SUPPLEMENTARY INFORMATION

# Taxonomic Reassessment of *Phytolacca* species through Comparative Chloroplast Genome, Morphological, and Phylogenetic Analysis

**Table S1.** Raw and trimmed read data.

| Scientific name | Input reads | Trimmed reads | | Raw bases | Trimmed bases |
|---|---|---|---|---|---|
| *P. acinosa* | 208,993,788 | 204,511,620 | 94.98% | 31,558,061,988 | 29,973,847,276 |

**Table S2.** Genome assembly data for the *Phytolacca acinosa* chloroplast genome.

| Scientific name | Aligned reads (#) | Coverage (x) | Cp genome length (bp) |
|---|---|---|---|
| *P. acinosa* | 913,560 | 878 | 156,433 |

**Table S3.** Read data and sequencing platform information for the *Phytolacca acinosa* chloroplast genome.

| Scientific name | Reads | Bases | Sequencing platform |
|---|---|---|---|
| *P. acinosa* | 208,993,788 | 31,558,061,988 | Novaseq 6000 |

**Table S4.** Gene groups and their corresponding genes identified in the chloroplast genome of *Phytolacca acinosa*, categorized by functional roles such as photosynthesis, ribosomal proteins, and other essential cellular processes.

| **Group of genes** | **Name of genes** |
|---|---|
| Photosystem I | *psaA*, *B*, *C*, *I*, *J*, *pafI*[2)], *pafII* |
| Photosystem II | *psbA*, *B*, *C*, *D*, *E*, *F*, *H*, *I*, *J*, *K*, *L*, *M*, *T*, *Z*, *pbf1* |
| Cytochrome b6/f | *petA*, *B*[1)], *D*[1)], *G*, *L*, *N* |
| ATP synthase | *atpA*, *B*, *E*, *F*[1)], *H*, *I* |
| Rubisco | *rbcL* |
| NADH oxidoreductase | *ndhA*[1)], *B*[1),3)], *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K* |
| Large subunit ribosomal proteins | *rpl2*[3)], *14*, *16*[1)], *20*, *22*, *23*[3)], *32*, *33*, *36* |
| Small subunit ribosomal proteins | *rps2*, *3*, *4*, *7*[3)], *8*, *11*, *12*[2),3),4)], *14*, *15*, *16*[1)], *18*, *19* |
| RNA polymerase | *rpoA*, *B*, *C1*[1)], *C2* |
| Unknown function protein coding gene | *ycf1*[3)], *ycf2*[3)] |
| Other genes | *accD*, *ccsA*, *cemA*, *clpP1*[2)], *infA*, *matK* |
| Ribosomal RNAs | *rrn4.5*[3)], *rrn5*[3)], *rrn16*[3)], *rrn23*[3)] |
| Transfer RNAs | *trnA-UGC*[1),3)], *trnC-GCA*[3)], *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, *trnG-GCC*, *trnG-UCC*[1)], *trnH-GUG*, *trnI-CAU*[3)], *trnK-UUU*[1)], *trnL-CAA*[3)], *trnL-GAU*[1),3)], *trnL-UAA*[1)], *trnL-UAG*, *trnM-CAU*, *trnN-GUU*[3)], *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*[3)], *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, *trnT-GGU*, *trnT-UGU*, *trnV-GAC*[3)], *trnV-UAC*[1)], *trnW-CCA*, *trnY-GUA* |

1) Gene containing a single intron, 2) Gene containing two introns, 3) Two gene copies in IRs, 4) Trans-splicing gene

**Table S5.** Intron-containing genes and their exon-intron structures in the chloroplast genome of *Phytolacca acinosa*. Genes are listed with their respective regions (LSC, SSC, IR), and the lengths of exon and intron segments are provided.

| No. | Gene | Region | Exon I | Intron I | Exon II | Intron II | Exon III |
|---|---|---|---|---|---|---|---|
| 1 | *trnK-UUU* | LSC | 37 | 2499 | 35 | | |
| 2 | *rps16* | LSC | 40 | 902 | 233 | | |
| 3 | *trnG-UCC* | LSC | 23 | 703 | 48 | | |
| 4 | *atpF* | LSC | 144 | 757 | 411 | | |
| 5 | *rpoC1* | LSC | 430 | 824 | 1613 | | |
| 6 | *pafI* | LSC | 124 | 773 | 228 | 748 | 155 |
| 7 | *trnL-UAA* | LSC | 35 | 555 | 50 | | |
| 8 | *trnV-UAC* | LSC | 37 | 596 | 37 | | |
| 9 | *rps12* | LSC | 114 | | 232 | | 26 |
| 10 | *clpP1* | LSC | 71 | 886 | 292 | 596 | 228 |
| 11 | *petB* | LSC | 6 | 783 | 642 | | |
| 12 | *petD* | LSC | 8 | 693 | 475 | | |
| 13 | *rpl16* | LSC | 9 | 981 | 399 | | |
| 14 | *ndhB* | IR | 777 | 668 | 756 | | |
| 15 | *trnL-GAU* | IR | 37 | 946 | 35 | | |
| 16 | *trnA-UGC* | IR | 38 | 823 | 35 | | |
| 17 | *ndhA* | SSC | 552 | 1058 | 540 | | |

**Table S6.** dN/dS values for various genes across six *Phytolacca* species, showing evolutionary pressure on protein-coding genes. Conserved genes have similar dN/dS values across species, while genes with varying values may be under different selective pressures, suggesting species-specific adaptations or functional divergence.

| | *P. acinosa* | *P. insularis* | *P. polyandra* | *P. japonica* | *P. latbenia* | *P. americana* |
|---|---|---|---|---|---|---|
| ***accD*** | 0.2468 | 0.2468 | 0.2503 | 0.2355 | 0.263 | 0.2287 |
| ***atpA*** | 0.0955 | 0.0955 | 0.0955 | 0.0955 | 0.0955 | 0.0821 |
| ***atpB*** | 0.021 | 0.021 | 0.021 | 0.0203 | 0.021 | 0.0306 |
| ***atpE*** | 0.1513 | 0.1513 | 0.1513 | 0.1513 | 0.1513 | 0.0707 |
| ***atpF*** | 0.1953 | 0.1953 | 0.1953 | 0.1953 | 0.1953 | 0.1193 |
| ***atpI*** | 0.3432 | 0.0898 | 0.3432 | 0.3432 | 0.3432 | 0.118 |
| ***ccsA*** | 0.2498 | 0.2498 | 0.2498 | 0.2533 | 0.2666 | 0.2668 |
| ***cemA*** | 0.2904 | 0.2904 | 0.2992 | 0.2992 | 0.31 | 0.2797 |
| ***clpP1*** | 0.0458 | 0.0458 | 0.0535 | 0.0535 | 0.0535 | 0.0467 |
| ***infA*** | 0.0878 | 0.0878 | 0.0878 | 0.0878 | 0.0878 | 0.0878 |
| ***matK*** | 0.3298 | 0.3298 | 0.3298 | 0.3298 | 0.3298 | 0.3203 |
| ***ndhA*** | 0.2854 | 0.3031 | 0.3031 | 0.3031 | 0.3031 | 0.274 |
| ***ndhB*** | 0.6533 | 0.6533 | 0.6533 | 0.6533 | 0.6533 | 0.6533 |
| ***ndhD*** | 0.1443 | 0.1443 | 0.1443 | 0.1443 | 0.1225 | 0.1715 |
| ***ndhE*** | 0.0212 | 0.0212 | 0.0212 | 0.0212 | 0.0212 | 0.0257 |
| ***ndhF*** | 0.1717 | 0.1717 | 0.1717 | 0.1631 | 0.1793 | 0.3969 |
| ***ndhG*** | 0.7412 | 0.7412 | 0.7412 | 0.8394 | 0.6524 | 1.0871 |
| ***ndhH*** | 0.0271 | 0.0271 | 0.0248 | 0.0271 | 0.0265 | 0.0683 |
| ***ndhI*** | 0.0675 | 0.0675 | 0.0675 | 0.0675 | 0.0675 | 0.0691 |
| ***ndhJ*** | 0.0607 | 0.0607 | 0.0607 | 0.0607 | 0.0607 | 0.0842 |
| ***ndhK*** | 0.1312 | 0.1284 | 0.1284 | 0.1284 | 0.1478 | 0.1356 |
| ***pafI*** | 0.0443 | 0.0443 | 0.0443 | 0.0443 | 0.0443 | 0.0387 |
| ***pafII*** | 0.1594 | 0.1594 | 0.1594 | 0.1848 | 0.1844 | 0.1271 |
| ***petA*** | 0.2621 | 0.2621 | 0.2621 | 0.2621 | 0.2405 | 0.1982 |
| ***petB*** | 0.0679 | 0.0606 | 0.0606 | 0.0606 | 0.0606 | 0.0468 |
| ***petD*** | 0.0351 | 0.0351 | 0.0351 | 0.041 | 0.0485 | 0.0417 |
| ***psaA*** | 0.0414 | 0.0414 | 0.0414 | 0.0414 | 0.0502 | 0.0367 |
| ***psaB*** | 0.0501 | 0.0501 | 0.053 | 0.0623 | 0.0501 | 0.0699 |
| ***psaC*** | 0.2082 | 0.2082 | 0.2082 | 0.2082 | 0.2082 | 0.0986 |
| ***psaI*** | 0.1366 | 0.1366 | 0.1366 | 0.1366 | 0.1366 | 0.1366 |
| ***psaJ*** | 0.4177 | 0.4177 | 0.4177 | 0.4177 | 0.4177 | 0.3974 |
| ***psbA*** | 0.0274 | 0.0274 | 0.0274 | 0.0274 | 0.0274 | 0.0856 |
| ***psbB*** | 0.0402 | 0.0402 | 0.0402 | 0.0402 | 0.0402 | 0.0398 |
| ***psbC*** | 0.0191 | 0.0191 | 0.0191 | 0.0191 | 0.0191 | 0.0178 |
| ***psbD*** | 0.0302 | 0.0302 | 0.0302 | 0.0302 | 0.0302 | 0.0549 |
| ***psbI*** | 0.2039 | 0.2039 | 0.2039 | 0.2039 | 0.2039 | 0.2109 |
| ***psbK*** | 0.1574 | 0.1574 | 0.1574 | 0.2205 | 0.2205 | 0.2205 |
| ***rbcL*** | 0.3096 | 0.3096 | 0.3096 | 0.2981 | 0.2862 | 0.2962 |
| ***rpl2*** | 0.0673 | 0.0673 | 0.0673 | 0.1353 | 0.0673 | 0.0578 |
| ***rpl14*** | 0.065 | 0.065 | 0.065 | 0.065 | 0.065 | 0.0597 |
| ***rpl20*** | 0.1964 | 0.1964 | 0.1964 | 0.1964 | 0.1964 | 0.2599 |
| ***rpl22*** | 0.3833 | 0.3833 | 0.3833 | 0.3833 | 0.3872 | 0.3375 |
| ***rpl32*** | 0.3595 | 0.3595 | 0.3595 | 0.522 | 0.3595 | 0.2416 |
| ***rpl33*** | 0.3237 | 0.3237 | 0.3237 | 0.3237 | 0.3237 | 0.1022 |
| ***rpoA*** | 0.2538 | 0.2538 | 0.2538 | 0.2746 | 0.2538 | 0.2101 |

| | | | | | | |
|---|---|---|---|---|---|---|
| ***rpoB*** | 0.1082 | 0.1082 | 0.1082 | 0.1115 | 0.117 | 0.0915 |
| ***rpoC1*** | 0.104 | 0.104 | 0.1109 | 0.104 | 0.104 | 0.1252 |
| ***rpoC2*** | 0.3388 | 0.3391 | 0.3391 | 0.3512 | 0.3189 | 0.3064 |
| ***rps2*** | 0.1239 | 0.1239 | 0.1239 | 0.1239 | 0.1239 | 0.1228 |
| ***rps3*** | 0.1084 | 0.1084 | 0.1084 | 0.1079 | 0.1079 | 0.1177 |
| ***rps4*** | 0.2687 | 0.2687 | 0.2687 | 0.2687 | 0.2687 | 0.2091 |
| ***rps8*** | 0.7359 | 0.7359 | 0.1076 | 0.1076 | 0.1076 | 0.7491 |
| ***rps11*** | 0.1011 | 0.1011 | 0.1011 | 0.1011 | 0.1011 | 0.1151 |
| ***rps14*** | 0.2765 | 0.2765 | 0.2765 | 0.4731 | 0.2765 | 0.1495 |
| ***rps15*** | 0.5071 | 0.5071 | 0.5071 | 0.5071 | 0.5071 | 0.2486 |
| ***rps19*** | 0.0385 | 0.0385 | 0.0385 | 0.0385 | 0.0385 | 0.0515 |
| ***ycf1*** | 0.8309 | 0.8317 | 0.8461 | 0.8728 | 0.8421 | 0.7624 |
| ***ycf2*** | 1.2226 | 1.2235 | 1.2218 | 1.1738 | 1.2056 | 0.8952 |

**Table S7.** Voucher specimen information for chloroplast genome used in this study.

| Species | Collection information | Coordinates | Voucher number |
|---|---|---|---|
| *P. acinosa* | Cheongnyangsan-gil, Myeongho-myeon, Bonghwa-gun, Gyeongsangbuk-do, South Korea | 36°47'18.9"N 128°54'26.1"E | CBU-C-1 |

**Table S8.** GenBank accession numbers of species used for maximum likelihood (ML), Bayesian inference (BI) and divergent phylogenetic tree construction in this study.

| No. | Subfamily | Taxon | GenBank accession number |
|---|---|---|---|
| 1 | | *Phytolacca acinosa* | PQ584746 |
| 2 | | *Phytolacca acinosa* | OM403713 |
| 3 | | *Phytolacca acinosa* | OM403715 |
| 4 | | *Phytolacca insularis* | NC_041113 |
| 5 | | *Phytolacca polyandra* | NC_068263 |
| 6 | | *Phytolacca japonica* | NC_068264 |
| 7 | Phytolaccoideae | *Phytolacca latbenia* | NC_068262 |
| 8 | | *Phytolacca americana* var. *americana* | MH286315 |
| 9 | | *Phytolacca americana* | OM403707 |
| 10 | | *Phytolacca thyrsiflora* | NC_068265 |
| 11 | | *Phytolacca icosandra* | NC_068266 |
| 12 | | *Phytolacca rivinoides* | NC_068267 |
| 13 | | *Phytolacca dioica* | NC_068261 |
| 14 | Agdestidoideae | *Agdestis clematidea* | MH286339 |

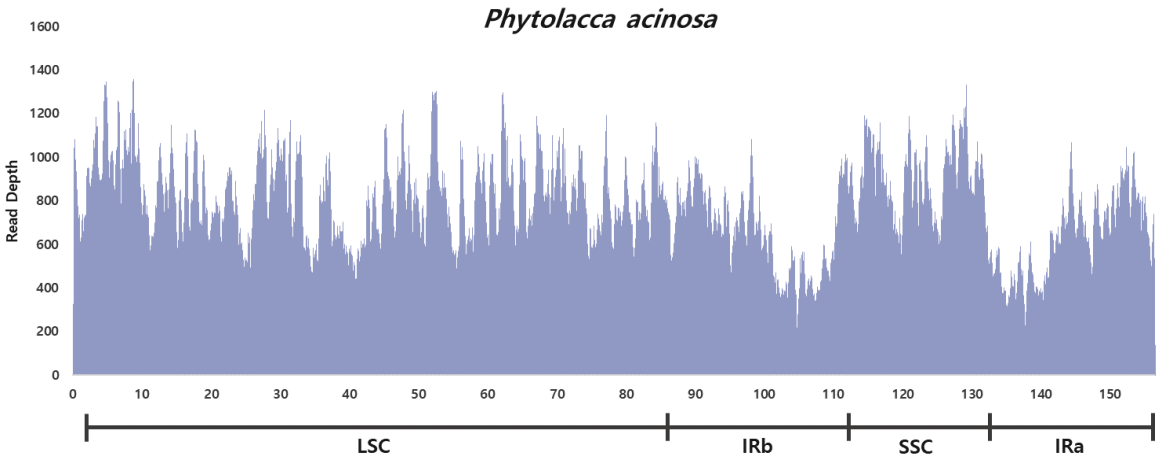


**Figure S1.** Coverage showing the number of paired-end reads mapped to the complete chloroplast genome of *Phytolacca acinosa*. LSC, large single-copy region; SSC, small single-copy region; IRa, inverted repeat a; IRb, inverted repeat b.

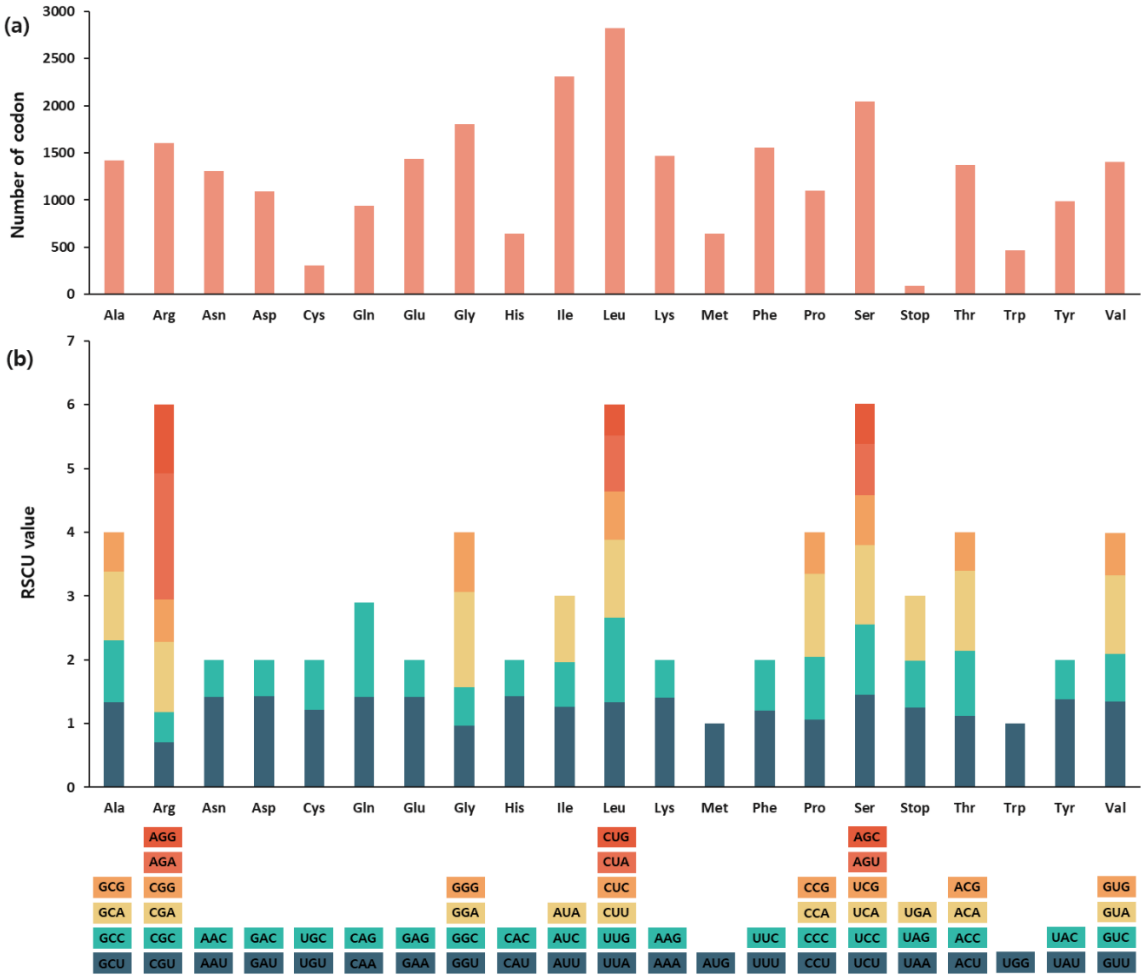


**Figure S2.** Codon usage and Relative Synonymous Codon Usage (RSCU) values for the 20 amino acids and stop codons in *P. acinosa*. (a) Frequency of amino acids in the protein-coding sequences. (b) RSCU values for 20 amino acids and stop codons from 79 protein-coding genes.

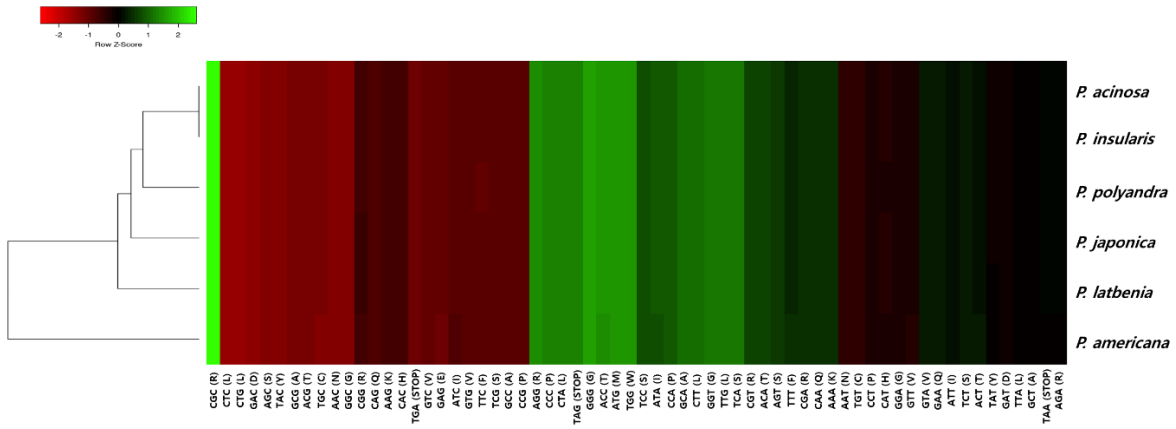


**Figure S3.** Codon usage heatmap of protein-coding genes across six *Phytolacca* species. Green indicates high RSCU values, and red indicates low RSCU values. Codon usage patterns were analyzed using hierarchical clustering, shown by the dendrogram on the left.

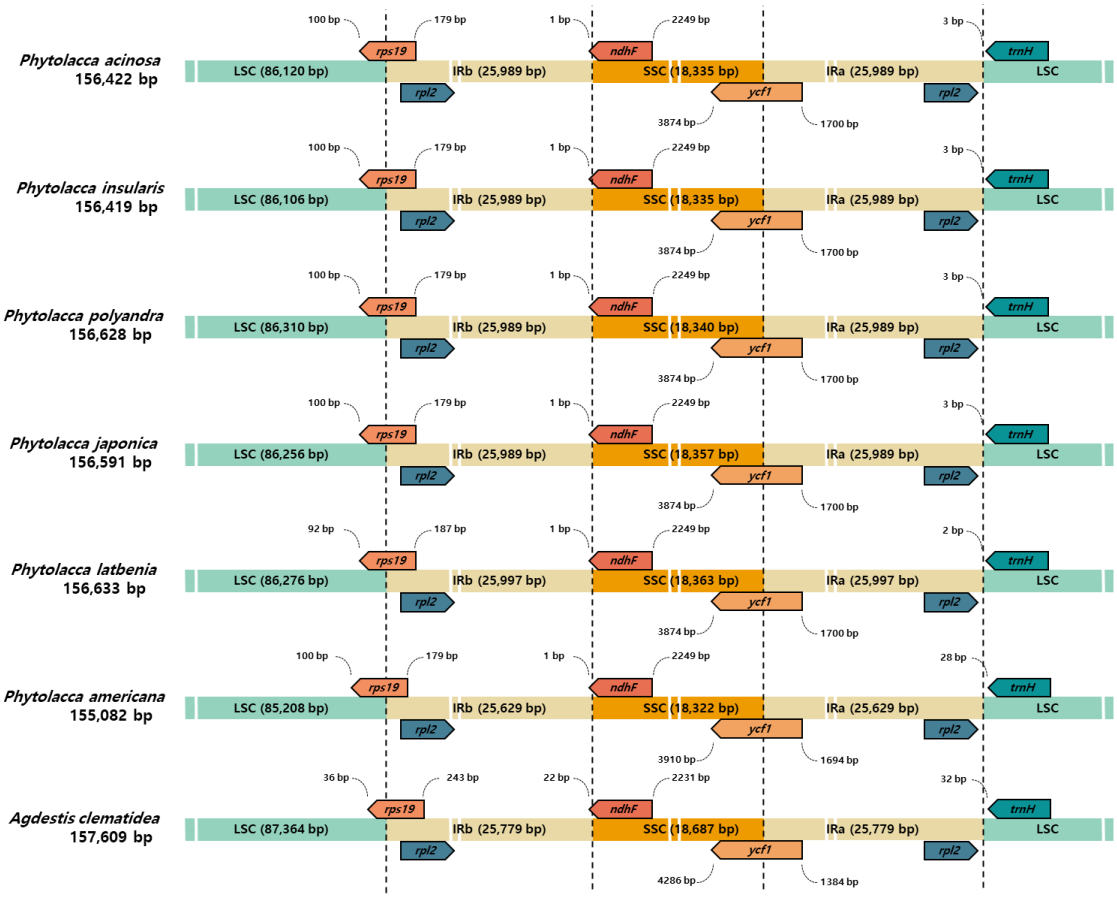


**Figure S4.** Comparison of chloroplast genome structures in six *Phytolacca* species and *Agdestis clematidea*. Each genome was composed of an LSC region, an SSC region, and IRa and IRb regions. The lengths of these regions are indicated, with structural variations, such as expansions, contractions, and shifts of the IR boundaries across the species, highlighted. LSC, large single-copy region; SSC, small single-copy region; IRa, inverted repeat a; IRb, inverted repeat b.