3. Material and methods

3.1 DNA extraction and purification
Voucher specimen information and GenBank accession numbers of plant material used are listed in Table 3.1. DNA was extracted from 0.01–0.3 g herbarium and silica-dried leaf tissue (Chase & Hills, 1991), using the 2X CTAB method of Doyle and Doyle (1987). In the precipitation of DNA, iso-propanol was added to herbarium leaf material and ethanol to silica-dried leaf tissue. Samples were stored at –20°C for at least three weeks (Fay et al., 1998), after which it was centrifuged for 5 min at 3000 rpm. QIAquick silica columns were used for purifying total DNA following the purification protocol provided by the manufacturer (QIAGEN Inc.).

3.2 Gene region amplification and purification
Polymerase chain reactions (PCR) were used to amplify the different gene regions in 50 µl reactions containing: 25 µl PCR Mastermix [50 units/ml Taq DNA Polymerase (pH 8.5), 400 µM each: dATP, dGTP, dCTP, dTTP and 3 mM MgCl$_2$ (Promega Corporation)] or GeneAmp Fast PCR Master Mix (2X) [AmpliTaq DNA polymerase, GeneAmp PCR Buffer, dNTP, MgCl$_2$, Stabilizers (Applied Biosystems Inc.)]; 1.0 µl bovine serum albumin (BSA) (0.004%; Savolainen et al., 1995); 0.5 µl of the forward as well as reverse primers (see Table 3.2 for the primer sequences and references for each region studied); 0.5–5.0 µl DNA template; 1.0 µl dimethyl sulfoxide (DMSO) for nuclear genes and Nuclease-free water (Promega Corporation). The target regions were amplified in a GeneAmp PCR System 9700 or in a 9800 Fast Thermal Cycler, following the protocols listed in Table 3.3. The PCR products were purified with QIAquick silica columns, using the purification protocol provided by the manufacturer (QIAGEN Inc.).

3.3 DNA sequencing
Cycle sequencing reactions consisted of: 1.0 µl Big Dye Terminator v. 3.1 (Applied Biosystems Inc.); 0.3 µl primer; 1.5 µl 5 x sequencing buffer; 0.5 µl DMSO for nuclear reactions; DNA template and Nuclease-free water (Promega Corporation) to make up a volume of 10 µl. Cycle sequencing was performed in a GeneAmp PCR System 9700 or in a 9800 Fast Thermal Cycler, using the following thermal profile: denaturation, 96°C (10 sec); annealing, 50°C (5 sec); extension, 60°C (4 min) for 26 cycles. Sodium acetate precipitation was used to purify the cycle sequencing product of excess dye terminator. The purified products were directly sequenced on an ABI PRISM 3130x/ Genetic Analyzer after adding 10 µl Hi-Di Formamide Genetic Analysis Grade (Applied Biosystems Inc.).
3. Material and methods

3.4 Data analysis

3.4.1 Choice of outgroups

Representatives of Sphaerosepalaceae (*Grielum humifusum, Rhopalocarpus*) and Neuradaceae (*Dialyceras coriaceum*) were selected as outgroups, due to their sister relationship to Thymelaeaceae (Fay et al., 1998; Bayer et al., 1999). According to Van der Bank et al. (2002), Thymelaeaceae is monophyletic with four subfamilies: Synandrodaphnoideae and Gonystyloideae successively sister to Aquilarioideae and Thymelaeoideae. Therefore representatives of Aquilarioideae, Gonystyloideae and Synandrodaphnoideae were also selected as outgroups for the same analyses as mentioned above. Due to alignment problems with ITS and *trnL-F*, representatives of Gonystyloideae (*Octolepis* spp.), and Thymelaeoideae (*Edgeworthia, Wikstroemia, Thymelaea* and *Daphne*) were selected as outgroups respectively. Voucher information and GenBank accession numbers of outgroups are listed in Table 3.4.

3.4.2 Maximum parsimony (MP)

The complimentary strands of the sequenced genes were assembled and edited in “Sequencher” version 3.1.2 (Gene Codes Corporation). Sequences were manually aligned and cladistic analyses performed using the parsimony algorithm of the software package PAUP version 4.0b10 (Swofford, 1998) on a Macintosh G4. Deletions of nucleotides did not contribute to the analysis as they were coded as missing data. The separate as well as the combined (ITS, *rbcL* and *trnL-F*) data matrices were analysed using a heuristic search with 1000 random sequence additions, keeping only 10 trees per replicate in order to reduce the time spent on branch swapping in each replicate. Tree bisection-reconnection (TBR) was performed as branch swapping algorithm with MULPARS (keeping multiple equally parsimonious trees) and all character transformations treated as equally likely (Fitch parsimony; Fitch, 1971). The trees collected in the 1000 replicates were then used as starting trees for another search without a tree limit. For the illustration of branch lengths throughout, DELTRAN (Delayed transformation) character optimisation was used instead of ACCTRAN (accelerated transformation) due to reported errors with the latter in PAUP 4.0b10. Internal support was assessed with 1000 bootstrap replicates (Felsenstein, 1985) with Fitch weights using TBR swapping and holding 10 trees per replicate. Only the clades with a frequency greater than 50% are reported.

Due to weak DNA, I was unable to amplify all of the regions thus the data matrices did not contain identical taxa sets (Table 3.5). The effects of these missing data on the support in the combined analyses as well as the patterns of relationship were investigated. Two combined analyses were performed: 1) all data present in all taxa, and 2) all taxa for which at least some data were present. The missing data did not have any obvious effect
on the second analysis. Through visual inspection of the individual bootstrap consensus trees, congruence of the separate datasets was assessed. Bootstrap trees were considered incongruent only when “hard” (i.e., with high bootstrap support) instead of “soft” (with low bootstrap support) incongruence was displayed (Seelanan et al., 1997; Wiens, 1998). No “congruence tests” such as ILD were used, due to their possible unreliability (Yoder et al., 2001; Reeves et al., 2001). The following scale for support percentages was used: 50–74%, low; 75–84%, moderate; 85–100%, strong.

3.4.3 Successive weighting (SW)
Successive approximations weighting (Farris, 1969) was used in the combined analysis to down-weight base positions that changed the tree topology (Chase et al., 2000). For SW the ‘reweight characters’ command based on the RI, using the maximum value (best fit) criterion and a base weight of one was used. The shortest Fitch trees were used as the basis for calculating the initial weights and the search–reweighting process was repeated until the same tree length was obtained twice in succession.

3.4.4 Bayesian analysis
A test was performed using Modeltest version 3.06 to determine which model of DNA substitution (out of 56 possible models specified in ‘modelblock’ for PAUP) should be used for the dataset at hand. Modeltest uses log likelihood scores to estimate which model of DNA evolution is more suited to a specific set of data (Posada & Crandall, 1998). In order to perform a modeltest, the combined PAUP matrix, containing one tree generated by the heuristic searches, was tested against the models of DNA evolution in modelblock for PAUP. The model selected was GTR+I+G (Akaike Information Criterion). A Bayesian analysis was performed using MRBAYES version 2.01 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). Settings for the selected model were nst = 6, rates = gamma, basefrequency = empirical, clock = unconstrained and number of generations = 1 000 000. Resulting trees were plotted against their likelihoods to determine the point where the likelihoods converged on a maximum value, and all the trees before the convergence were discarded as the ‘burn-in’ phase. All the remaining trees were imported into PAUP 4.0b10 and a majority rule consensus tree was produced showing the frequencies (i.e. posterior probabilities or PP) of all observed bi-partitions.