2.* Divergence between orthologs (shaded boxes) is lower than between paralogs.
(C) Assembly of the cDNA sequence of the *T. repens LPD2-accD* fusion gene from EST data. Five EST reads cross the junction between *LPD2* and *accD*.
(D) Alignment of the predicted *T. repens* LPD2-AccD fusion protein (Tr_LPD2-accD) to other LPD and AccD proteins. The complete sequence of the fusion protein (805 residues) is shown. Residues 1-512 are aligned to plastid lipoamide dehydrogenases from *M. truncatula* (Mt_LPD1 and Mt_LPD2) and *T. repens* (Tr_LPD1). Residues 513-805 are aligned to the *Lotus japonicus* chloroplast *accD* gene translation (Lj_accD). Protein domains in LPD are indicated (Lutziger and Oliver 2000), including a

conserved motif (HAHPT) in the interface domain that is highly conserved among plastid, mitochondrial and bacterial LPD proteins but has been lost from the *T. repens* fusion protein. The site corresponding to the location of the intron between exons 13 and 14 of *M. truncatula LPD2* is shown by a triangle. In the AccD alignment, similarity to the *T. repens* fusion protein begins at residue 209 of *L. japonicus* AccD. The location of the carboxyl transferase domain (Zhang et al. 2003) in AccD is marked. There are no known functional domains in AccD upstream of this domain. An arrow indicates the site corresponding to position 267 of pea AccD, where mRNA editing converts a Ser codon to a Leu codon (Sasaki et al. 2001). This site is edited in one of the two *Lotus* ESTs available (accession numbers GO021515 and DC597685).

(E) Confirmation of the junction between *LPD2* and *accD* in *T. repens* mRNA by reverse transcriptase PCR and Sanger sequencing. The arrow marks the junction. The primers used for amplification were 5'- AAAATGAAGGGGAGGGACAT and 5'- GCACTGAACCCACAAATGG, amplifying a 200 bp fragment centered on the junction (positions 1465 to 1664 of GenBank accession HM029367).

(F) Phylogenetic relationship of *Trifolium* nuclear AccD sequences to chloroplast AccD sequences from other angiosperms. The tree was constructed from amino acid sequences of the C-terminal part of the protein (beginning at residue 513 of the *T. repens* LPD2-AccD protein). Sequences were aligned using Muscle and the tree was constructed using PhyML (WAG+Γ substitution model, 8 rate classes) as implemented in SeaView (Gouy et al. 2010).

Figure S7. Gene maps of (A) the *Lathyrus sativus* (grasspea) chloroplast genome, and (B) the *Pisum sativum* (pea) chloroplast genome.

Supplementary Tables.

Table S1 is a separate Excel file.

Table S2. Synonymous and nonsynonymous divergence between pea and sweetpea nuclear genes.

Pea GenBank acc. number	Sweetpea EST contig	Description	*S*	*N*	Omega (*dN/dS*)	*dN*	S.E.	*dS*	S.E.
D89619	CL884	Cycloartenol synthase	113.6	282.4	0.0000	0.0000	0.0000	0.0178	0.0126
AY166633	CL64	Malate dehydrogenase (MDH)	266.6	678.4	0.0871	0.0059	0.0030	0.0680	0.0171
AF095284	CL564	Tic22	180.9	416.1	0.5556	0.0421	0.0103	0.0758	0.0214
AF043905	CL118	Plastoglobule associated protein PG1	259.7	613.3	0.2124	0.0165	0.0052	0.0776	0.0182
Z31559	CL1383	AccA acetyl-CoA carboxylase	129.9	422.1	0.8786	0.0682	0.0132	0.0777	0.0258
AF095285	CL1536	Tic20	131.9	390.1	0.7360	0.0573	0.0125	0.0778	0.0263
X89828	CL26	Fructose-1, 6-biphosphate aldolase	217.9	601.1	0.1614	0.0134	0.0048	0.0832	0.0207
M94558	CL523	ATP synthase delta subunit	200.1	522.9	1.1445	0.0966	0.0143	0.0844	0.0221
AJ630104	CL810	Galactokinase (galK gene)	244.5	661.5	0.3449	0.0299	0.0068	0.0866	0.0202
AJ250769	CL822	Cytosolic phosphoglucomutase (PGM gene)	133.1	445.9	0.1284	0.0113	0.0051	0.0880	0.0271
AY367058	CL240	SAT5 mRNA	155.8	429.2	0.1579	0.0141	0.0058	0.0894	0.0255
AB104529	CL589	LKA mRNA for brassinosteroid receptor	293.6	726.4	0.2149	0.0196	0.0053	0.0911	0.0187
U56697	CL402	Pyruvate dehydrogenase E1beta	210.4	689.6	0.1894	0.0176	0.0051	0.0930	0.0229
L29077	CL80	Ubiquitin conjugating enzyme (UBC4)	120.4	323.6	0.0314	0.0031	0.0031	0.0987	0.0308
AF275639	CL1669	Cytosolic phosphoglycerate kinase (PGK)	175.9	517.1	0.1168	0.0117	0.0048	0.1002	0.0264
M73744	CL337	IM30 protein mRNA	156.7	461.3	0.1268	0.0131	0.0054	0.1035	0.0276
X63604	CL831	AtpC (gamma subunit of ATP synthase)	177.6	395.4	0.2748	0.0284	0.0086	0.1035	0.0259
X60170	CL48	Mn superoxide dismutase (SOD)	137.0	496.0	0.1238	0.0132	0.0052	0.1068	0.0300
AJ251646	CL1446	beta-1,3-glucanase GNS2	171.8	416.2	0.3448	0.0370	0.0096	0.1072	0.0268
AJ001009	CL643	OEP24 preprotein	141.6	446.4	0.1045	0.0113	0.0051	0.1080	0.0301
AJ005589	CL1292	Protein tyrosine phosphatase	146.7	405.3	0.5498	0.0617	0.0127	0.1122	0.0302
X63605	CL46	PetC mRNA for chloroplast Rieske FeS protein	126.6	380.4	0.2615	0.0295	0.0090	0.1129	0.0327
X53035	CL800	P34 protein kinase	73.9	346.1	0.0489	0.0058	0.0041	0.1186	0.0435
DQ535894	CL1210	Tic21	233.1	501.9	0.2340	0.0285	0.0077	0.1216	0.0246
Y17186	CL664	Translation initiation factor eIF-4A	161.3	462.7	0.0713	0.0087	0.0044	0.1219	0.0297
AF002698	CL1095	NADPH-cytochrome P450 reductase (PSC450R1)	186.6	575.4	0.0677	0.0087	0.0039	0.1293	0.0285
AJ000520	CL1365	Tic55	205.7	646.3	0.1324	0.0172	0.0052	0.1301	0.0280
DQ026703	CL425	WD-40 repeat protein (MSI1)	109.9	673.1	0.0114	0.0015	0.0015	0.1303	0.0376

M71235	CL1699	Aminolevulinic acid dehydratase (ALAD)	142.4	457.6	0.1183	0.0155	0.0059	0.1307	0.0327
U60592	CL178	S-adenosylmethionine decarboxylase	263.3	795.7	0.1892	0.0256	0.0057	0.1352	0.0250
AF148506	CL1175	Hs1pro-1 homolog	64.5	217.5	0.0334	0.0046	0.0046	0.1383	0.0514
DQ845340	CL323	CRY mRNA	116.6	453.4	0.1087	0.0156	0.0059	0.1435	0.0388
AF109922	CL201	Sucrose transport protein SUT1	196.7	601.3	0.1526	0.0222	0.0062	0.1457	0.0297
X54359	CL832	Clone 26g	160.9	529.1	0.1714	0.0250	0.0070	0.1458	0.0333
X82404	CL558	SecA	204.6	662.4	0.0512	0.0076	0.0034	0.1483	0.0301
AY299688	CL353	RPA 32kDa mRNA	165.9	419.1	0.2946	0.0443	0.0105	0.1504	0.0341
X60169	CL343	Catalase	230.2	675.8	0.1382	0.0210	0.0056	0.1520	0.0289
AY822467	CL869	AT-rich element binding factor 2 (ATF2)	94.2	196.8	0.0326	0.0051	0.0051	0.1566	0.0461
AF191098	CL197	Nucleoside diphosphate kinase (NDPK)	180.7	518.3	0.1242	0.0195	0.0062	0.1573	0.0335
AF284759	CL221	Chloroplast TatC protein	126.1	497.9	0.0879	0.0142	0.0054	0.1615	0.0402
X60373	CL1027	Glutathione reductase	161.4	399.6	0.1403	0.0229	0.0077	0.1633	0.0364
L34578	CL264	Histone H1 (H1-41)	189.6	356.4	0.2079	0.0346	0.0101	0.1665	0.0330
X70703	CL485	MAP kinase homologue	126.6	443.4	0.0395	0.0068	0.0039	0.1720	0.0421
M60952	CL383	Chloroplast ribosomal protein (CL22)	150.2	428.8	0.5346	0.0922	0.0154	0.1724	0.0379
AF079850	CL362	Nodule-enhanced malate dehydrogenase (NEMDH)	151.9	418.1	0.1807	0.0318	0.0089	0.1758	0.0397
X66061	CL833	Thiolprotease	190.0	623.0	0.1985	0.0354	0.0077	0.1781	0.0349
X65155	CL179	Ribosomal protein L9	114.7	464.3	0.0407	0.0076	0.0041	0.1863	0.0463
AF323104	CL1455	Phosphatase-like protein Psc923	152.8	456.2	0.1552	0.0291	0.0081	0.1872	0.0395
X54358	CL767	Clone 15a	155.1	477.9	0.0547	0.0105	0.0047	0.1927	0.0400
AF002248	CL732	PSI LHC Chl a/b-binding protein (lhcA-P4)	200.7	555.3	0.0992	0.0201	0.0061	0.2025	0.0365
X62077	CL501	ApxI mRNA for ascorbate peroxidase	167.4	582.6	0.0933	0.0191	0.0058	0.2051	0.0416
AF144684	CL682	SecY	242.4	813.6	0.1606	0.0339	0.0066	0.2114	0.0339
Y15253	CL1298	Phospholipase C	100.6	424.4	0.0853	0.0191	0.0068	0.2238	0.0556
AY830931	CL1039	FVE mRNA	113.8	471.2	0.0374	0.0085	0.0043	0.2281	0.0522
AJ315851	CL232	2-Cys peroxiredoxin	174.6	530.4	0.0894	0.0239	0.0068	0.2677	0.0471
AJ243759	CL746	Outer envelope protein OEP37	139.7	391.3	0.2253	0.0640	0.0132	0.2841	0.0559
		Median values and total numbers of sites	9340	27788	0.1353	0.0191		0.1305	

Genes are sorted in order of increasing *dS.* An arbitrary cutoff of $dS = 0.3$ was applied to remove probable paralogs. *S* and *N* are the numbers of synonymous and nonsynonymous sites in each sequence.

References for Supplementary Information

Cai Z, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) Is associated with numerous repeated sequences and novel DNA insertions. *J Mol Evol* **67:** 696-704.

Conant GC, Wolfe KH. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* **24:** 861-862.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27:** 221-224.

Lutziger I, Oliver DJ. 2000. Molecular evidence of a unique lipoamide dehydrogenase in plastids: analysis of plastidic lipoamide dehydrogenase from *Arabidopsis thaliana*. *FEBS Lett* **484:** 12-16.

Palmer JD, Osorio B, Thompson WF. 1988. Evolutionary significance of inversions in legume chloroplast DNAs. *Curr Genet* **14:** 65-74.

Sasaki Y, Kozaki A, Ohmori A, Iguchi H, Nagano Y. 2001. Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *J Biol Chem* **276:** 3937-3940.

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, et al. 1986. The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. *EMBO J* **5:** 2043-2049.

Tesler G. 2002. GRIMM: genome rearrangements web server. *Bioinformatics* **18:** 492-493.

Zhang H, Yang Z, Shen Y, Tong L. 2003. Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase. *Science* **299:** 2064-2067.

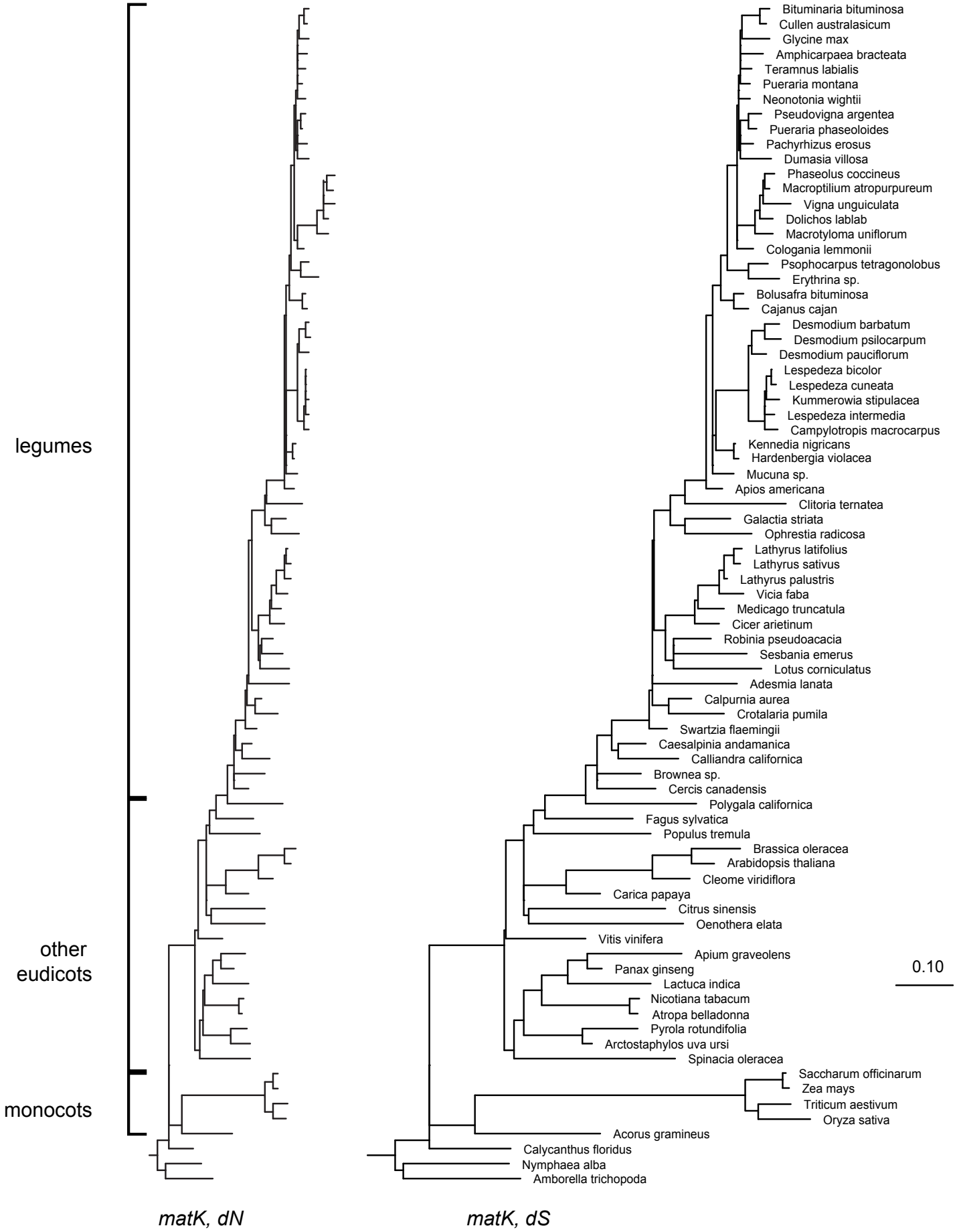

legumes
other eudicots
monocots
Bituminaria bituminosa
Cullen australasicum
Glycine max
Amphicarpaea bracteata
Teramnus labialis
Pueraria montana
Neonotonia wightii
Pseudovigna argentea
Pueraria phaseoloides
Pachyrhizus erosus
Dumasia villosa
Phaseolus coccineus
Macroptilium atropurpureum
Vigna unguiculata
Dolichos lablab
Macrotyloma uniflorum
Cologania lemmonii
Psophocarpus tetragonolobus
Erythrina sp.
Bolusafra bituminosa
Cajanus cajan
Desmodium barbatum
Desmodium psilocarpum
Desmodium pauciflorum
Lespedeza bicolor
Lespedeza cuneata
Kummerowia stipulacea
Lespedeza intermedia
Campylotropis macrocarpus
Kennedia nigricans
Hardenbergia violacea
Mucuna sp.
Apios americana
Clitoria ternatea
Galactia striata
Ophrestia radicosa
Lathyrus latifolius
Lathyrus sativus
Lathyrus palustris
Vicia faba
Medicago truncatula
Cicer arietinum
Robinia pseudoacacia
Sesbania emerus
Lotus corniculatus
Adesmia lanata
Calpurnia aurea
Crotalaria pumila
Swartzia flaemingii
Caesalpinia andamanica
Calliandra californica
Brownea sp.
Cercis canadensis
Polygala californica
Fagus sylvatica
Populus tremula
Brassica oleracea
Arabidopsis thaliana
Cleome viridiflora
Carica papaya
Citrus sinensis
Oenothera elata
Vitis vinifera
Apium graveolens
Panax ginseng
Lactuca indica
Nicotiana tabacum
Atropa belladonna
Pyrola rotundifolia
Arctostaphylos uva ursi
Spinacia oleracea
Saccharum officinarum
Zea mays
Triticum aestivum
Oryza sativa
Acorus gramineus
Calycanthus floridus
Nymphaea alba
Amborella trichopoda
0.10
matK, dN
matK, dS

Figure S1A

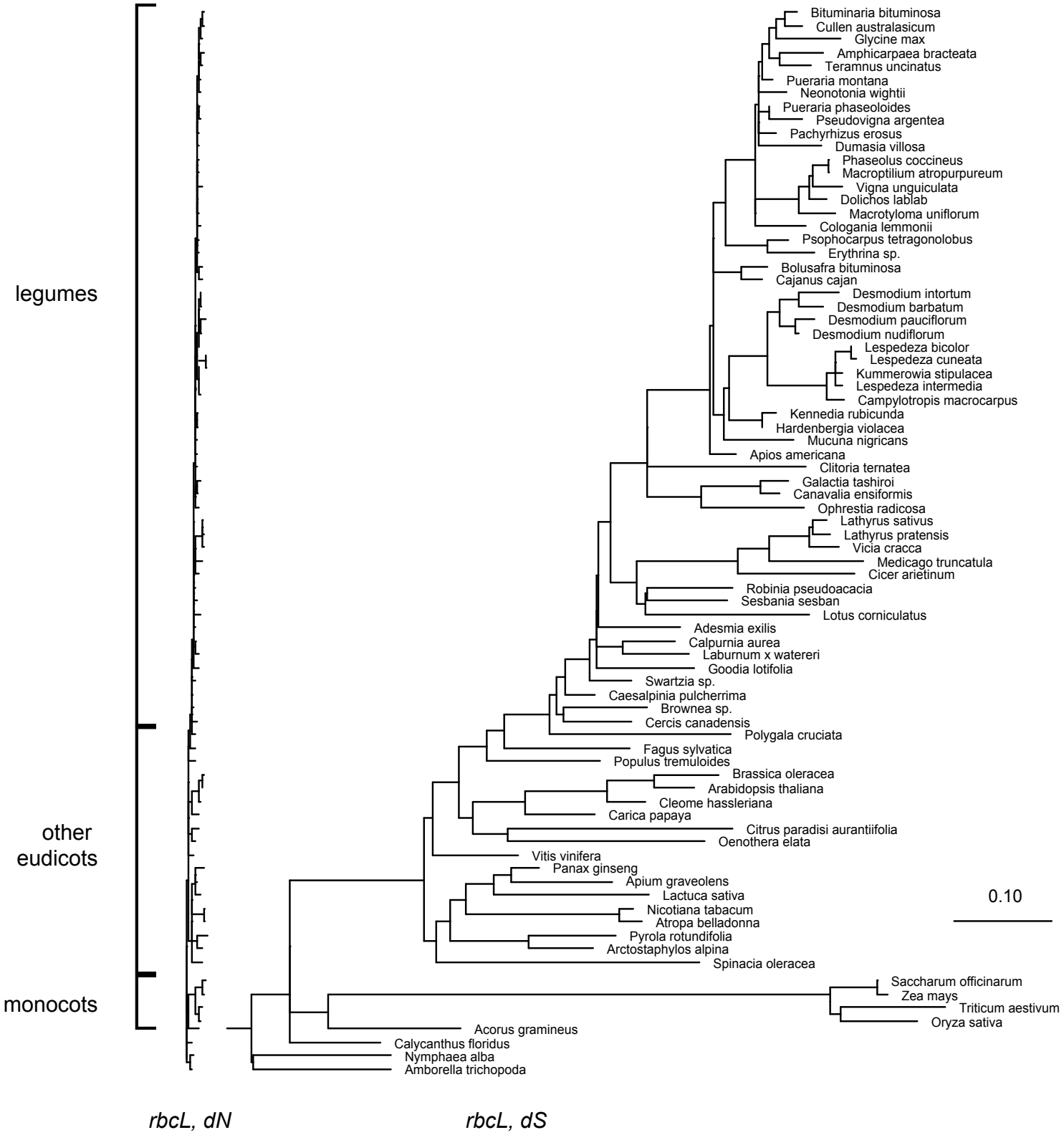

legumes
other eudicots
monocots
Bituminaria bituminosa
Cullen australasicum
Glycine max
Amphicarpaea bracteata
Teramnus uncinatus
Pueraria montana
Neonotonia wightii
Pueraria phaseoloides
Pseudovigna argentea
Pachyrhizus erosus
Dumasia villosa
Phaseolus coccineus
Macroptilium atropurpureum
Vigna unguiculata
Dolichos lablab
Macrotyloma uniflorum
Cologania lemmonii
Psophocarpus tetragonolobus
Erythrina sp.
Bolusafra bituminosa
Cajanus cajan
Desmodium intortum
Desmodium barbatum
Desmodium pauciflorum
Desmodium nudiflorum
Lespedeza bicolor
Lespedeza cuneata
Kummerowia stipulacea
Lespedeza intermedia
Campylotropis macrocarpus
Kennedia rubicunda
Hardenbergia violacea
Mucuna nigricans
Apios americana
Clitoria ternatea
Galactia tashiroi
Canavalia ensiformis
Ophrestia radicosa
Lathyrus sativus
Lathyrus pratensis
Vicia cracca
Medicago truncatula
Cicer arietinum
Robinia pseudoacacia
Sesbania sesban
Lotus corniculatus
Adesmia exilis
Calpurnia aurea
Laburnum x watereri
Goodia lotifolia
Swartzia sp.
Caesalpinia pulcherrima
Brownea sp.
Cercis canadensis
Polygala cruciata
Fagus sylvatica
Populus tremuloides
Brassica oleracea
Arabidopsis thaliana
Cleome hassleriana
Carica papaya
Citrus paradisi aurantiifolia
Oenothera elata
Vitis vinifera
Panax ginseng
Apium graveolens
Lactuca sativa
Nicotiana tabacum
Atropa belladonna
Pyrola rotundifolia
Arctostaphylos alpina
Spinacia oleracea
Saccharum officinarum
Zea mays
Triticum aestivum
Oryza sativa
Acorus gramineus
Calycanthus floridus
Nymphaea alba
Amborella trichopoda
0.10
rbcL, dN
rbcL, dS

Figure S1B

ycf4, dN

ycf4, dS

Figure S1C

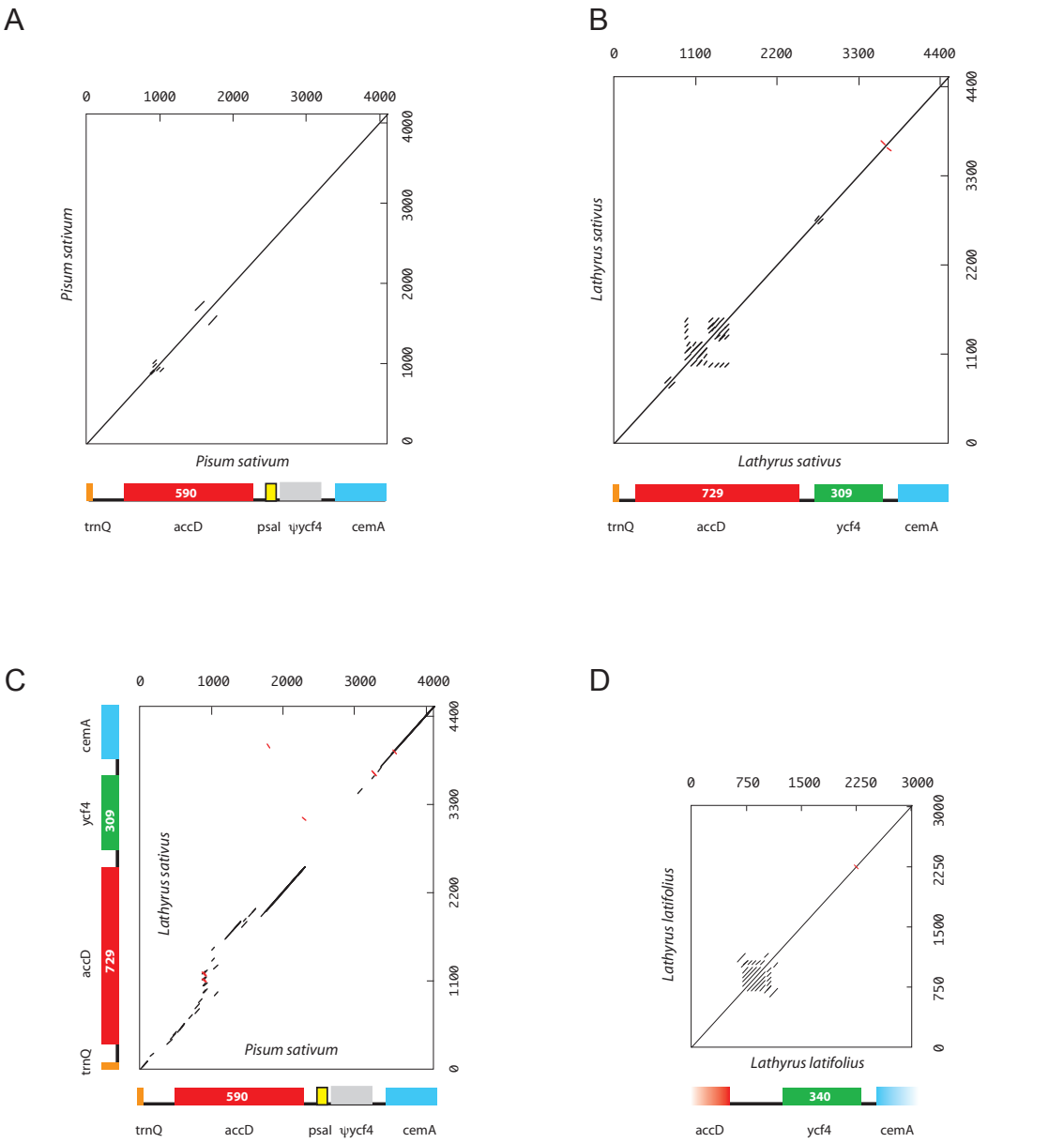
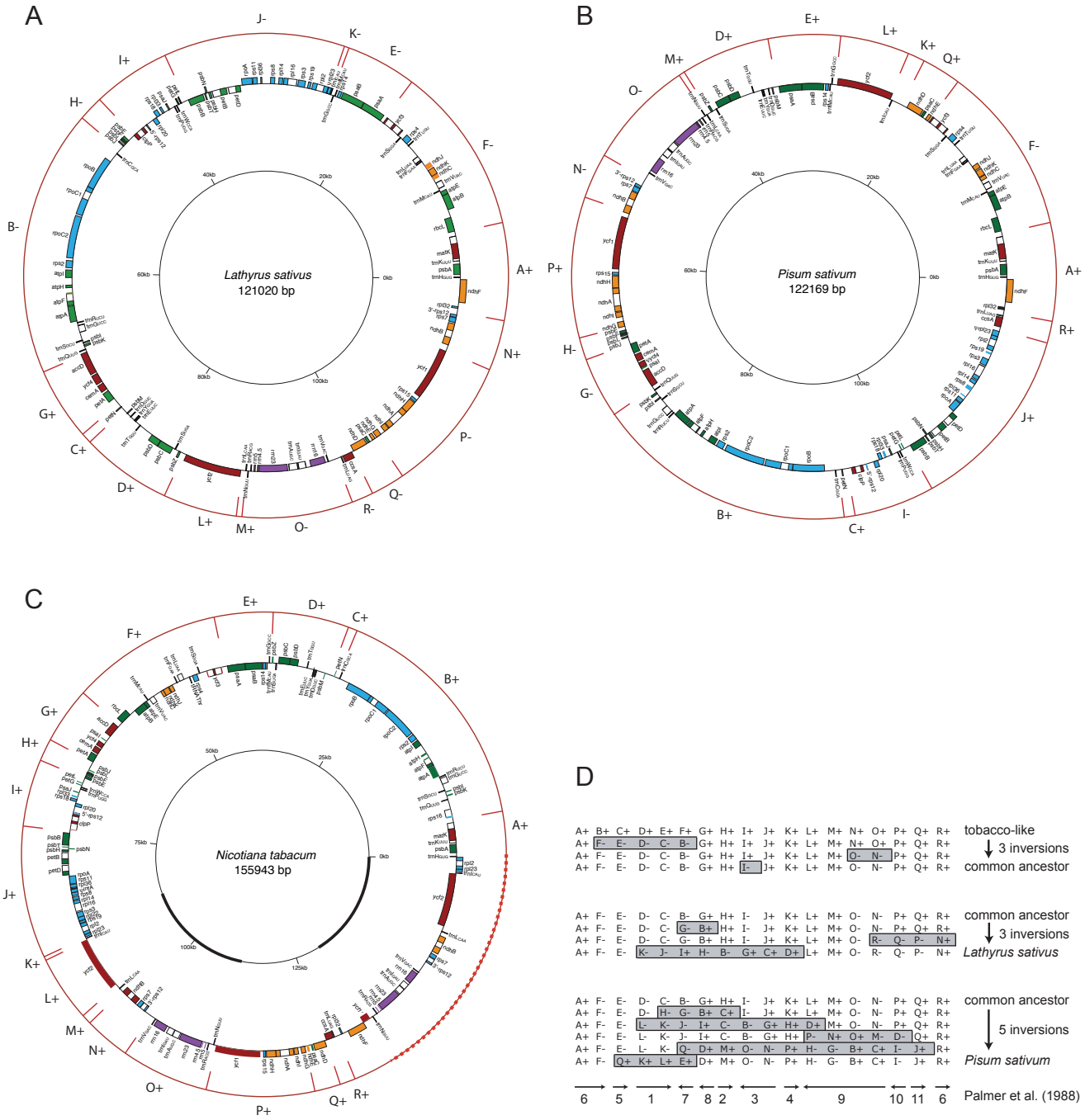

A
0
1000
2000
3000
4000
Pisum sativum
Pisum sativum
590
trnQ
accD
psal
ψycf4
cemA
B
0
1100
2200
3300
4400
Lathyrus sativus
Lathyrus sativus
729
309
trnQ
accD
ycf4
cemA
C
cemA
ycf4
309
accD
729
trnQ
Lathyrus sativus
Pisum sativum
590
trnQ
accD
psal
ψycf4
cemA
D
0
750
1500
2250
3000
Lathyrus latifolius
Lathyrus latifolius
340
accD
ycf4
cemA

Figure S2



D

A+ B+ C+ D+ E+ F+ G+ H+ I+ J+ K+ L+ M+ N+ O+ P+ Q+ R+		tobacco-like
A+ F- E- D- C- B- G+ H+ I+ J+ K+ L+ M+ N+ O+ P+ Q+ R+	↓ 3 inversions	
A+ F- E- D- C- B- G+ H+ I+ J+ K+ L+ M+ O- N- P+ Q+ R+		common ancestor
A+ F- E- D- C- B- G+ H+ I- J+ K+ L+ M+ O- N- P+ Q+ R+		common ancestor
A+ F- E- D- C- G- B+ H+ I- J+ K+ L+ M+ O- N- P+ Q+ R+	↓ 3 inversions	
A+ F- E- D- C- G- B+ H+ I- J+ K+ L+ M+ O- R- Q- P- N+		
A+ F- E- K- J- I+ H- B- G+ C+ D+ L+ M+ O- R- Q- P- N+		*Lathyrus sativus*
A+ F- E- D- C- B- G+ H+ I- J+ K+ L+ M+ O- N- P+ Q+ R+		common ancestor
A+ F- E- D- H- G- B+ C+ I- J+ K+ L+ M+ O- N- P+ Q+ R+	↓ 5 inversions	
A+ F- E- L- K- J- I+ C- B- G+ H+ D+ M+ O- N- P+ Q+ R+		
A+ F- E- L- K- Q- D+ M+ O- N- P+ H- G- B+ C+ I- J+ R+		
A+ F- Q+ K+ L+ E+ D+ M+ O- N- P+ H- G- B+ C+ I- J+ R+		*Pisum sativum*
6 5 1 7 8 2 3 4 9 10 11 6		Palmer et al. (1988)

E

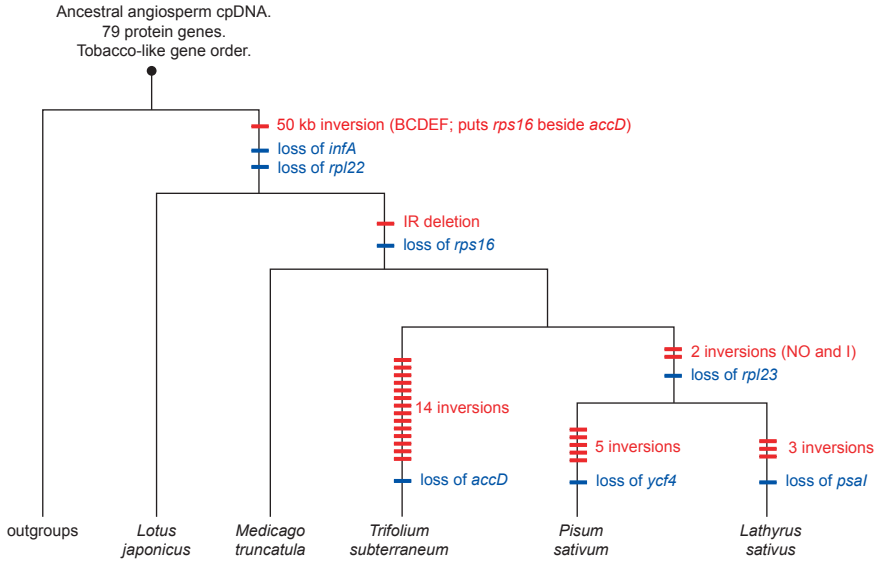


Figure S3

0
30250
60500
90750
121000
Lathyrus sativus
ycf4
accD
Lathyrus sativus
0
30500
61000
91500
122000
Pisum sativum
accD
psaI
Pisum sativum
0
37500
75000
112500
150000
Lotus japonicus
Lotus japonicus
0
38750
77500
116250
155000
Nicotiana tabacum
Nicotiana tabacum
0
37500
75000
112500
150000
Spinacia oleracea
Spinacia oleracea
0
33500
67000
100500
134000
Oryza sativa
Oryza sativa

Figure S4

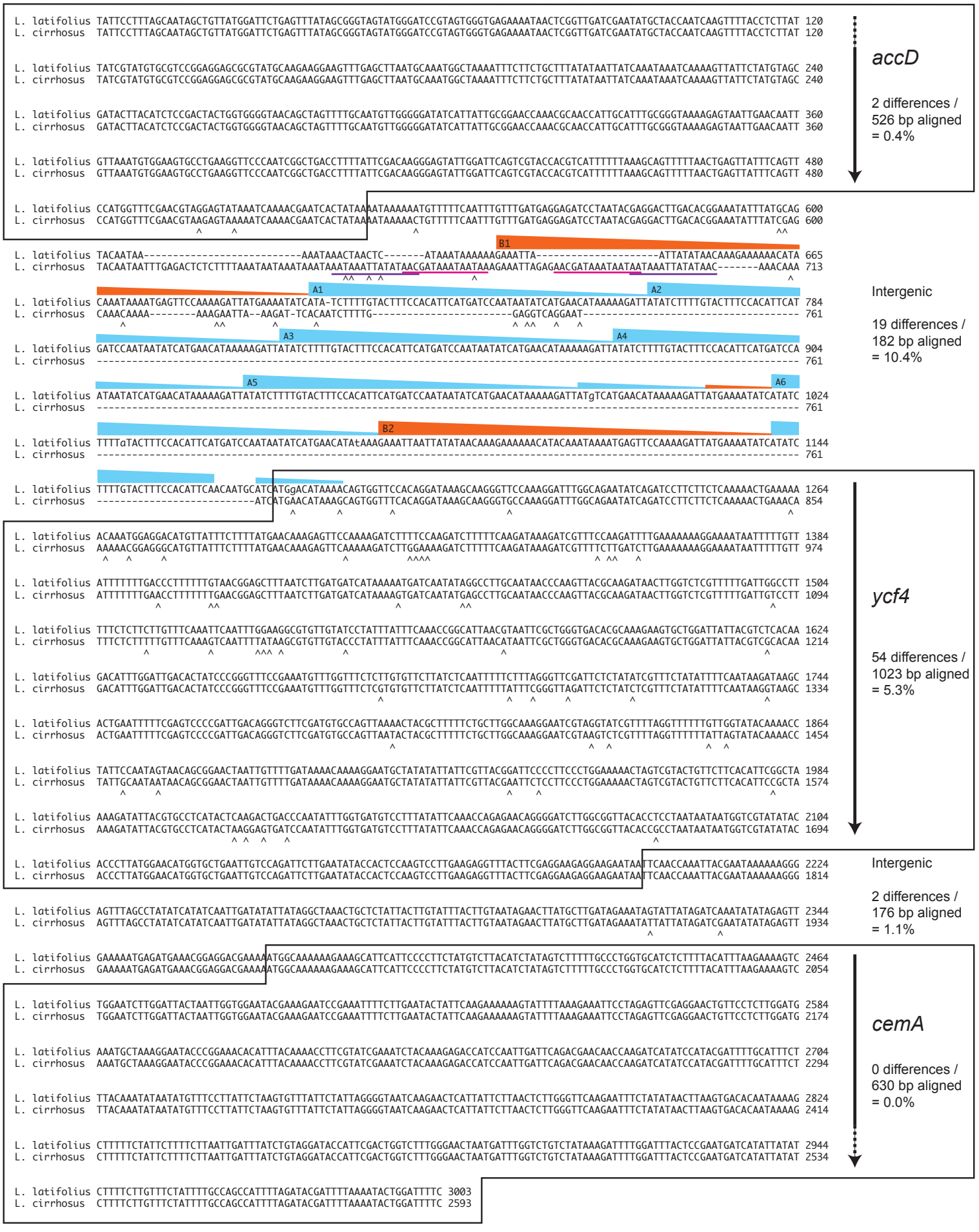
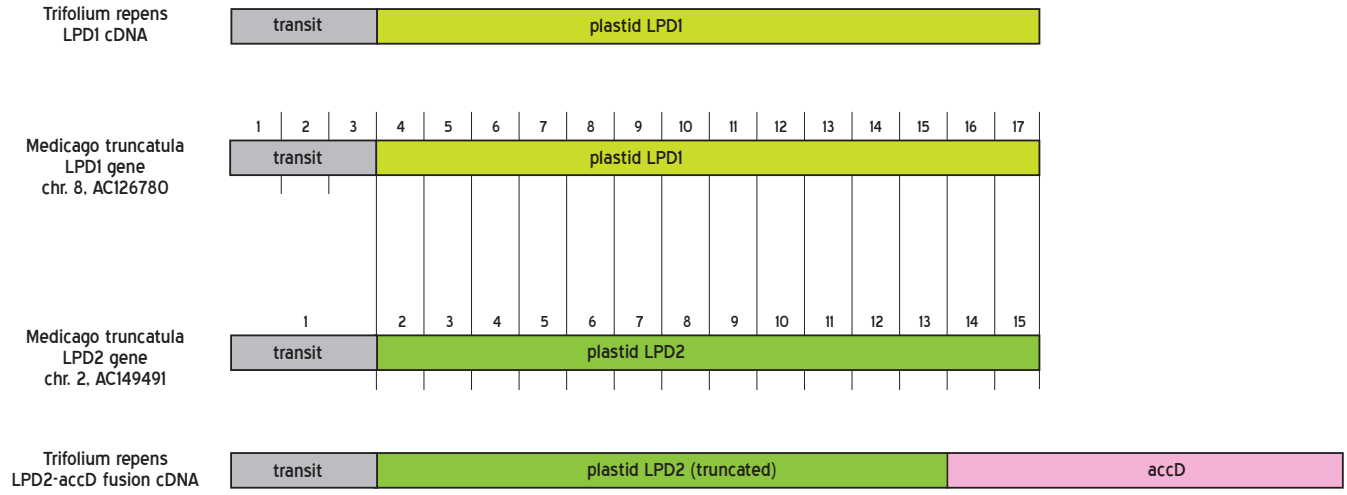

accD
2 differences / 526 bp aligned = 0.4%
Intergenic
19 differences / 182 bp aligned = 10.4%
ycf4
54 differences / 1023 bp aligned = 5.3%
Intergenic
2 differences / 176 bp aligned = 1.1%
cemA
0 differences / 630 bp aligned = 0.0%

Figure S5

A



B

dN \ dS	M. truncatula LPD1	T. repens LPD1	M. truncatula LPD2	T. repens LPD2-accD
M. truncatula LPD1		0.1939	0.6453	0.5115
T. repens LPD1	0.0087		0.5454	0.5497
M. truncatula LPD2	0.0275	0.0341		0.3035
T. repens LPD2-accD	0.0372	0.0439	0.0180	

C

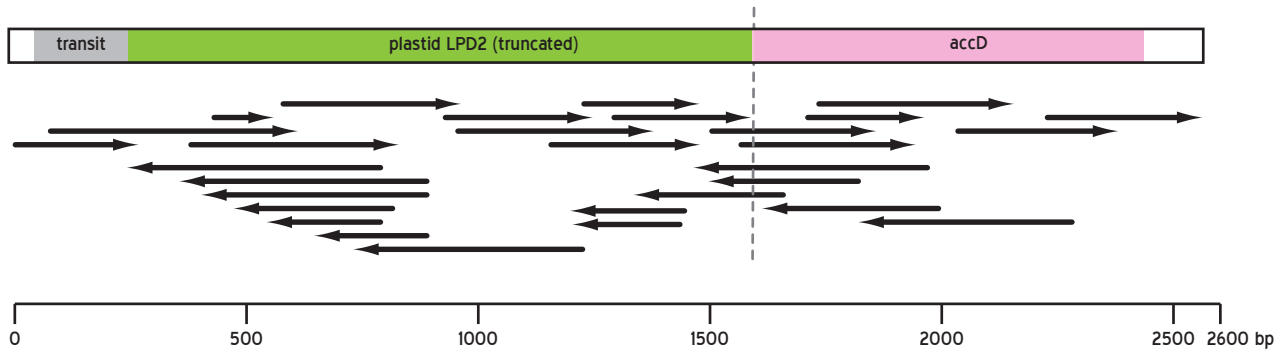
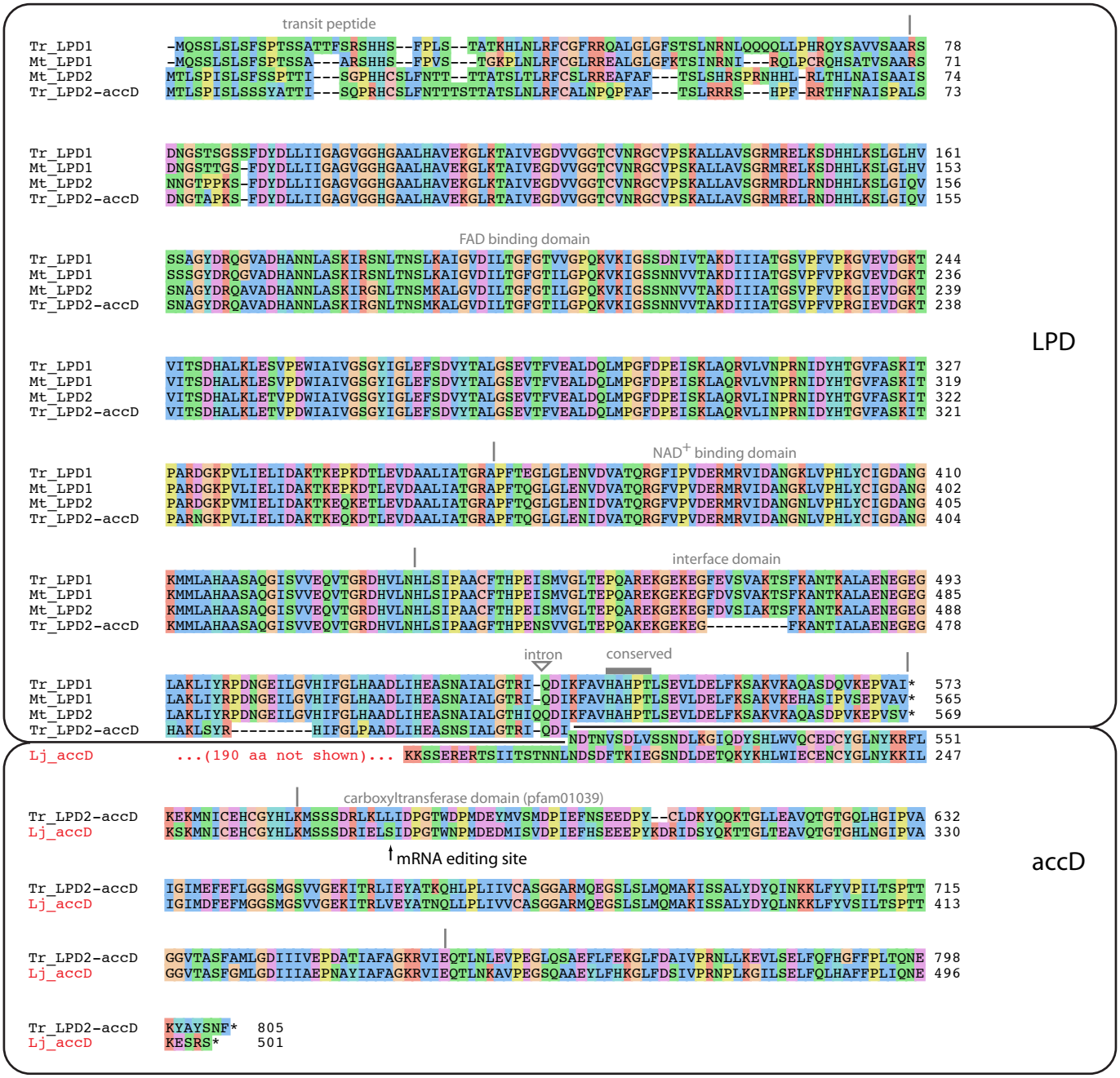


Figure S6 A,B,C

D



E

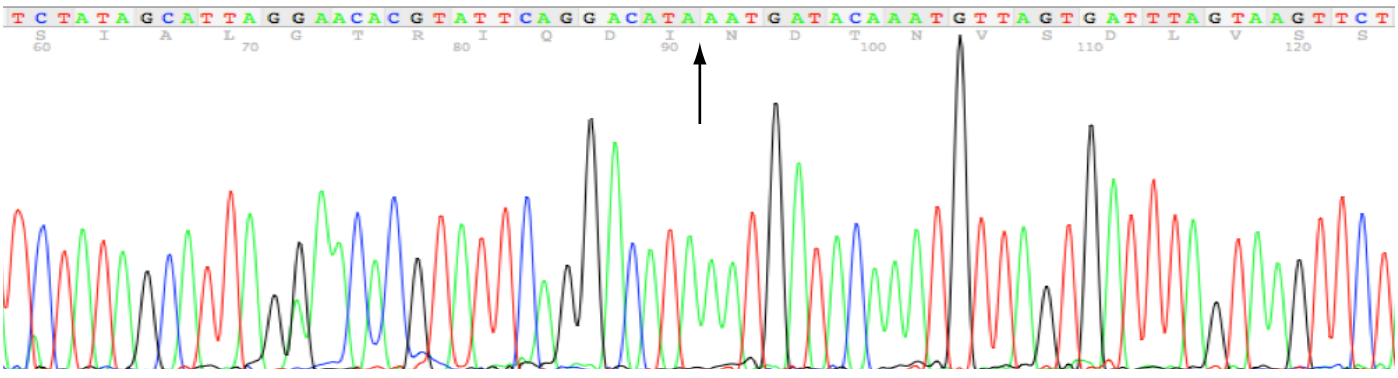


Figure S6 D, E

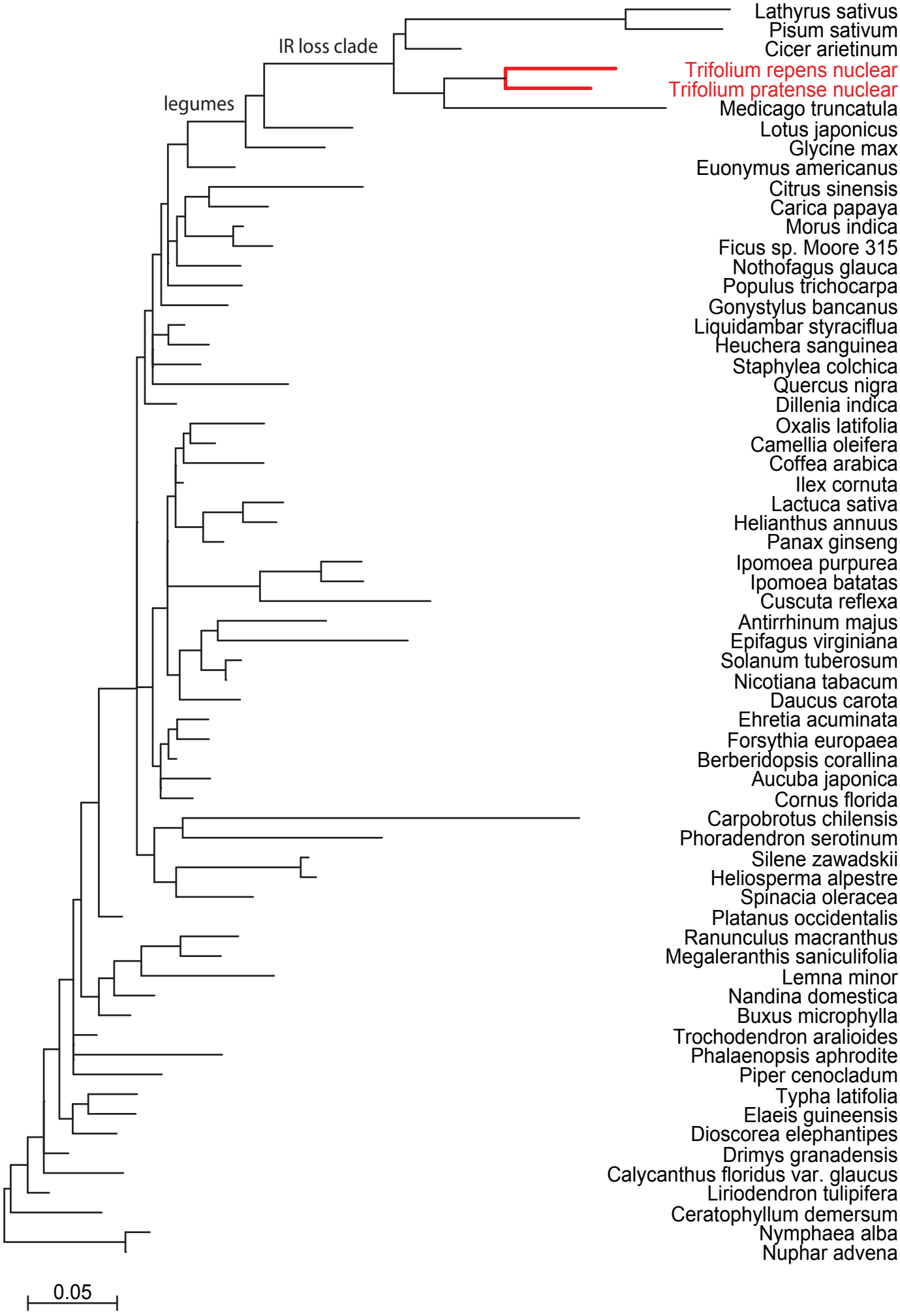

IR loss clade
legumes
Lathyrus sativus
Pisum sativum
Cicer arietinum
Trifolium repens nuclear
Trifolium pratense nuclear
Medicago truncatula
Lotus japonicus
Glycine max
Euonymus americanus
Citrus sinensis
Carica papaya
Morus indica
Ficus sp. Moore 315
Nothofagus glauca
Populus trichocarpa
Gonystylus bancanus
Liquidambar styraciflua
Heuchera sanguinea
Staphylea colchica
Quercus nigra
Dillenia indica
Oxalis latifolia
Camellia oleifera
Coffea arabica
Ilex cornuta
Lactuca sativa
Helianthus annuus
Panax ginseng
Ipomoea purpurea
Ipomoea batatas
Cuscuta reflexa
Antirrhinum majus
Epifagus virginiana
Solanum tuberosum
Nicotiana tabacum
Daucus carota
Ehretia acuminata
Forsythia europaea
Berberidopsis corallina
Aucuba japonica
Cornus florida
Carpobrotus chilensis
Phoradendron serotinum
Silene zawadskii
Heliosperma alpestre
Spinacia oleracea
Platanus occidentalis
Ranunculus macranthus
Megaleranthis saniculifolia
Lemna minor
Nandina domestica
Buxus microphylla
Trochodendron aralioides
Phalaenopsis aphrodite
Piper cenocladum
Typha latifolia
Elaeis guineensis
Dioscorea elephantipes
Drimys granadensis
Calycanthus floridus var. glaucus
Liriodendron tulipifera
Ceratophyllum demersum
Nymphaea alba
Nuphar advena
0.05

Figure S6 F

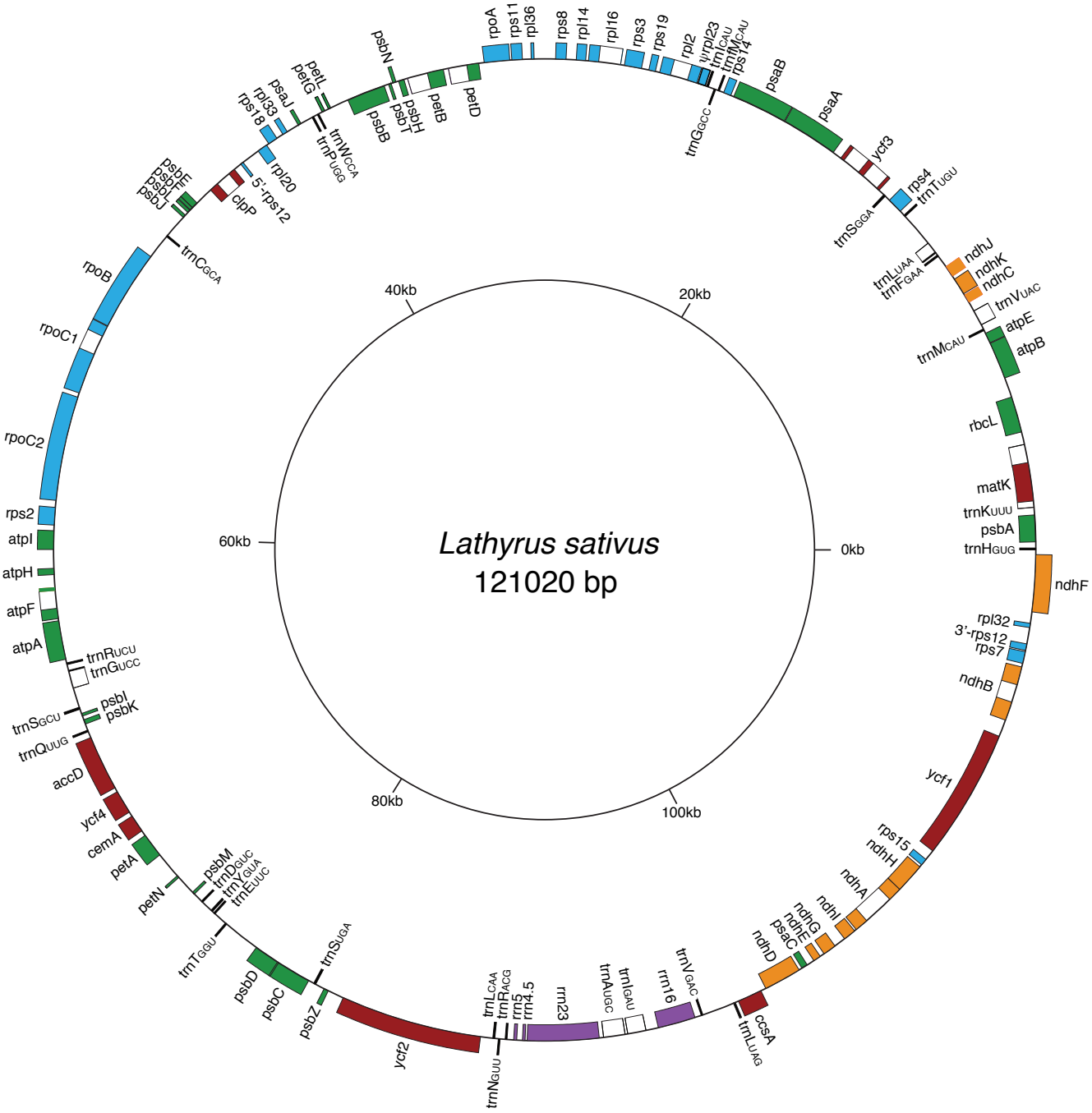

Lathyrus sativus
121020 bp
0kb
20kb
40kb
60kb
80kb
100kb
rpoA
rps11
rpl36
rps8
rpl14
rpl16
rps3
rps19
rpl2
rpl23
trnI CAU
trnM CAU
rps14
psaB
psaA
ycf3
trnG GCC
rps4
trnT UGU
trnS GGA
trnL UAA
trnF GAA
ndhJ
ndhK
ndhC
trnV UAC
trnM CAU
atpE
atpB
rbcL
matK
trnK UUU
psbA
trnH GUG
ndhF
rpl32
3'-rps12
rps7
ndhB
ycf1
rps15
ndhH
ndhA
ndhI
ndhG
ndhE
psaC
ndhD
ccsA
trnL UAG
trnV GAC
rrn16
trnI GAU
trnA UGC
rrn23
rrn4.5
rrn5
trnR ACG
trnL CAA
ycf2
trnN GUU
psbA
trnS UGA
psbC
psbD
trnT GGU
petN
psbM
trnD GUC
trnY GUA
trnE UUC
petA
cemA
ycf4
accD
trnQ UUG
trnS GCU
psbK
psbI
trnG UCC
trnR UCU
atpA
atpF
atpH
atpI
rps2
rpoC2
rpoC1
rpoB
trnC GCA
psbJ
psbF
psbE
petL
petG
trnW CCA
trnP UGG
psaJ
rpl33
rps18
rpl20
5'-rps12
clpP
psbB
psbT
psbH
petB
petD
psbN

Figure S7A

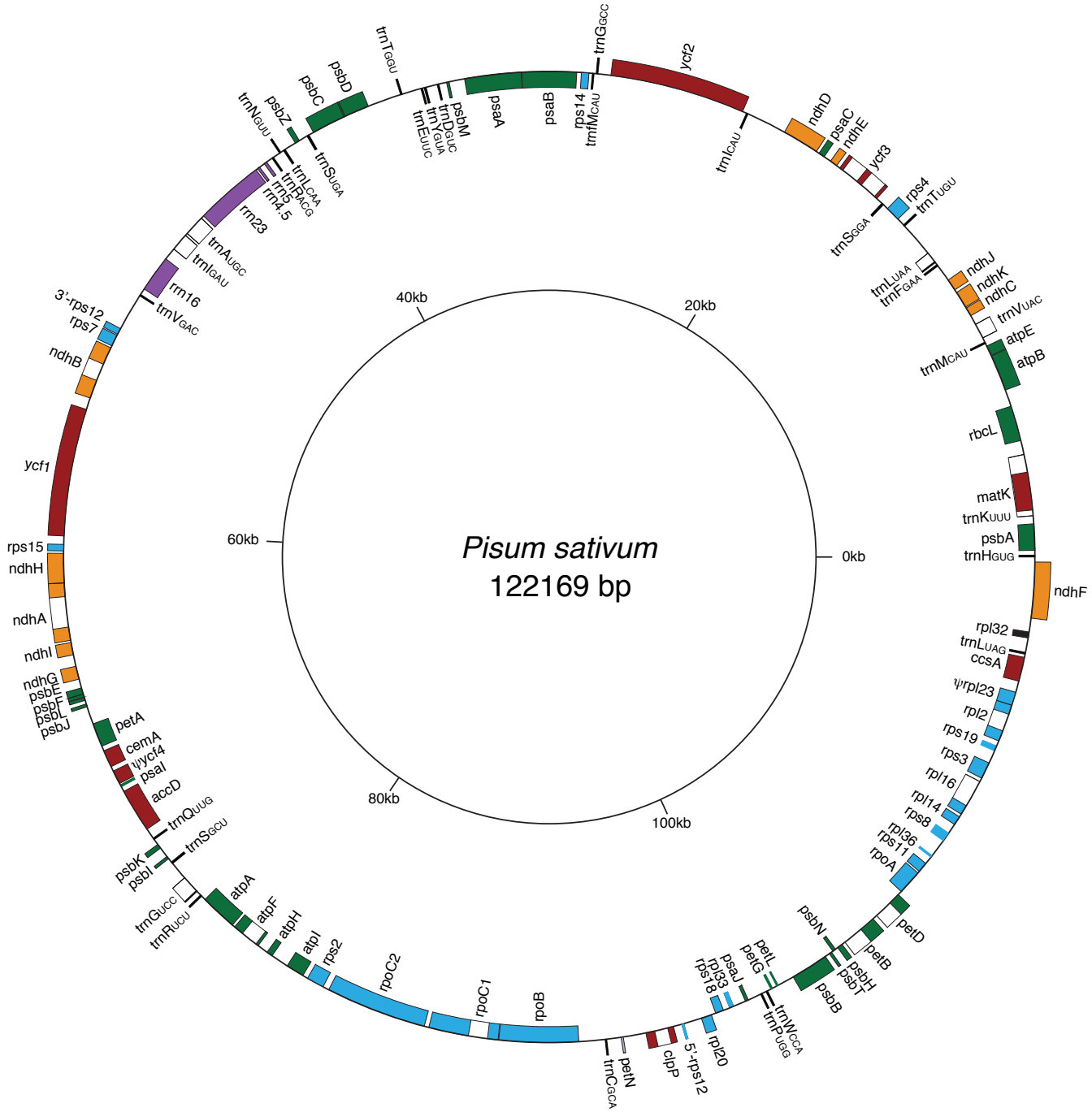

Pisum sativum
122169 bp
0kb
20kb
40kb
60kb
80kb
100kb
trnGGCC
ycf2
trnICAU
ndhD
psaC
ndhE
ycf3
rps4
trnTUGU
trnSGGA
trnLUAA
trnFGAA
ndhJ
ndhK
ndhC
trnVUAC
trnMCAU
atpE
atpB
rbcL
matK
trnKUUU
psbA
trnHGUG
ndhF
rpl32
trnLUAG
ccsA
ψrpl23
rpl2
rps19
rps3
rpl16
rpl14
rps8
rpl36
rps11
rpoA
petD
petB
psbH
psbT
psbB
psbN
trnWCCA
trnPUGG
petL
petG
psaJ
rpl33
rps18
rpl20
5'-rps12
clpP
petN
trnCGCA
rpoB
rpoC1
rpoC2
rps2
atpI
atpH
atpF
atpA
trnRUCU
trnGUCC
psbI
psbK
trnSGCU
trnQUUG
accD
psaI
ycf4
cemA
petA
psbJ
psbL
psbF
psbE
ndhG
ndhI
ndhA
ndhH
rps15
ycf1
ndhB
rps7
3'-rps12
trnVGAC
rrn16
trnIGAU
trnAUGC
rrn23
rrn4.5
rrn5
trnRACG
trnLCAA
trnSUGA
psbZ
trnNGUU
psbC
psbD
trnTGGU
trnEUUC
trnYGUA
trnDGUC
psbM
psaA
psaB
rps14
trnfMCAU

Figure S7B