

Chloroplast genomes are the main source of molecular data for phylogenetic analyses among plants. Here we take the set of chloroplast gene families annotated in ChloroplastDB[1] and analyze them with primers4clades[2], an easy-to-use web server that implements a fully automatic PCR primer design pipeline for cross-species amplification of novel sequences belonging to user-specified clades or taxa. **Results:**

- 1) Assessment of the phylogenetic value of chloroplast genes present in 10+ genomes finds that *matK* and *rbcL*, broadly accepted as universal barcoding candidates for land plants[3], are among the most informative genes. In addition several other loci are found with similar phylogenetic information content.
- 2) Analysis of 32 *matK* gene sequences illustrates how the server can help targeting primer design to selected clades of interest, and how the number and quality of obtained pairs of primers depends on the variability of the input nucleotide sequences.

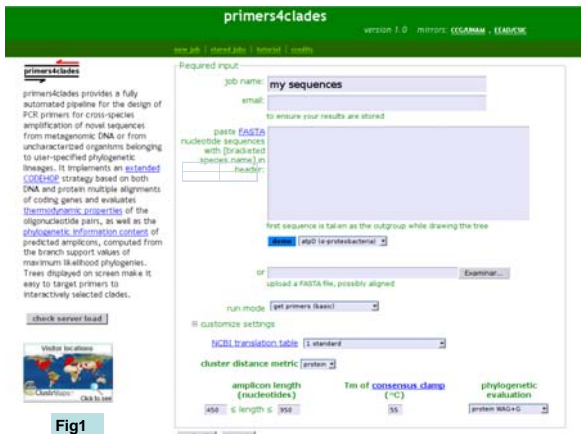
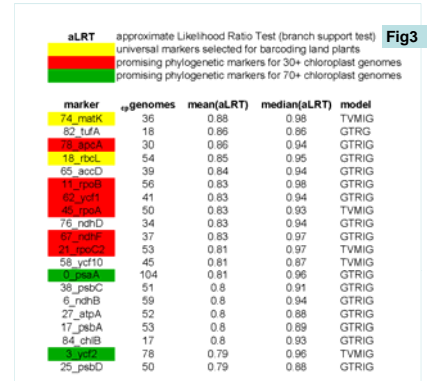
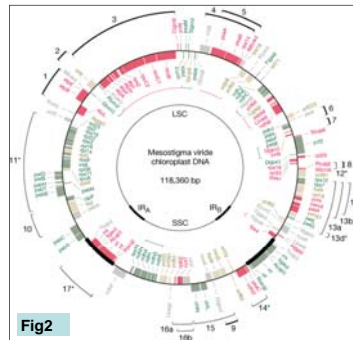


Fig1



primer design example: 32 matK homologues

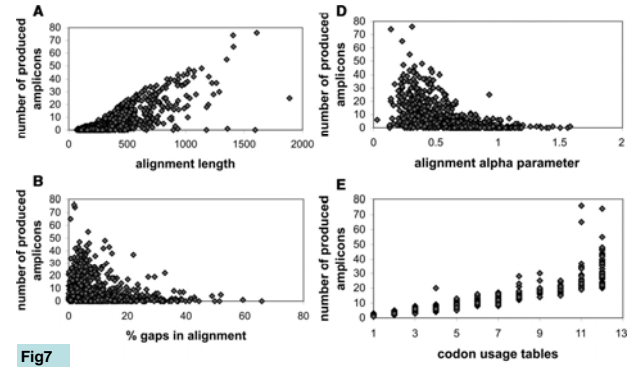
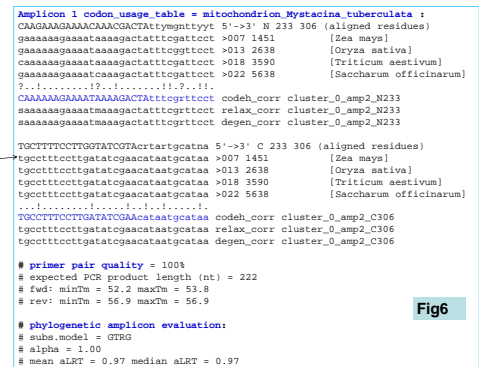
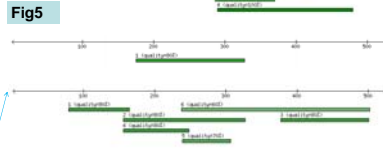


Figure 1. Web interface of primers4clades .
Figure 2. Genomic map of *Mesostigma viride* .
Figure 3. Evaluation of phylogenetic information content of genes present in 10+ chloroplast genomes, using ChloroplastDB gene families[1]. Note that accepted barcoding markers *matK* and *rbcL* [3] are high in the list. Best-fit DNA-substitution models are reported, where GTRIG(8 free params) is frequent.
Figure 4. DNA-metric neighbour-joining tree generated by primers4clades for 32 *matK* genes.
Figure 5. Amplicon maps for three different clusters of *matK* sequences.
Figure 6. Sample *matK* primer pair for cereals, inferred from corn, sugar cane, rice and wheat.
Figure 7. Large scale benchmark of primers4clades highlighting alignment-related variables that affect performance[2].

- [1] Cui, L. *et al.* (2006) ChloroplastDB: the chloroplast genome database. *Nucleic Acids Research*, **34**: D692-696. (<http://chloroplast.cbio.psu.edu>)
- [2] Contreras-Moreira, B. *et al.* (2009) primers4clades: a web server that uses phylogenetic trees to design lineage-specific PCR primers for metagenomic and diversity studies. *Nucleic Acids Research*, **37**(Web Server issue):W95-W100.
- [3] CBOL Plant Working Group (2009) A DNA barcode for land plants. *PNAS*, **106**(31): 12794-12797.