

## Supplementary Legends

**Supplementary Figure S1.** The distribution of contig lengths.

The lengths of the 120,586 contigs generated by hybrid assembly are shown.

**Supplementary Figure S2.** A neighbor-joining tree of amino acid sequences of casbene synthase genes.

The list of gene names with accession numbers is shown in Supplementary Table S9.

Bootstrap values were calculated from 1000 replicates.

**Supplementary Figure S3.** A neighbor-joining tree of amino acid sequences of disease resistance genes.

The list of gene names with accession numbers is shown in Supplementary Table S11.

Bootstrap values were calculated from 500 replicates.

**Supplementary Figure S4.** A neighbor-joining tree of amino acid sequences of MIKC type II MADS-box genes.

The list of gene names with accession numbers is shown in Supplementary Table S12.

Bootstrap values were calculated from 1000 replicates.

**Supplementary Figure S5.** A neighbor-joining tree of amino acid sequences of flowering-related genes.

The list of gene names with accession numbers is shown in Supplementary Table S14.

Bootstrap values were calculated from 1000 replicates.

**Supplementary Figure S6.** A neighbor-joining tree of amino acid sequences of COL genes.

The list of gene names with accession numbers is shown in Supplementary Table 15.

Bootstrap values were calculated from 1000 replicates.

**Supplementary Figure S7.** Amplification ratio of the 100 genome derived microsatellite markers tested.

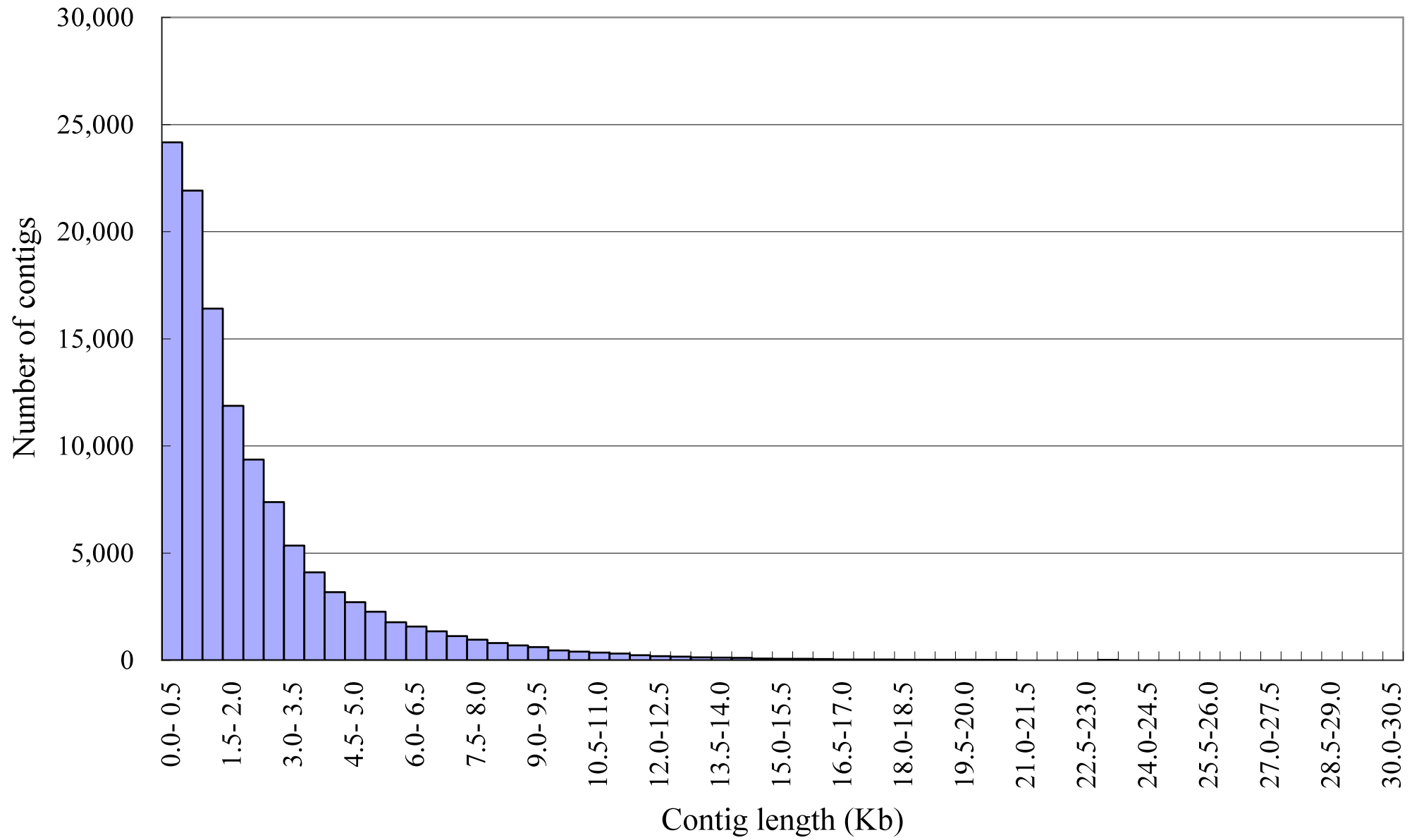
The dark gray and light gray colors show the number of markers indicating non-amplification and non-specific amplification, respectively. The light green, red,

orange and yellow colors represent the number of amplified markers detecting single, double, triple and quad alleles among the 12 *Jatropha* lines, respectively.

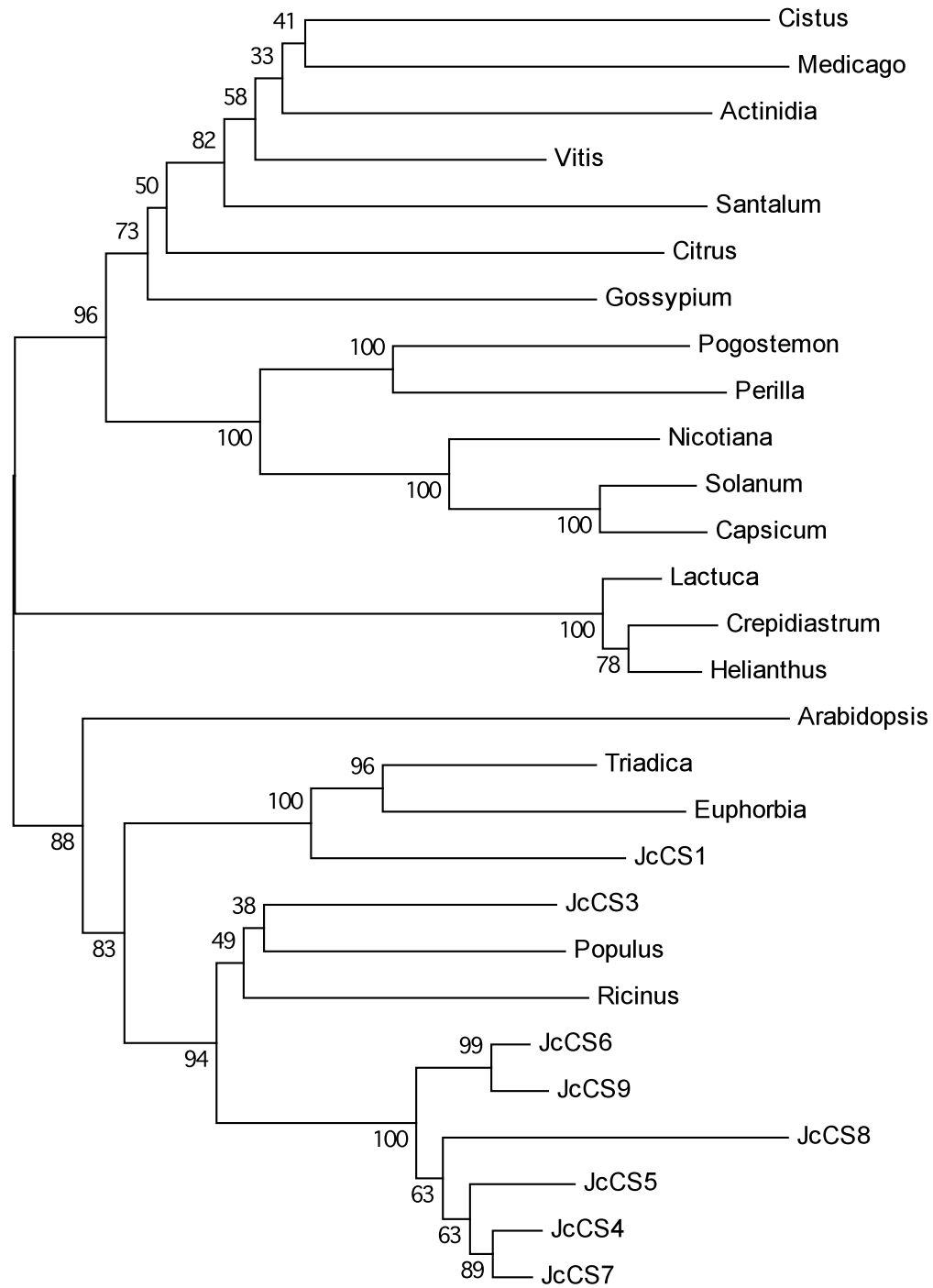
**Supplementary Figure S8.** Distribution of PIC values.

**Supplementary Figure S9.** An UPGMA genetic tree of the 12 *Jatropha* lines.

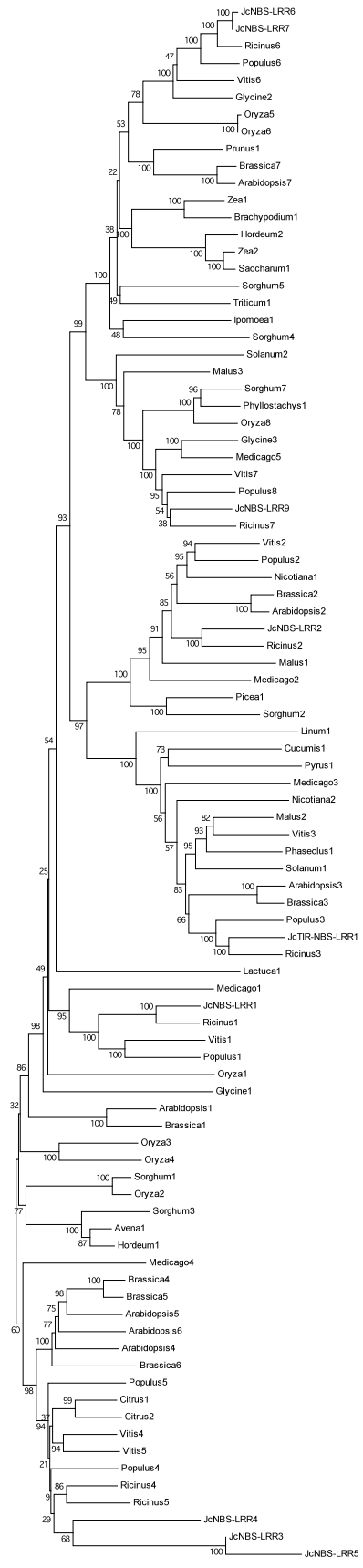
Blue, green and red squares indicate lines derived from Asia, Africa and meso-America, respectively.



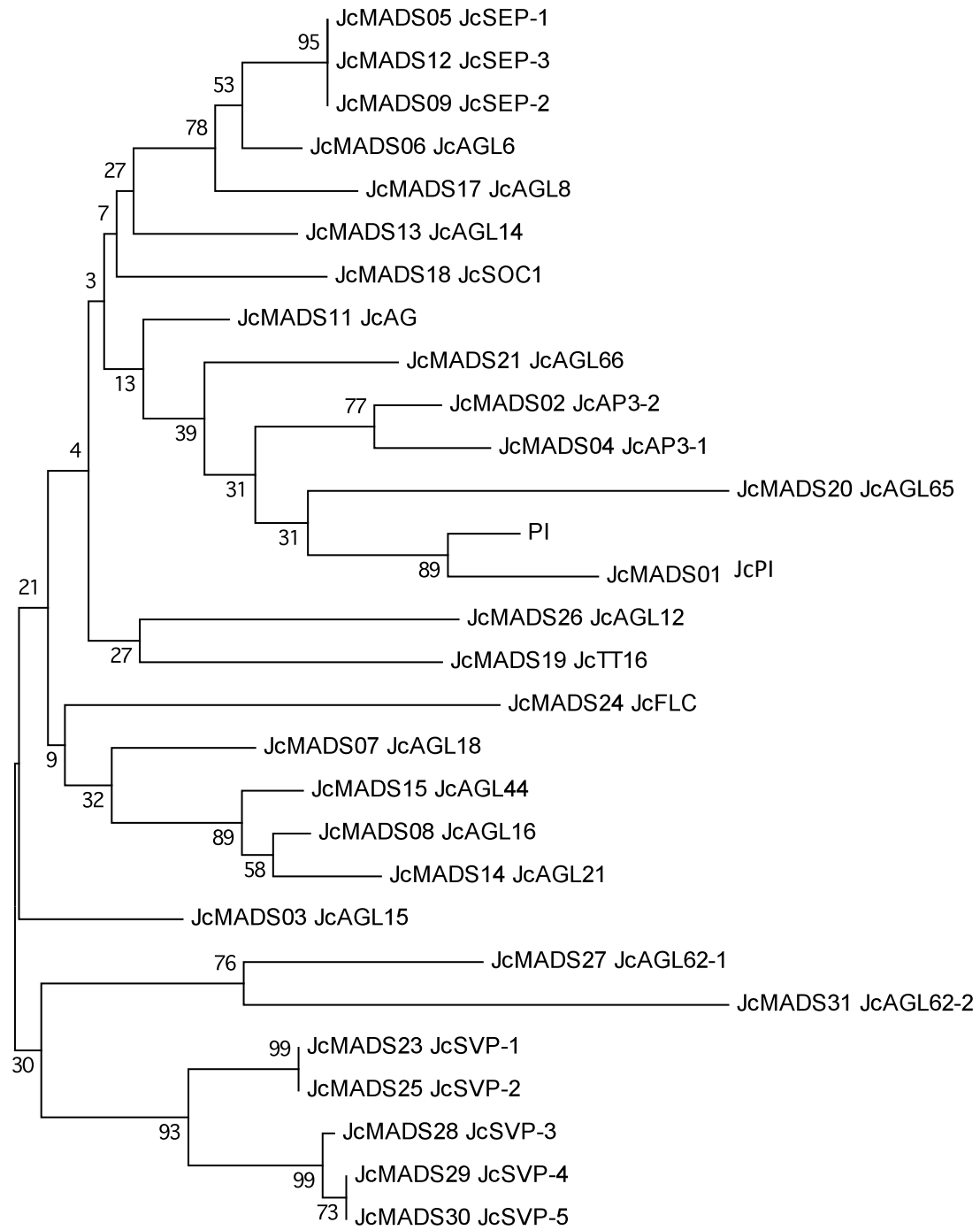
# Casbene synthase



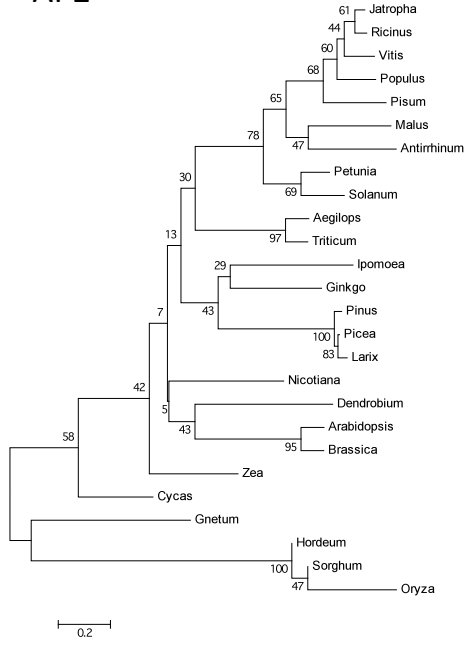
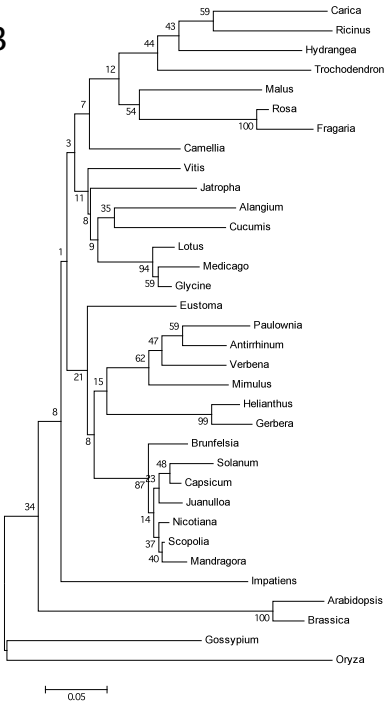
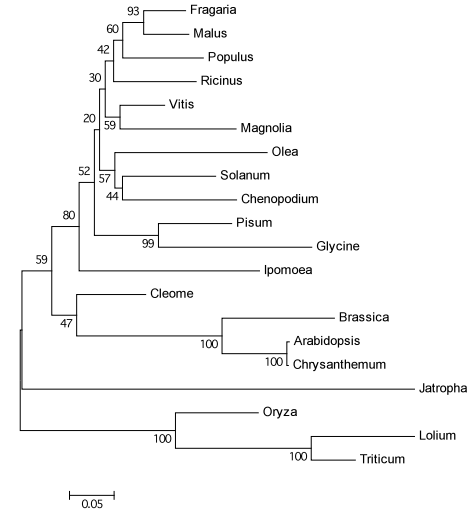
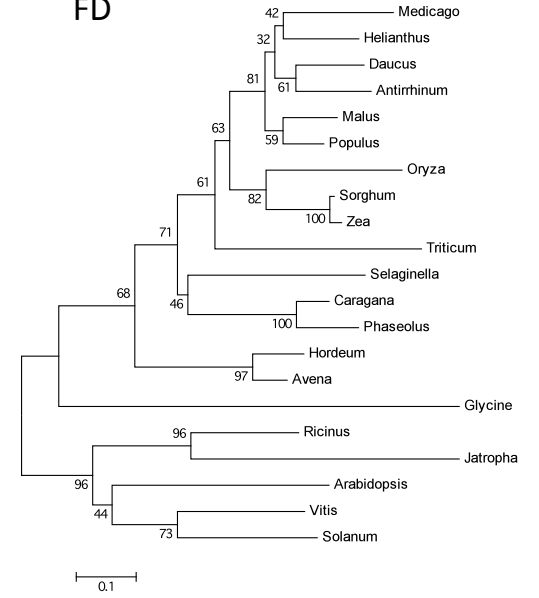
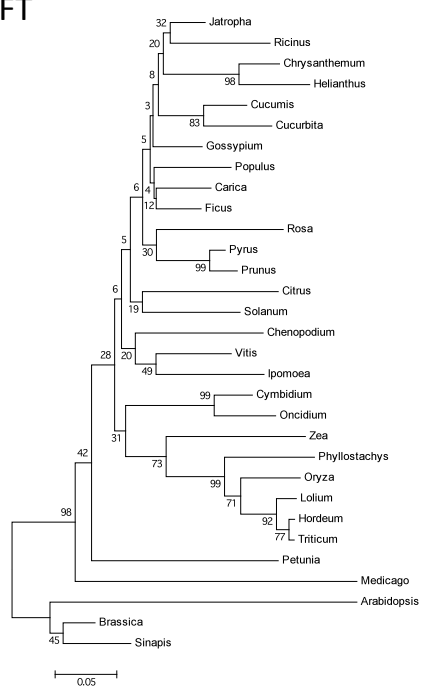
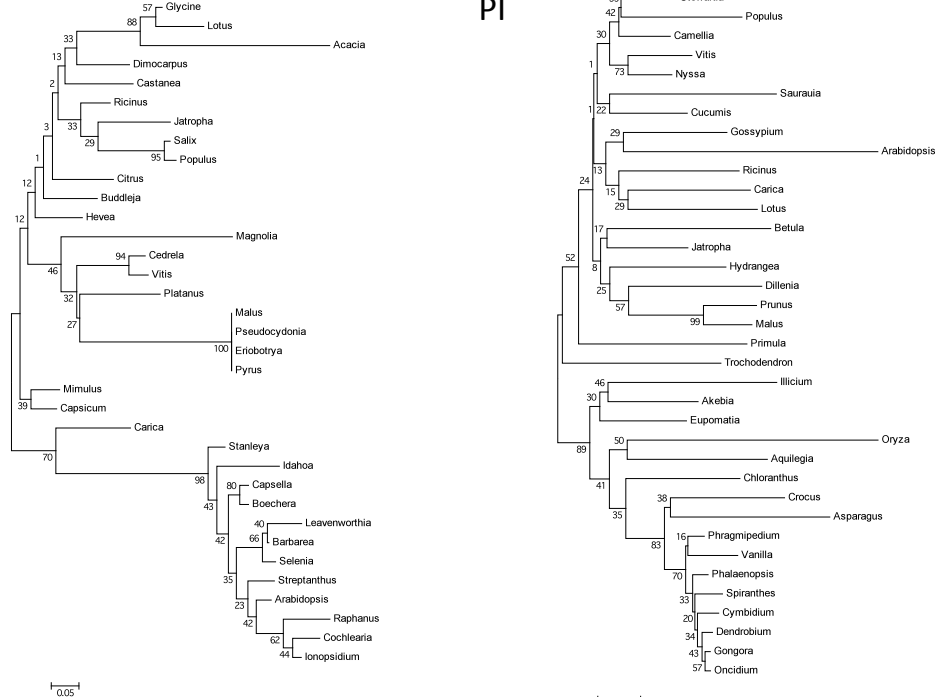
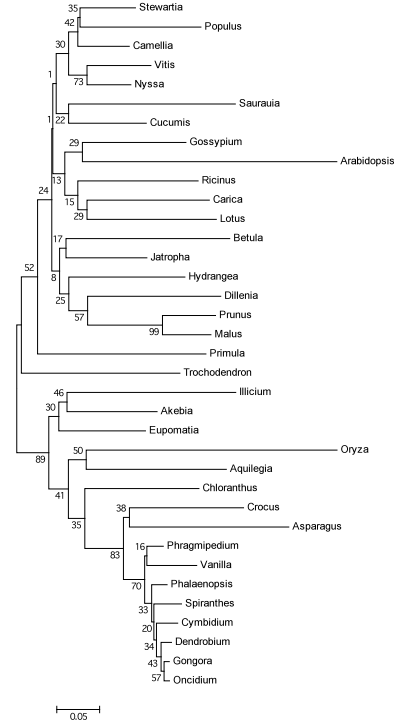
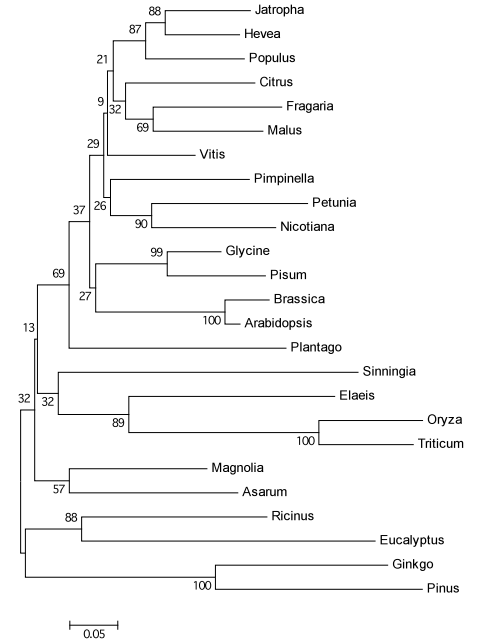
0.1



# MADS domain



0.05

**AP2****AP3****CO****FD****FT****LFY****PI****SOC1 (AGL20)**

# COL

