Available online at www.sciencedirect.com
ScienceDirect
Organisms, Diversity & Evolution 7 (2007) 55–77
ORGANISMS DIVERSITY & EVOLUTION
www.elsevier.de/ode

Phylogeny of basal eudicots: Insights from non-coding and rapidly evolving DNA

Andreas Worberg[a], Dietmar Quandt[b], Anna-Magdalena Barniske[b], Cornelia Löhne[a], Khidir W. Hilu[c], Thomas Borsch[a,*]

[a]*Nees-Institut für Biodiversität der Pflanzen, Rheinische Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany*
[b]*Institut für Botanik, Technische Universität Dresden, Zellescher Weg 20b, 01217 Dresden, Germany*
[c]*Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24061, USA*

Received 13 June 2006; accepted 10 August 2006

Dedicated to Wilhelm Barthlott on the occasion of his 60th birthday

Abstract

Sequence data of the *trnL* group I intron, the *petD* group II intron, the *trnL-F* and *petB-D* spacers, and the rapidly evolving *matK* gene were analysed from all families of the basal eudicot grade and from representatives of 19 core eudicot orders. The dataset comprised 5654 positions of aligned sequence plus a matrix of 1087 binary indel characters. Mutational hotspots correspond in number and extension to hotspots already known from basal angiosperms and, with respect to secondary structure, are generally located in terminal parts of stem-loop regions. Parsimony, Bayesian, and likelihood analyses depict Ranunculales as sister to all remaining eudicots with maximum support. The branching order in the basal eudicot grade is further resolved as Sabiales, Proteales, Trochodendrales, and Buxales. Nearly all of the backbone nodes gain high confidence, except for the node showing Proteales diverging before Trochodendrales, which is only moderately supported (83% JK). In Ranunculales, the woody Eupteleaceae are first-branching, with Papaveraceae plus Fumariaceae coming next. Within Proteales, *Nelumbo* is clearly resolved as sister to a Platanaceae–Proteaceae clade. Gunnerales are found as the first branch in core eudicots, with maximum support in the combined analysis. This node is also resolved with *matK* alone, but unsupported. It appears that the combined analysis of sequence data from rapidly evolving and non-coding genomic regions leads to significantly improved statistical support values in comparison to earlier studies of basal eudicots using multiple conserved genes.

Keywords: Angiosperms; *matK*; *trnL-F*; *petD*; Molecular evolution
See also **Electronic Supplement** at doi:10.1016/j.ode.2006.08.001

Introduction

The eudicot clade comprises the vast majority of angiosperm diversity, with an estimated 200,000 species (Drinnan et al. 1994). The clade was first recognized by Donoghue and Doyle (1989) and Doyle and Hotton

*Corresponding author. Tel.: +49 228 732681; fax: +49 228 733120.
E-mail address: borsch@uni-bonn.de (T. Borsch).

1439-6092/$ - see front matter
doi:10.1016/j.ode.2006.08.001

(1991) based on morphological characters. Broad-scale molecular analyses of angiosperms using single genes (Chase et al. 1993; Savolainen et al. 2000a) have consistently recovered the eudicots. The clade has gained maximum support when three or more genes were combined (Qiu et al. 2000; Soltis et al. 2000; Kim et al. 2004). More recently, the analysis of partial *matK* sequences alone yielded 96% JK (Hilu et al. 2003). The eudicots share tricolpate and tricolpate-derived pollen (Donoghue and Doyle 1989; Nandi et al. 1998; Hoot et al. 1999). Thus they have also been called the tricolpate clade (Donoghue and Doyle 1989).

Using DNA data, a number of lineages (Ranunculales, Proteales, Sabiaceae, Buxaceae plus Didymelaceae, and Trochodendraceae plus Tetracentraceae) have been identified as representing the earliest branches in eudicots (Chase et al. 1993; Savolainen et al. 2000b; Soltis et al. 2000, 2003; Hilu et al. 2003), whereas large groups such as asterids, Caryophyllales, rosids, Santalales, and Saxifragales were shown to belong to the core eudicots. The core eudicot node is one of the best-supported nodes within the angiosperm tree (Hilu et al. 2003; Soltis et al. 2003) and obviously marks a major shift in angiosperm evolution. The core eudicot node is also identified by recent analyses of MADS-box genes, where non-core tricolpate clades only have the eu*FUL* gene family and lack the eu*AP1* gene family. Thus, this gene duplication is a synapomorphy for core eudicots (Litt and Irish 2003; Kim et al. 2004; Kramer et al. 2004). Molecular clock dating inferred the eudicots to have an age of 131–125 mya (Magallon et al. 1999; Anderson et al. 2005), whereas the core eudicot node is estimated at 113–116 mya (Magallon et al. 1999; Anderson et al. 2005; Leebens-Mack et al. 2005).

So far, five different coding genes, analysed alone or in combination, have been used to reconstruct relationships of early branching eudicots. The first genes to be analysed were *rbcL* (Chase et al. 1993) and *atpB* (Savolainen et al. 2000a). Their use recovered all lineages belonging to the "basal eudicots", but support for their inter-relationships was not evident. Nevertheless, terminal clades like Ranunculales, Proteales, or Buxaceae–Didymelaceae were identified, and both genes converged on the first-branching position of Ranunculales in eudicots. Hoot et al. (1999) and Soltis et al. (2000) added nuclear 18S sequences. Their analyses showed improved support for most terminal clades. Buxaceae–Didymelaceae and Trochodendraceae were depicted either as successive sisters to core eudicots or in a tritomy with the core eudicots. The clade including Buxaceae–Didymelaceae, Trochodendraceae, and core eudicots gained 87–88% JK support. The respective positions of Sabiaceae and Proteales were not resolved with confidence. Even adding nr26S sequences for a four-gene dataset (Kim et al. 2004) did not improve resolution in the basal eudicot grade. Phylogenetic analysis of a dataset comprising two thirds of the rapidly evolving *matK* gene (Hilu et al. 2003) provided a picture similar to that of the multi-gene analyses. Moreover, *matK* indicated that Buxaceae are sister to core eudicots (91% JK, 1.00 posterior probability (PP)) and provided moderate support (82% JK) for the first-branching position of Ranunculales in eudicots.

Recently, sequences of introns such as the group I intron in *trnL*, and the group II intron in *petD* were used to infer relationships among basal angiosperms (Borsch et al. 2003, 2005; Löhne and Borsch 2005). The same applies to the *trnT-L* and *trnL-F* spacers (Borsch et al. 2003) which, like the above-mentioned introns, are located in the large single-copy region of the chloroplast genome, and are rapidly evolving. It was shown that mutational dynamics in these spacers and introns follows complex patterns related to structural constraints. Extreme length variability in introns and spacers is confined to certain mutational hotspots which correspond to the least constrained stem-loop elements P6 and P8 in the secondary structure of the group I intron (Quandt et al. 2004), and to the least constrained terminal stem-loop elements of domains I, II, and IV in the group II intron (Löhne and Borsch 2005). Moreover, the *petD* intron dataset yielded one of the largest indel matrices so far generated for angiosperms. Reconstructing the evolution of the underlying microstructural mutations, involving one to many nucleotides, showed a large number of them to be synapomorphic for deep to terminal nodes. Thus, microstructural mutations in rapidly evolving spacers and introns can be expected to be of high phylogenetic utility (Kelchner 2000), as has been shown for indels supporting shallower nodes (Müller and Borsch 2005) as well as for indels in the conserved chloroplast-inverted repeat (Graham et al. 2000). In basal angiosperms it was evident that combining *trnT-F* and *petD* sequences with *matK*, which also is rapidly evolving and has provided good signal in an overall angiosperm analysis (Hilu et al. 2003), can lead to further improved resolution and support of phylogenetic trees (Borsch et al. 2005; Müller et al. 2006). Combining such datasets could therefore have the potential of providing further insight into some of the nodes that are notoriously difficult to resolve in the basal eudicot grade. In comparison to analyses of basal angiosperms, where gymnosperms had to be used as outgroup, a *petD* and *trnL-F* eudicot dataset with basal angiosperms as outgroups could be expected to entail lower *p*-distances, and thus to be easier to align. Because mutational dynamics is strongly influenced by structural constraints inherent to the respective genomic region, at least in introns, hotspots were to be expected in similar positions in eudicots as compared to basal angiosperms.

The aims of the present study were: (1) to produce an alignment of rapidly evolving group I and group II introns, and of spacers, for a taxon sampling