The petD group II intron as a species level marker: utility for tree inference and species identification in the diverse genus Campanula (Campanulaceae)

Authors: Thomas Borsch, Nadja Korotkova, Thomas Raus, Wolfram Lobin, and Cornelia Löhne

Source: Willdenowia, 39(1) : 7-33

Published By: Botanic Garden and Botanical Museum Berlin (BGBM)

URL: https://doi.org/10.3372/wi.39.39101
THOMAS BORSCH1*,2, NADJA KOROTKOVA3, THOMAS RAUS1, WOLFRAM LOBIN4 & CORNELIA LÖHNE1

The petD group II intron as a species level marker: utility for tree inference and species identification in the diverse genus Campanula (Campanulaceae)

Abstract

Chloroplast introns have a high potential as tools for phylogeny inference and DNA barcoding. This study examines the molecular evolution of the petD group II intron in Campanulaceae based on a sequence data set of 114 ingroup taxa. Three small mutational hotspots had to be excluded from phylogenetic analysis, the two most variable being located in the D4 loop (domain IV). A (GT)4-7 microsatellite in domain II is conserved at species level but of limited phylogenetic use due to unclear homology of individual repeat units. Sequences of the petD group II intron depict Cyphioideae, Lobelioideae and Campanuloideae as major Campanulaceae clades. Core Campanuloideae comprise two major radiations of Campanula species: a Musschia clade (including C. lactiflora) and a jacqion clade. Campanula is highly paraphyletic to a number of smaller genera such as Azorina, Michauxia and Edraianthus. The closed-tubular flowered taxa (Phyteuma and allies) are resolved sister to C. persicifolia. Within core campanuloids petD sequences identify 90 % of the taxon samples included in this study. Considering the ease of amplification and sequencing, and its high information content, the petD intron appears to be a good candidate in a two-tailed approach integrating molecular phylogenetics and species identification in the needed sampling of all core Campanuloideae species.

Additional key words: chloroplast genome, phylogenetic structure, molecular evolution, Eurasia, endemics, DNA barcoding

1. Introduction
One of the greatest challenges in plant biodiversity research is to generate robust phylogenetic frameworks for the many species-rich genera as a prerequisite for reliable inference of character evolution, historical biogeography and to examine modes of speciation in time and space. Reconstructing phylogenetic relationships in large and species-rich genera as well as analysing character evolution and biogeographic patterns in these genera require the inclusion of many taxa. Typically, the need for extensive taxon sampling limits the amount of sequence data that can be generated per individual taxon. As a consequence, efficient markers are needed that provide a maximum of phylogenetic signal per base sequenced.

In angiosperms, chloroplast spacers and introns and also the rapidly evolving matK gene were used mostly for species level phylogenetic inference (Borsch & Quandt 2009). The group I intron in trnL and the trnL-F spacer (together the “trnL-F region”) were proposed early on as markers that can be amplified and sequenced using universal primers (Taberlet & al. 1991) and are now the most frequently used non-coding plastid regions (Quandt & Stech 2003; Shaw & al. 2007). Other frequently used spacers are those between atpB and

1 Botanic Garten and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Straße 6-8, D-14195 Berlin, Germany; *e-mail: t.borsch@bgbm.org (author for correspondence).
2 Institute for Biology, Freie Universität Berlin, Altensteinstraße 6, D-14195 Berlin, Germany.
3 Nees Institute for Biodiversity of Plants, Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 170, D-53115 Bonn, Germany.
4 Bonn University Botanical Gardens, Meckenheimer Allee 171, D-53115 Bonn, Germany.
More recently, group II introns were discovered as further powerful tools (e.g., Kelchner 2002), with the two introns in *rpl16* and *rps16* currently being the most widely used. Considering that the greatest variability of plastid genomic regions between closely related taxa is found in the non-coding parts of the chloroplast genome’s large and small single copy regions (LSC and...
...more and more spacers and introns from that part are used in plant evolutionary studies every year. Small & al. (1998) started a series of studies comparing the variability of sequences from 3-taxon sets of closely related species across angiosperms and compared so-called PIC values (parsimony informative characters) for many spacers and introns. More recently, such comparisons of variability are also based on completely sequenced genomes (e.g., Timme & al. 2007).

On the other hand, it is obvious that there are considerable differences in phylogenetic utility (corresponding to differences in phylogenetic structure) of genomic regions. These are determined by specific mutational dynamics of a genomic region governing the accumulation of character state changes during the region’s evolutionary history. A comparison of sequence data sets from \( rbcL \), \( matK \) and the non-coding parts of the \( trnT-trnF \) region (Müller & al. 2006) showed that phylogenetic structure is not only determined by the number of potentially informative sites (i.e., the amount of information) but also by the distribution of mutations across the tree (i.e., quality of information). Sequences of \( trnT-trnF \) performed significantly better than \( matK \) in a basal angiosperm taxon set. Consequently, phylogenetic structure in sequence data sets cannot be easily deduced from PICs in a small sequence set (Timme & al. 2007), but is determined by molecular evolutionary patterns, which again are strongly influenced by structural and functional constraints in a genomic region. Searching highly variable markers with the best phylogenetic structure is therefore a worthwhile approach (Borsch & Quandt 2009).

In addition, there can of course be practical limitations, since not all genomic regions are easily amplified from PICs in a small sequence set (Timme & al. 2007) or frequently cause sequencing problems due to long strands of A’s or T’s (>10 nt, such as the \( psbA-trnH \) spacer; Devey & al. 2009).

However, resolution and internal statistical support of the recovered trees are not sufficient in many cases. The trend therefore is to combine more regions for increased quality of phylogenetic hypotheses. Most current studies include at least two regions in combined data sets, and multiple data sets combining four to five chloroplast regions are becoming standard (e.g., Barfuss & al. 2005; Kocyan & al. 2007; Löhne & al. 2007; Käreheid & al. 2008) to infer a robust plastid tree. There are many cases where the combination of many non-coding and rapidly evolving chloroplast regions led to considerably improved tree resolution and support within and among genera of various angiosperm lineages (e.g., Barfuss & al. 2002; Tesfaye & al. 2007; Löhne & al. 2007; Käreheid & al. 2008). Tree reconstruction in \( Campanulaceae \) and especially in \( Campanula \) L. has so far only yielded partly supported trees, including large polytomies (nrITS: Eddie & al. 2003; Roquet & al. 2008; \( trnL-F \): Roquet & al. 2008). Most recently combination of \( rbcL \), \( atpB \) and \( matK \) (approx. 4200 nt) yielded improved but still not fully resolved trees (Cellinese & al. 2009).

The \( petD \) gene is part of the \( petB \) operon consisting of five genes, \( psbB \), \( psbT \), \( psbH \), \( petB \) and \( petD \), in the LSC region of the chloroplast genome (Westhoff & Hermann 1988). The group II intron in \( petD \) is located in the upstream part of the gene, following an 8 nt 5’ exon. Like other group II introns it is composed of six domains with conserved helical elements that are arranged around a central wheel, whereas the distal parts of especially domains I and IV are more variable. Universal primers for amplifying the \( petD \) intron have been designed by Löhne & Borsch (2005). Since then these primers have been successfully applied to inferring deep level relationships in eudicots (Worberg & al. 2007), rosids (Worberg & al. 2009) and asterids (Salomo & al., unpubl. data) as well as in \( Nymphaeales \) (Löhne & al. 2007) and \( Malpighiales \) (Korotkova & al. 2009). Compared to other chloroplast regions (e.g., \( trnK-matK \), \( trnT-F \), \( rpl16 \), \( atpB-rbcL \)) the amplification of \( petD \) was always easy and yielded large amounts of PCR products. The alignment always was straightforward, with only small and well defined mutational hotspots. As a phylogenetic marker \( petD \) performed well in all the studies carried out so far. More recently, Käreheid & al. (2008) and Groeninkx & al. (2009) showed \( petD \) intron sequences in \( Rubiaceae \) to contain much better phylogenetic signal than all other previously used markers.

It was thus promising to ask for variability and performance of \( petD \) intron sequences in phylogenetic analyses of a species-rich lineage. The \( Campanulaceae \) (the bell flower family) and the genus \( Campanula \) (Fig. 1) are diverse in northern temperate and Mediterranean regions. About 150 of the 300–400 species of the genus \( Campanula \) occur in Europe (Fedorov & Kovanda 1976). More than 260 plus many infraspecific taxa occur in the Mediterranean area (Greuter & al. 1984), of which more than 80 % are endemic to that area. Hotspots of endemism are found in the E Mediterranean (Phitos 1964, 1965; Carlström 1986), the Balkan region (Kovačić 2004), the Caucasus (e.g., Gagnidze 2005) and Turkey (Damboldt 1978). Moreover, generic concepts in \( Campanulaceae \) vary considerably, with between 40 and 85 recognized genera (Kolakovskiy 1987, 1994; Eddie & al. 2003; Lammers 2007), many of the smaller ones being European or W Asian endemics. Phylogenies published so far agree on high paraphyly or polyphyly of the genus \( Campanula \), resulting in the need for additional dense taxon sampling. At the same time there are many taxonomically difficult groups of morphologically closely allied taxa with unclear status (see Eddie & al. 2003). In various lineages of \( Campanula \) there is high phenotypic plasticity and there are polyphyletic lineages (Podlech & Damboldt 1964; Kovanda 1970a, b, 1977). As a consequence, taxonomic treatments at the species level often differ considerably among different authors and for different countries, causing serious problems for
assessing distribution and conservation status of the species. A striking example is the *C. rotundifolia* group (Podlech 1965; Fedorov & Kovanda 1976). DNA sequence data may complement morphology and help to back up taxonomic decisions and conservation measures. With a perspective on the conservation of the palearctic flora, *Campanula* s.l. is therefore an important genus in need of considerable research in order to effectively plan conservation measures.

The *Campanulaceae* belong to the eudicot order *Asterales* and contain approximately 2300 species (Lammers 2007). Classification systems either treat the closely related *Lobeliaceae* as separate family or include a subfamily *Lobelioidae* into a broadly defined family *Campanulaceae* (e.g., Cronquist 1988; APG II 2003; Lammers 2007; followed here). In other classification systems, *Lobeliaceae* are considered as an own family close to *Campanulaceae* (Takhtajan 1997). Nevertheless, phylogenetic relationships within the broad *Campanulaceae* alliance were never tested using a dense taxon sampling from all putative major lineages of the bell flower family, but rather focused on *Campanulolideae* (Eddie & al. 2003; Roquet & al. 2008) or *Lobelioidae* (Antonelli 2008), using the respective other lineage as outgroup. In analyses of *Asterales* based on nuclear ITS sequences, Eddie & al. (2003) first provided evidence that *Campanula* is not monophyletic. Due to its high variability, the nrrTS region has been widely used for phylogeny inference at the species and genus levels. However, the backbone of ITS trees in *Campanulaceae* is poorly resolved and many nodes did not gain good statistical support. Pollen (Dunbar 1975 a,b) and seed (Belyaev 1984 a,b) characters are informative in *Campanulaceae* and were used among other morphological data in a phylogenetic analysis by Shulkina & al. (2003). Larger taxon sets were further generated of the trnL-F region (Roquet & al. 2008) and the rbcL+apetB+matK genes (Cellinese & al. 2009; hereafter referred to as 3-gene analysis or data set). However, taxon sampling differs considerably between the two last-named studies, and major parts of the topology are not receiving high statistical support, even in the 3-gene analysis. Thus, there is a need for sampling more taxa and more characters, ideally based on molecular markers chosen for high phylogenetic signal.

Aims of this study were (1) to generate a well sampled *petD* group II intron data set for the genus *Campanula* and the *Campanulaceae* and to infer phylogenetic hypotheses based on both substitutions and micro-structural mutations; (2) to calculate a secondary structure of the *petD* intron for *Campanula* and to evaluate mutational hotspots in an accurate structural context; (3) to evaluate the phylogenetic utility of *petD* sequences for reconstructing species level phylogenies and the utility of *petD* sequences for species identification in *Campanula* s.l.

### 2. Material and methods

#### Taxon sampling and plant material

Sequences of the *petD* intron and the *petB-petD* intergenic spacer were obtained for 114 taxa currently classified as *Campanulaceae* incl. *Lobeliaceae* and one representative of *Styliadiaceae* and *Pentaphragmataceae*, which are close relatives of *Campanulaceae* (Albach & al. 2001; Bremer & al. 2002; Lundberg & Bremer 2003). Plants for this project were collected, predominantly in Germany, Georgia, Italy and Slovakia. A further important source of material were the living collections and the herbarium of the Botanic Garden and Botanical Museum Berlin-Dahlem (B) and the living collection of the Bonn University Botanical Gardens. Four taxa were sampled as outgroups, one from *Asterales* (*Lactuca sativa* L.; *Asteraeae*) and three from more distantly related asterids (*Atropa belladonna* L. and *Nicotiana tabacum* L., *Solanaceae*; *Sphenoclea zeylanica* Gaertn., *Sphenocleaceae*). The sequence for *Sphenoclea* was generated in this study, whereas sequences for the other three taxa were available through completely sequenced chloroplast genomes. Vouchers have been deposited in the herbaria at Berlin (B) or Bonn (BONN). Detailed information on the sampled material is provided in Appendix 1.

#### DNA isolation, amplification and sequencing

Genomic DNA was isolated from silica dried leaf tissue or herbarium specimens using the modified CTAB method with triple extractions described by Borsch & al. (2003) or using the NucleoSpin Plant II Kit (Macherey-Nagel). The *petB-petD* spacer and the *petD* intron were co-amplified using the universal primers PpetB1411F and PpetD738R designed by Löhne & Borsch (2005). Amplification reactions contained 2 µl DNA template with a concentration of 10 ng/µl, 5 µl Taq buffer S (PeqLab [including 15 mM MgCl2], 2 µl of each primer (20 pm/µl), 10 µl dNTPs (each 1.25 mM), and 1.5 units of Taq DNA Polymerase (PeqLab). Ultrapure H2O was added to obtain a total volume of 50 µl. PCR amplifications were performed in a T3 Thermocycler (Biometra). Amplification conditions were as follows: an initial denaturation step of 1.5 min at 96 °C, followed by 0.5 min at 95 °C, 1 min at 50 °C, 1.5 min at 72 °C, for 34 cycles, and a final extension step of 20 min at 72 °C. The PCR products were stained with 100x Gelstar nucleic acid stain and electrophoresed on a 2% agarose gel at 80 V for 2 hours. Fragments were excised from the gel and DNA was extracted and purified using the Qiaquick Gel Extraction Kit (QIAGEN) or the Gel/PCR DNA Fragment Extraction Kit (Avegene). The purified products were directly sequenced using the DCTS Quick Start Kit (Beckman Coulter) and run on an automated capillary sequencer (CEQ 8000 Genetic Analysis System, Beckman Coulter) or sequenced via Macrogen Inc. (Seoul, South Korea). Pherograms were
edited manually with PhyDe992 and later versions (Müller & al. 2005+).

Sequence alignment, coding of length mutational events — Sequences were aligned manually using PhyDe0992 and later versions (Müller & al. 2005+). Rules for the alignment of non-coding regions as outlined by Borsch & al. (2003) and Löhne & Borsch (2005) are followed here. Regions of uncertain homology (hotspots sensu Borsch & al. 2003) were excluded from the analysis. Indels were coded according to the Simple Indel Coding method (Simmons & Ochoterena 2000) using SeqState 1.40 (Müller 2005a).

Phylogenetic analyses — A parsimony ratchet was performed using PRAP (Müller 2004a). Ratchet settings were 200 iterations with 25 % of the positions randomly upweighted (weight = 2) during each replicate and 10 random addition cycles. The number of steps for each tree and the consistency, retention and rescaled consistency indices (CI, RI and RC) were calculated using PAUP*+ v. 4.0b10 (Swofford 1998). Support for the nodes found by the parsimony ratchet was inferred using jackknifing with the optimal jackknife parameters described by Müller (2005b). A total number of 10 000 jackknife replicates was performed using the TBR branch swapping algorithm with 36.788 % of characters deleted in each replicate. One tree was held during each replicate.

Bayesian Inference (BI) was carried out using MrBayes 3.1 (Huelsenbeck & Ronquist 2001). Nucleotide substitution models for the data set were evaluated using Modeltest 3.7 (Posada & Crandall 1998) with spacer and intron sequences analysed separately. The hierarchical likelihood ratio test (hLRT) suggested the GTR+I+1 model. Four simultaneous runs of Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) analyses, each with four parallel chains, were performed for 1 million generations, saving one tree every 100th generation, starting with a random tree. Other MCMC parameters were left with the program’s default settings. Burn-in was set at generation 50 000. The remaining trees were summarised in a majority rule consensus tree.

Inference of RNA secondary structure — The sequence of Campanula trachelium L. was chosen to represent a typical petD intron sequence of Campanula without large indels or other deviant features. Due to the large size of the intron, current available algorithms for RNA folding are not able to fold the entire intron sequence in one step, since too many alternative foldings are possible (Mathews & al. 2006). To ensure that the adequate secondary structure was inferred, structural partitions of the intron, i.e., domains, subdomains, were first identified based on the annotated petD intron alignment of Michel & al. (1989). Then, each of the domains was folded separately following Kelchner (2002). Additionally, constraints for the two exon binding sites and the branch-point “A” were defined to leave these sequence parts single-stranded. Foldings were done using RNAstructure 4.5 (Mathews & al. 1999+) that is based on the algorithm of Mathews & al. (2004). The intron structure was then drawn using RnaViz 2.0 (Rijk & al. 2003).

Calculation of substitutional rates — Relative rate tests were performed using GRate version 1.0 (Müller 2004b). This software uses the DNA substitution models implemented in phylogenetic analyses and allows to compare rates between pre-defined groups of taxa. All major clades of Campanulaceae s.l. found with parsimony and Bayesian tree inference in this study were defined as groups for the comparison of molecular rates. Atropa bella-donna was set as outgroup and Nicotiana tabacum as reference taxon.

3. Results

The petD intron and the upstream petB-petD spacer were easily amplified and sequenced for DNAs isolated from both silica gel dried and herbarium material. Sequencing was not hindered by long stretches of As/Ts (chloroplast microsatellites) and single reads were covering the whole region when using an ABI3100 system. Therefore, the use of both amplification primers for sequencing generated a double coverage of the marker.

The petD intron ranged from 654-772 nt in Campanulaceae and the petB-petD spacer from 172-231 nt. The secondary structure of the petD intron calculated for Campanula trachelium is given in Fig. 2. Examples of the secondary structure of its domain II, representing different sizes of the [GT]_n microsatellite, are shown in Fig. 3.

In the overall alignment three small hotspots (HS1-HS3) had to be excluded (Fig. 2, Appendix 3) from the phylogenetic analysis. The alignment can be downloaded (at www.eudicots.de) and the EMBL/GenBank accession numbers of the sequences are given in Appendix 1. The matrix comprised 1362 characters in total, of which 182 were variable but not informative and 372 were variable and parsimony informative. Maximum Parsimony reconstruction and Bayesian Inference yielded strict consensus and majority rule trees for Campanulaceae and Campanula (Fig. 4, 5) that were largely resolved and well supported except in the C. latifolia clade. Parsimony analysis based on substitutions only resulted in 350 shortest trees (1303 steps) with a CI = 0.629, a RI = 0.869 and a RC = 0.546. A list of postulated microstructural mutations is provided in Appendix 2. Simple sequence repeats accounted for most of these mutations, whereas inversions were not found. Addition of a matrix of 192 indels raised the number of variable but not informative characters to 299 and of variable and parsimony-informative characters to 437. From the combined matrix 229 trees of a length of 1525
steps and a CI = 0.656, a RI = 0.868 and a RC = 0.570 were recovered. The strict consensus is shown in Fig. 4. Bayesian Inference sampled 9500 trees from four chains of four independent runs employing a GTR+Γ+F model. The majority rule tree is shown in Fig. 5 as a phylogram.

The average relative substitutional rates in the petD intron data set calculated for all major Campanulaceae groups in the present study are given in Fig. 6.

4. Discussion
RNA secondary structure of the petD intron in Campanula

As an example for the core Campanuloideae, the secondary structure of the petD intron in Campanula trachelium was calculated (Fig. 2). The sequence of C. trachelium showed comparatively few microstructural mutations and was therefore considered to be a representative example for the study group. The overall structure reflects the typical consensus structure of group II B introns with six domains arranged around a central wheel (Michel & al. 1989; Kelchner 2002). Domain II contains the highlighted branch-point “A”. EBS = “Exon Binding Site” and IBS = “Intron Binding Site”.

Fig. 2. Secondary structure of the petD group II intron from Campanula trachelium. – Nomenclature of structural elements follows Michel & al. (1989). Roman numerals I to VI highlight domains. Subdomains of domain I are labelled with letters A-D(2). Lower case roman numerals i to iii mark conserved helical elements. Domain IV contains the highlighted branch-point “A”. EBS = “Exon Binding Site” and IBS = “Intron Binding Site”.

ΔG = -137,7
and eudicots the complete domain II and the terminal stem-loop of domain IV deviate most, whereas domains V and VI are largely conserved.

**Patterns of variability in the petD group II intron**

The petD intron and the upstream petB-petD spacer yielded a reliable alignment with the majority of microstructural mutations being simple sequence repeats (SSRs; Appendix 2) that allowed unambiguous motif recognition. The sequences analysed in this study cover a much smaller spectrum of lineage diversity than all petD data sets previously published in angiosperms (Löhne & Borsch 2005; Worberg & al. 2007; Korotkova & al. 2009). Even though the distances between the sequences are much smaller, three mutational hotspots of unclear sequence homology across all Campanulaceae and the outgroups had to be excluded from the analysis (Fig. 2, Appendix 3). Compared to Löhne & Borsch (2005), who used a consensus model for group II introns (Michel & al. 1989) to annotate domain borders in the alignment and to map hotspots onto the secondary structure, this study allows a more precise examination of domain structure and of the exact location of mutational hotspots within the domains.

Compared to other rapidly evolving spacer and intron data sets the percentage of nucleotides in hotspots is extremely low (Appendix 3). On average hotspots comprise less than 4% of nucleotides in the petD intron of Campanulaceae. One of the reasons is a distally scarcely extended subdomain D2 of domain I. In all petD intron secondary structures known so far this subdomain is usually AT-rich and length variable, as in most other asterids (Salomo & al., unpubl. data) and in Malpighiales (Korotkova & al. 2009).

Hotspot HS1 is a dinucleotide GT microsatellite (Fig. 3). It is also present in other asterids (Salomo & al., unpubl. data) and is constituted by three to six repeat units in Campanulaceae. Although homoplastic at the level of the family, repeat number is conserved within major clades. No variation within species or among closely allied taxa (e.g., *Campanula rotundifolia* L. and its allies) is observed. The *C. latifolia* and *Azorina* clades have a copy number of four, the *C. rotundifolia* clade of five, the *Phyteuma* clade also of five but raising to six in *Phyteuma* s.str. All other Campanuloideae also have five repeat units with the exception of *Wahlenbergia* (three). The conserved nature of this microsatellite may be caused by its position in a stem element of domain II where it is stabilised through complementary base-pairing. An incrementation of the repeat-unit copy number extends the length of the proximal helix in domain II and leads to a considerably altered distal part of...
this domain in *C. trachelium* (one of the taxa with four repeat units; Fig. 3). The structure shown is thermodynamically suboptimal with a ΔG value of -11.3, whereas the optimal folding (ΔG = -13.0) forms a second helix departing from the intron’s central wheel (not in line with a group II intron consensus structure). High mutational activity in domain II appears to be present in core eudicots, as evidenced by this study and observations by Korotkova & al. (2009) and Salomo & al. (unpubl. data), but not in early branching angiosperms (Löhne & Borsch 2005). Domain II is the least important for correct intron splicing (Lehmann & Schmidt 2003), so that the mutational hotspot again correlates with an area of minimal evolutionary constraints.
The other two hotspots are located in the D4 loop of domain IV. In Campanulaceae this loop is the largest and most variable terminal stem-loop. High D4 loop variability appears to be a general pattern of plant chloroplast introns (Kelchner 2002; Löhne & Borsch 2005; Watts & al. 2008; Korotkova & al. 2009), as evolutionary constraints are low because of the degradation of the maturase open reading frames in all introns except trnK (Toor & al. 2001).

Hotspot HS 2 is mostly generated by insertion/deletion events that lead to different terminal loops (Fig. 2). However, within core Campanuloidae microstructural mutations are rather rare (a 4 nt deletion in Symphyandra hofmannii Pant. and a “AAAGAA” SSR in C. rumeliana (Hampe) Vatke) so that the respective sequence parts could be utilized in phylogenetic analyses limited to that clade. The same regards to HS 3 that consists of three nucleotides in Campanuloidae (Fig. 2), whereas it is length variable in Cyphioideae, Lobelioideae and the other Asterales.

Caused by their core structure, group II introns exhibit mosaic patterns of sequence conservation (Kelchner 2002; Löhne & Borsch 2005). Helical elements tend to be conserved, whereas stem-loop elements are usually less constrained and may be expected to contain more variable sites as well as insertions, deletions and inversions. In line with the petD intron consensus structure in asterids (Salomo & al., unpubl. data) there are several conserved helical elements labelled i to iii throughout all intron domains in Campanulaceae (Fig. 2). In particular, the structure of the large domain I is upheld by these helical elements. For microstructural mutations and, as a consequence, for our alignment this results in an evolutionary constrained core structure that still exhibits many variable sites.

The petD intron is a rather small group II intron in the chloroplast genome, especially compared to the on average about 25 % bigger rpl16 intron (Borsch & Quandt 2009). Length differences among different intron sequences were primarily attributed to smaller

---

**Fig. 4A-B.** Maximum Parsimony tree of Campanulaceae (strict consensus of 229 shortest trees of 1525 steps) based on the petD data set including indels. – Jackknife values supporting individual nodes are shown above branches.
Borsch & al.: The petD group II intron: tree inference and species identification in *Campanula*

![Species tree diagram](https://bioone.org/journals/Wildenowia on 04 Jul 2019)

**Campanula latifolia clade**

**Azorina clade**

**Campanula rotundifolia clade**

**Plytsea clade**

**Muschia clade**

**Jestone clade**

**Wahlenbergia clade**

**Platyton clad**

**lobelids**

**cyprioids**

**OUTGROUP**

---

*Note: The diagram represents the phylogenetic relationships among species within the *Campanula* genus, highlighting the petD group II intron's role in inferring species trees and identifying species within the genus.*
stem-loop elements in domains I (especially subdomain D2) and IV. In a comparison of intron variability in identical taxon sets in Nymphaeales, Löhne & al. (2007) found the petD intron to be the least variable, also yielding lower phylogenetic structure R in comparison to the introns in rpl16 and trnK and also to trnT-trnF (comprising two spacers and the group I intron in trnL). It is therefore rather surprising that this study finds high phylogenetic structure of the petD intron at species level in Campanula (see also below). If the group II intron secondary structure is a mosaic of more conserved helical elements and more variable loops and bulges, one would expect that point mutations are more frequent in the less constrained loops and bulges when less distant sequences are compared (e.g., those spanning different species and closely related genera). The petD intron in Campanulaceae will provide a good model to determine structural elements (stems, loops, bulges) based on precisely calculated secondary structures, and to evaluate the distribution of mutations in relation to the structure.

Circumscription of Campanulaceae and phylogenetic relationships within the Campanulaceae s.l.

Within Asterales the families Pentaphragmataceae, Styliidaeae and Rousseaeae have been inferred to be the closest relatives of the Campanulaceae clade (including Lobeliaceae). Campanulaceae and Styliidaeae were sisters in the combined analysis of rbcL-atpB+18S+ndhF (Albach & al. 2001) but with only weak support (61 % JK). In a subsequent analysis using six chloroplast regions, Bremer & al. (2002) raised the confidence into this node to 82 % JK, whereas Pentaphragmataceae were hypothesised as sister to a Campanulaceae-Styliidaeae clade. Lundberg & Bremer (2003) found the same, adding a morphological matrix, but again with only weak support. Rousseaeae on the other hand were shown by Soltis & al. (2000) to be the sister group to Campanulaceae, but appear to be more distant in other broad-scale analyses (Savolainen & al. 2000; Albach & al. 2001; Bremer & al. 2002). Sampling a greater number of genomic regions, Winkworth & al. (2008) now provide increased evidence for Campanulaceae being sister to Rousseaeae, whereas Pentaphragma is sister to the remainder of Asterales. Sequence data of petD (this study) also depict Pentaphragma as distant from Campanulaceae.

The pre-cladistic assumption of close relations between Sphenoclea Gaertn. and Campanulaceae (e.g., Cronquist 1988) are refuted by all molecular phylogenetic analyses. In our petD trees that were rooted with Solanaceae (Atropa L., Nicotiana L.), Sphenoclea is shown as sister to the Asterales (Fig. 4, 5). Although our sampling beyond Campanulaceae is restricted and does not permit further insights into the exact position of Sphenoclea, it provides clear arguments against any close association with Campanulaceae.

This phylogenetic analysis of the bluebell family has a representative taxon sampling that includes Cyphioideae together with Lobelioidae and a broad sampling of Campanuloideae. Previous phylogenetic analyses either rooted Campanulaceae with Lobeliaceae (e.g., Eddie & al. 2003), just focused on Campanula and putative close relatives using Jasione L., Roella L. and Wahlenbergia Roth as outgroups (Roquet & al. 2008), or sampled Lobelioidae and very few Campanuloideae with other Asterales as outgroups (e.g., Antonelli 2008). Using molecular (atpB, ndhF, rbcL) and phenotypic characters to analyse phylogenetic relationships of Asterales, Lundberg & Bremer (2003) included one sample of Campanula and Codonopsis Wall. to represent Campanuloideae, and Cyphia Roger-see S. Moore, along with three genera of the Lobelioidae. Whereas Cyphia P. J. Bergius appeared either unresolved or as sister to Campanula plus Codonopsis, the monophyly of Campanulaceae s.l. was clearly established. Our Maximum Parsimony reconstruction of petD sequence data finds Cyphioideae as weakly supported sister to Campanulaceae (Fig. 4), whereas Bayesian Inference is inconclusive (Fig. 5). The other two small subfamilies Cyphocarpoideae and Nemacla-idoideae (Lammers 2007) have not been included so far in molecular analyses. APG II (2003) does not exclude the possibility of keeping Campanulaceae and Lobeliaceae as separate taxa, but it appears that the problem has not yet been thoroughly investigated. The position of the other three subfamilies in the Campanulaceae clade needs to be clarified before any decision on internal classification can be made.

Phylogenetic relationships within the core Campanuloideae and Campanula

Core Campanuloideae — Our study depicts a well supported clade of core Campanuloideae constituted by four major lineages (Fig. 4, 5). However, the relationships of the Jasione and Musschia clades with respect to the two other large Campanula clades are not yet clear. The Canarina and Wahlenbergia clades appear as successive sisters to core Campanuloideae with maximum support. The ITS trees of Eddie & al. (2003) and Roquet & al. (2008), which comprised a more representative sampling of Campanuloideae than previously available phylogenetic analyses, did not resolve the backbone under Maximum Parsimony. Nevertheless, Eddie & al. (2003) provided evidence for a so-called Raphunculus clade (corresponding to our well supported sister group of the Campanula rotundifolia + Phyteuma clades) with the genera Githopsis Nutt. and

Fig. 5. Bayesian phylogram of Campanulaceae (majority rule tree found with four independent runs) based on the petD data set including indels. – Posterior probabilities are given above branches.
Heterocodon Nutt. as weakly supported sister group. What we call the Campanula latifolia clade (Fig. 4, 5) is contained for the most part in their “Campanulaceae s.str. clade”. The Jasion and Musschia clades, including Wahlenbergia, appeared unresolved as “transitional taxa”. Bayesian Inference by Roquet (2008) then improved the picture for the Campanula s.str. and Rapunculus clades. However, in a combined ITS and trnL-F tree the latter authors showed an incongruent topology, with Wahlenbergia branching from inside the core campanuloids (Roquet & al. 2008). Since the authors also used outgroups that at least in the case of the Jasion clade belong to the core Campanuloideae, the respective results need to be viewed with care. To summarise, both the Jasion and the Musschia clades have to be included along with representatives of the two main Campanula clades into the core Campanuloideae for further testing the phylogeny of major clades of Campanula.

The two major campanuloid clades are the Campanula latifolia + Azorina clade (comprising the nomenclatural type of Campanula) and the C. rotundifolia + Phyteuma clade. In this study, we will not further discuss the subgeneric and sectional classification of Campanula as this has been done by Eddie & al. (2003) and Shulkina & al. (2003), who also provide detailed overviews of the classification history of Campanula and so far existing morphological, anatomical and cytological studies.

Campanula lactiflora M. Bieb. is sister to Musschia Dumort. with high confidence (Fig. 4, 5), corroborating the ITS tree of Eddie & al. (2003). Shulkina (1979) placed this species into its own genus Gadellia Schulkina, endemic to the Caucasus. However, other close relatives (C. primulifolia Brot., C. peregrina L.) could not be sampled here. Segregation of these species from Campanula could be justified only if Musschia is upheld as a separate genus, and this again requires clarification of the relationships of the four major lineages in core Campanuloideae.

Campanula latifolia clade — Campanula rumeliana (Hampe) Vatke (≡ C. jacquinii subsp. rumeliana (Hampe) Kit Tan) is resolved as isolated sister lineage (94 % JK, 1.0 PP) to all other members of this clade (90 % JK, 1.0 PP; Fig. 4, 5). Our topology is congruent with the nrITS tree of Roquet & al. (2008) and the 3-gene analysis of Cellinese & al. (2009) but petD yields better node support (nrITS: no statistical confidence; rbcL+atpB+matK: 80 % JK, 1.0 PP and 81 % JK, 1.0 PP, respectively). C. rumeliana has a capitate inflorescence with very narrow-infundibuliform corolla (less than 2 mm broad) and was therefore described and traditionally filed under Trachelium L. (Tan & Iatrour 2001) or under its segregate Diosphaera Buser (Hayek 1928-31). It is endemic to S Bulgaria and NE Greece. Other molecular studies found Trachelium caeruleum L. (the nomenclatural type of Trachelium) as sister to the C. latifolia + Azorina clade (Roquet & al. 2008; Cellinese & al. 2009), underscoring the distinctness of Diosphaera from Trachelium. Thus, Trachelium, as treated for example in Flora Europaea (Tutin 1976), is not a natural group. When restricted to T. caeruleum (as in Greuter & al. 1984), it is a W Mediterranean genus.

Within the Campanula latifolia clade several groups of species are resolved, but their interrelationships still remain unclear. In the following we will discuss such lineages that can be phylogenetically defined. These lineages either comprise species that are also morphologically allied or species from certain geographical regions, which can be considered to represent smaller radiations within Campanula.

The latter is true for the largest subclade (70 % JK, 1.0 PP, Fig. 4, 5), which, alongside with the widespread Campanula bononiensis L. and C. rapunculoides L. (see below), is mostly made up of Caucasian species (C. alliariifolia Willd. [extending to Anatolia], C. armazica Kharadze, C. ciliata Stev., C. daliricia Kharadze, C. federovii Kharadze, C. grossheimii Kharadze, C. hohenackeri Fisch. & al., C. hypopolia Trautt., C. kolenatiana C. A. Mey ex Rupr., C. makaschvili E. Busch and C. petrophila Rupr., C. saxifraga M. Bieb.), indicating a secondary radiation of Campanula in that region. Relationships between C. alliariifolia, C. grossheimii and C. hohenackeri were also depicted in the ITS trees of Eddie & al. (2003) and Roquet & al. (2008), in the latter gaining higher support with a different, improved taxon sampling. A second, smaller lineage of predominantly Caucasian species is composed by C. annae Kolak., C. collina Sims, C. sarmatica Ker Gawl. and the amphi-Adriatic C. foliosa Ten. (Fig. 4, 5).

Campanula latifolia L. and C. trachelium are resolved in a clade, in the Bayesian tree of our study (Fig 5; 0.8 PP), that has not been found by other molecular phylogenetic analyses. Eddie & al. (2003) point out that these belong to a morphologically distinctive group of Eurasian mesophytic species with elongate, spicate inflorescences and lacking appendages, also including C. bononiensis and C. rapunculoides. However, the latter species belong to the aforementioned mostly Caucasian clade in our study, and also in the ITS tree of Roquet & al. (2008), as far as C. rapunculoides is concerned (C. bononiensis was not sampled; see below for the surprisingly different position of C. trachelium in the trnL-F tree of Roquet & al. 2008). C. bononiensis and C. rapunculoides are species with a European to W Siberian distribution that also occur in the Caucasus (Podlech 2008; Kolakovsky 1992).

Edraianthus (A. DC.) DC. clearly is nested within the Campanula latifolia clade in our trees, corroborating earlier findings (Eddie & al. 2003; Shulkina & al. 2003; Roquet & al. 2008). Phylogenetic studies thus agree on its placement within Campanuloideae, as sug-
gested by Fedorov (1957) and Kolkovsky (1987, 1994) and not close to Wahlenbergia where it was placed historically. In a detailed analysis of this Balkan centred group of 10-28 species, Stefanović et al. (2008) showed the monophyly of Edraianthus except the Caucasian species E. overinianus Rupr. (= Muehlbergella overiniana (Rupr.) Feer; also accepted as genetically distinct by Lammers 2007) and the N Greek E. parnassicus (Boiss. & Spruner) Halácsy (= Halactyrella parnassica (Boiss. & Spruner) Janch.), using trnL-F and atpB-rbcL spacer sequences. The apical capsule dehiscence that was considered characteristic for Edraianthus (e.g., by Lakušić 1974) thus is a derived feature of several species, whereas a clear phenotypic synapomorphy for the emended monophyletic treatment of the genus is still not known (Stefanović & al. 2008). Hypotheses about the next relatives of Edraianthus are, however, difficult. Our study depicts Campanula bartata as sister lineage albeit with low support (0.7 PP). Roquet et al. (2008) show Edraianthus in a well supported clade with C. medium, C. spicata, C. incurva A. DC. and C. hofmannii (Pant.) Greuter & Burdet (= Symphyandra hofmannii Pant.). The ITS trees in Eddie & al. (2003) are inconsistent, with C. latifolia and Edraianthus as sisters (no statistical support). Bayesian reconstruction on the 3-gene data set (Cellinese & al. 2009) shows E. graminifolius A. DC. in a grade of C. medium and followed by a lineage of Symphyandra armena A. DC. and C. saxatilis L. albeit again without statistical confidence. Stefanović & al. (2008) show C. betulifolia C. Koch, C. radicosa Bory & Chaub., C. seraglio Kit Tan & Sorger and C. tymphaea Hausskn. as close relatives, but the latter three have not been included in any other phylogenetic study so far. Since none of the so far published analyses includes all of the aforementioned species, a sensible comparison of phylogenetic hypotheses derived from the different data sets is difficult and may have to await a more complete species sampling of Campanula s.l. A group with Campanula spicata, C. thrysoides L., and C. tridentata Schreb. is well resolved with petD but C. spicata has a different, incongruent position in the ITS trees, appearing in the same subclade as C. medium and Edraianthus (Roquet & al. 2008). The trnL-F tree depicts C. spicata in yet another place, as sister to C. allia-risifolia and C. trachelium. On the other hand, C. spica-ta, C. thrysoides and C. tridentata share morphological characters such as long, more or less compact inflorescences with upright flowers and acuminate calyx teeth, supporting a hypothesis of their close relationship. Campanula cervicaria L. and C. glomerata L. are found to be closely related. C. glomerata, a taxonomically complex species with different subspecies in various parts of Europe (Fedorov & Kovanda 1976), and C. cervicaria also show morphological similarities with the previous lineage. Much additional work will be necessary to clarify species limits within this group and to test its relationships to C. sicca and its allies. The lineage of Campanula carpatha Halácsy and C. tomentosa Vent. is also newly found in this petD study. C. tomentosa is a chasmophytic species with quinelocular ovary, endemic to the calcareous massif of Samsun Dağ in coastal W Anatolia, between the ancient cities of Ephesus and Priene. C. carpatha, endemic to the SE Aegean islands of Karpathos and Saria, belongs to a group of biennial or short-lived perennial, five-carpelled chasmophytes confined to the S Aegean island arc (Phitos 1965). Its nearest relatives are C. pelvi-formis Lam. and C. tubulosa (Boiss.) Engl. All three are endemics of the Cretan area and are considered to form a radiation of neo-endemics (Cellinese & al. 2009). Five other E Mediterranean species belong to the Campanula latifolia clade (Fig. 4, 5, see bottom part of uppermost clade), but still with unclear relationships. (1) C. incurva A. DC. is a perennial, large-flowered endemic of E Central Greece. Among the three-carpelled species of Campanula with appendiculate calyx sinuses it is unique by its cordate leaf blades and is excluded by Phitos (1964) from the Greek tricarpellary species of C. sect. Medium studied by him. (2) C. sartorii Boiss. & Heldr. is endemic to the island of Andros (W Kiklades). It is three-carpelled, with an exappendiculate calyx, which is the normal condition in Campanula. In habit and leaf shape it recalls some amphi-Adriatic members of the C. elatines aggregate (Greuter & al. 1984), from which it was, however, excluded by Damboldt (1965) due to differences in seed morphology. The studied members of the latter group belong in the C. rotundifolia clade (see below). (3) C. thessala Maire belongs to the Greek tricarpellary species of C. sect. Medium studied by Phitos (1964). Greuter & al. (1984) place it in the C. rupestris aggregate, a group of 20 closely allied, predominantly five-carpelled taxa of facultative chasmophytes with petiolar, lyrate to lobate rosette leaves and appendiculate calyx sinuses, confined to Greece and the Aegean islands, revised by Phitos (1964, 1965). (4) C. hierapetrae Rech. f., a local endemic of Mt Afendis Kavousi in E Crete, belongs to the C. heterophylla aggregate (Greuter & al. 1984): four rare and local Aegean island endemics, to which the recently described C. koyuncu H. Duman of Baba Dağ in SW Anatolia is to be added (Fielding & al. 2005). Cellinese & al. (2009) found Michauxia tchihatchewii Fisch. & Heldr. to be sister to C. hierapetrae. (5) In our petD tree, M. campanuloides L’Hér. is also unresolved in the C. latifolia clade. Michauxia L’Hér. is a genus of seven species growing in SW Asia, predominantly Turkey (Damboldt 1978; Lammers 2007). Its main characteristic, the conspicuous stigma with 8-10 recurved lobes, appears to be a derived feature, and likely relationships to E Mediterranean Campanula taxa have to be further tested. Surprisingly, ITS data resolve M. tchihatchewii sister to C. barbata (Alps and Pyrenees,
Fig. 1A), with variable support (Eddie & al. 2003, 98 % BS; Roquet & al. 2008, 0.79 PP). Future studies will have to evaluate whether there might be incongruence between plastid and nuclear DNA trees that would point to ancient reticulations.

**Azorina clade** — Our study shows a well defined Azorina clade, indicating close relationships of the Azorean endemic genus Azorina Feer, Cape Verdean Campanula species (C. bravensis Bolle, C. jacobaea C. Sm.; both included for the first time in a molecular phylogenetic study) and an E African species (C. edulis Forssk.). This clade was anticipated by Eddie & al. (2003), because all these species belong to a group with basal capsule dehiscence, appendages between the calyx lobes (e.g., Thulin 1976), and a total area ranging from the mid Atlantic archipelagos via the E Mediterranean to E Africa. The clade is not readily comparable with results from Roquet & al. (2008) due to a strongly deviating taxon sampling, but their ITS and trnL-F trees show the Ibero-Maghrebine C. mollis L. (a relative of C. edulis) as a close relative of Azorina.

**Campanula rotundifolia clade** — This well supported clade as understood in our study (Fig. 4, 5) includes most species with bell-shaped flowers and is sister to the Phyteuma clade, which contains many species with a corolla deeply split into narrow segments or with a closed tubular corolla (see below). The C. rotundifolia aggregate (Greuter & al. 1984; Fedorov & Kovala 1976; Fig. 1C) forms a well supported clade including species such as C. rotundifolia (widespread), C. scheuchzeri Vill. (Pyrenees, Alps, W Carpathians, Balkans) and C. baumgartenii J. Becker (W Central Europe; Buttler 2002). Relationships within this clade are difficult to resolve, since it contains many polyploids (Kovala 1970a, b, 1977) and species limits are unclear in many cases. Further work will thus have to evaluate possible reticulate patterns and incomplete lineage sorting, using a geographically dense sampling. The dwarf mountain species, represented by C. cespitosa Scop. and C. cochlearifolia Lam., appear as close but independent relatives (Fig. 4, 5) of the phylogenetically defined C. rotundifolia alliance. ITS data did not provide resolution for these species (Park & al. 2006). The North American C. divaricata Michx. is a third member of this group, corroborating earlier findings of Eddie & al. (2003) based on ITS and indicating a relatively recent migration to North America from Eurasia. The C. elatines aggregate (Greuter & al. 1984), comprising the isophyllous species studied by Damboldt (1965), is polyphylectic in the petD trees of our study, in agreement with the nuclear ITS trees (Park & al. 2006): C. elatinoideas Moretti, C. fragilis Cyr. and C. isophylla Moretti are resolved as relatives to the C. rotundifolia alliance, whereas a second clade with C. elatines L. and C. portenschlagiana is more distant. The latter species was found within the highly supported C. garganica clade by Park & al. (2006).

A novel result of our study is the well supported sister-group relationship of Campanula patula L. and C. spatulata Sm. Previous phylogenetic analyses never included both species together. C. patula, predominantly biennial but with a perennatial subspecies to the south (Carpathians and N Balkans), is widespread in European woodlands, whereas C. spatulata, a perennial geophyte with a napiform taproot, is endemic to the S Balkan Peninsula (two altitudinally vicarious subspecies) and Crete (a third subspecies). The surface parts of these two vicarious species are similar. Cano-Maqueda & al. (2008) found C. patula and C. rapunculus to be sisters, with the two in turn being the sister group to the W Mediterranean monophyletic C. lusitanica alliance. The whole clade (including C. patula and C. rapunculus) was called C. lusitanica lineage by Cano-Maqueda & al. (2008), distant from the C. dieckii lineage that is geographically restricted to the Baetic range. More extensive taxon sampling is needed to assess the relationship of the E Mediterranean C. spatulata to the C. lusitanica lineage, suggested by the position of C. patula in our study (Fig. 4, 5). Roquet & al. (2007) also found that W Mediterranean taxa of C. sect. Rapunculus often have a sister-group relationship with E Mediterranean and W Asian taxa.

A clade containing Campanula carpathica Jacq., C. pyramidalis L. and C. versicolor is of note. All have large flowers and the latter two are sister species and share very open flowers (Fig. 1D). C. versicolor is a highly variable, sturdy perennial with a woody rootstock, mostly growing as a chasmophyte from NE Italy to the Balkan Peninsula and the Ionian Islands. C. carpathica (Carpathian mountains) and C. pyramidalis (N Italy and NW Balkan Peninsula) grow in similar habitats. The lineage thus appears to reflect a SW European to E Mediterranean species radiation. C. carpathica and C. pyramidalis had been found without support as sister groups with ITS (Eddie & al. 2003; Roquet & al. 2008), and with high support in the 3-gene analysis (Cellinese & al. 2009). C. versicolor had not been previously studied, but was considered as a member of the C. pyramidalis aggregate by, e.g., Greuter & al. (1984). A putative fossil relative of C. pyramidalis and C. carpathica, called C. palaeopyramidalis (Łańcucka-Środoniowa 1977, 1979), the only so far known Campanula fossil, can be used for internally calibrating the tree (Cellinese & al. 2009).

**Phyteuma clade** — The Campanula persicifolia lineage occupies an isolated position as sister to all remaining members of the Phyteuma clade in this study (Fig. 4, 5, maximum support) and all other published molecular studies (variable support). Eddie & al. (2003) and Roquet & al. (2008) furthermore show that C. stevenii M. Bieb. (E Europe) and C. pterocaula Hausskn. (Tur-
key) are part of the crown group for this lineage. *C. persicifolia* L. is a large-flowered species with a wide European range. Branching after *C. persicifolia* in the *Phyteuma* clade is *C. trichocalycina* Ten. It is a perennial peculiar in having a bluish violet corolla deeply divided into linear lobes, not unlike those of *Petromarula* R. Hedw. or *Asyneuma* Griseb. & Schenk. Its taxonomic position remains controversial: first described under *Campanula*, the species has variously been assigned to *Podanthum* (G. Don) Bois., *Phyteuma* L. or *Asyneuma* as recently again adopted by Conti & al. (2005). Lakušić & Conti (2004) subdivided it into *A. trichocalycinum* s.str. (S Italy and Sicily) and *A. pichleri* (Vis.) D. Lakušić & F. Conti (Balkan Peninsula and Crete), to which latter the material we studied pertains. *Petromarula pinnata* (Fig. 1E) is a monotypic genus endemic to Crete and is rather isolated morphologically, being unique among European *Campanulaceae* in having compound leaves. ITS data (Roquet & al. 2008) and the 3-gene analysis (Cellinese & al. 2009) resolve it as sister to the *Physoplexis-Phyteuma* clade, a position which, for once, is not fully corroborated by the *petD* trees (Fig. 4, 5). The astonishing position of *Legousia* Durande as sister to *Petromarula* (Fig. 4) is not corroborated by the Bayesian phylogram (Fig. 5) and should be viewed with care. *Phyteuma* appears monophyletic, bearing in mind that only four of the 24 European species (Damboldt 1976) were sampled, and appears sister to the monotypic *Physoplexis* (Endl.) Schur, endemic to the SE Alps.

**Phylogenetic utility of petD sequences in Campanula**

The *petD* marker was first characterized in a study of early branching angiosperms (Löhne & Borsch 2005) and has since been applied successfully to infer deep nodes in eudicots (Worberg & al. 2007) and relationships in the orders *Nymphaeales* (Löhne & 2007) and *Malpighiales* (Korotkova & al. 2009). Notably, recent analyses in *Rubiaceae*, tribe *Spermacoceae*, revealed *petD* sequences to contain much better phylogenetic signal than other non-coding regions (such as *trnL-F* or *rps16*) (Kårehed & al. 2008; Groeninkx & al. 2009). This is the first study using a larger set of *petD* sequences at the species level.

Although the comparison of phylogenetic structure in the *petD* data set with other data sets for *Campanulaceae* is currently difficult because *R* is strongly influenced by taxon sampling and published molecular data sets deviate considerably from our data set, a trend appears to be obvious. *R* is a measure for the average statistical support related to the total number of nodes in a tree that can be expected from a taxon set when the tree is fully resolved (Müller & al. 2006). The maximum parsimony tree based on ITS as shown by Eddie & al. (2003) depicts several lineages in *Campanuloideae*, such as the *Platyodon* clade (100 % JK), the sister-group relationship of the *Campanula rotundifolia + Phyteuma* clade (100 %), the *Jasione* clade (100 %) and the *C. latifolia* clade, each of which is hold together by its firmly similar sequence block. However, deep nodes describing relationships between these clades are not resolved with confidence. Within each clade, only few sublineages are supported. ITS trees in *Campanulaceae* are not very well resolved, and an examination of the sequences available in GenBank indicates that unlike plastid spacers and introns large parts of the ITS1 and ITS2 spacers cannot be aligned reliably between major clades. Cano-Maqueda & al. (2008) also hint to a lower performance of ITS compared to *trnT-L* sequence data. To the contrary, most nodes of the backbone of *Campanulaceae* in trees derived from *petD* sequence data receive high JK values.

Equal or better phylogenetic performance of *petD* alone is apparent when compared to trees inferred from combined genes (*rbcL-atpB+matK*) such as the 3-gene data set by Cellinese & al. (2009). Compared to the three genes *petD* has < 20 % of sequenced nucleotides. Nevertheless, resolution and support of the respective Bayesian reconstructions (Fig. 5 in this study; Cellinese & al. 2009; fig. 3) is largely similar in the overall topologies for the *Campanuloideae*. There are even some nodes with higher support in the *petD* tree, such as the crown group node for the *Campanula rotundifolia* clade (0.81 PP in the 3-gene analysis; 1.0 PP in the *petD* study), the crown group node for the *Phyteuma* clade (0.83 PP versus 1.0 PP) or the node depicting the sister-group relationship of the two aforementioned clades (0.81 PP versus 1.0 PP). The results obtained here indicate a similar performance as studies on basal angiosperms based on non-coding regions such as *trnT-trnF* or *petD* (Borsch & al. 2003; Löhne & Borsch 2005), which yielded phylogenetic hypotheses comparable to those based on five to nine genes, with a four to seven times higher amount of sequence data applied (Qiu & al. 1999, 2005). The practical aspects of “phylogenetic utility” such as easy amplification and sequencing will be addressed below.

**Relative mutational rates and resolution contrast between major lineages of Campanula**

The rate of molecular evolution in a genomic region (both substitutions and microstructural mutations) can vary over time within a lineage and mutational rates can shift with the evolution of new lineages, leading to the well known “rate heterogeneity”. As a first approximation we calculated average relative substitutional rates in the *petD* intron data set for all major *Campanulaceae* groups found in this study based on a maximum likelihood estimate (GTR+I+F+I model) of substitutions per site between taxa. Significantly lower relative substitution rates were found in the *Campanula latifolia* clade (Fig. 6), which also shows lower internal branch lengths. The causes for mutational rates are different. In addition to a widespread trend of short-lived plants to exhibit higher rates (e.g., Smith & Donoghue 2008; con-
firming the “generation-time hypothesis”), other factors influencing the underlying patterns of fixation of mutations such as differences in DNA repair efficiency, fidelity of polymerases and metabolic changes leading to altered selective pressures are discussed.

Mutational rates within genomic regions vary by a factor of six to eight throughout angiosperms as has been shown for example for the matK gene (Müller & al. 2004). Molecular mutational rates also influence the performance of phylogenetic markers, through different amounts of phylogenetic structure. When average mutational rates of a genomic locus in a radiating clade are high, this will lead to a higher probability for mutations to accumulate along internal branches of this radiation as compared to when mutational rates are low. Lineage specific resolution contrasts have been found in Piperaceae (good resolution in Peperomia versus poor resolution in Piper; Wanke & al. 2007) and in Lentibulariaceae (good resolution in the Genlisea-Utricularia clade versus poor resolution in Pinguiscula; Jobson & al. 2003; Müller & al. 2004). In both cases high mutational rates in the genome correlate to high resolution and support in the respective tree compartments. In a similar way, rate heterogeneity seems to lead to a resolution contrast in Campanulaceae. The poorly resolved Campanula latifolia clade also exhibits significantly lower molecular evolutionary rates in petD (Fig. 6).

Cellinese & al. (2009) found that a molecular clock in a Bayesian environment (with the program BEAST), the authors found much younger divergence times of the taxa within the Campanula latifolia clade (major splits around 10-15 Ma) as compared to the C. rotundifolia and Phyteuma clades (major splits around 20-40 Ma). The respective effects of time differences in diversification (i.e., the speed of organismic radiation) versus changes in underlying molecular evolutionary rates throughout the evolutionary history of Campanula need to be carefully evaluated. It will therefore be relevant to all dating analyses and to the analysis of species diversification how effectively shifts in molecular mutational rates can be corrected by means of approaches such as “rate smoothing” through penalized likelihood (Sanderson & al. 2002), or the application of “relaxed clocks” in a Bayesian framework (see Renner 2005 for review). Future work will, in particular, have to test the respective effects of rapid species diversification and lowered mutational rates within the C. latifolia clade.

Utility of petD for molecular species identification in Campanula s.l.

Sequences of petD (the group II intron and the petB-petD spacer) allow to recognize about 90 % of the species of Campanuloideae included in this study. Moreover, the marker is of practical utility as barcode (see, e.g., the recent discussion by Devey & al. 2009 and

Fig. 6. Rate plot comparing relative molecular substitutional rates across major clades in Campanulaceae and allies with Nicotiana tabacum as reference taxon and Atropa bella-donna as outgroup. – $K_{ij} =$ maximum likelihood estimate (GTR+Γ+I model) of substitutions per site between taxa.
Borsch & Quandt (2009). PetD is a genomic region that proved to be easily amplified in Campanulaceae using the universal primers of Löhne & Borsch (2005), also from herbarium specimens (up to 20 years old). Fragments of about 900 nt in size could be easily sequenced. The absence of larger polyA/T stretches is relevant for getting long and unambiguous reads. Using an ABI3730XL automated sequencer, reads of >800 nt could be obtained without any N, covering both strands with one primer respectively. For barcoding application, shorter areas could also be selected. Watts & al. (2008) designed universal primers for amplification of the D4 loop of chloroplast introns. A similar approach was carried out by Taberlet & al. (2007) for the P8 loop of the trnL intron. However, applicability depends on the level of taxonomic resolution needed for a particular question. Moreover, further work will be necessary to test performance of shorter petD fragments at the species level.

Contrary to phylogenetic analysis, sequence elements from mutational hotspots can be included for species identification. One of the most important issues for accurate species identification with DNA sequences, however, is to evaluate intraspecific variability, for example through assessing haplotype variation in an approach integrating phenotypic and molecular characters. The present study gives several examples (e.g. the Campanula rotundifolia alliance s.str., the C. patula + C. spatulata clade, the C. cervaria + C. glomerata clade, the C. spicata + C. thyrsoides clade) for lineages where species might have diversified faster than petD sequences but where, essentially, species limits are still unclear. Work for the coming years therefore needs to first integrate molecular and morphological approaches for achieving robust monographic information on the species in Campanula.

5. Conclusions and future work

Sequence data of the petD group II intron allowed to generate well resolved trees at the level of species and closely related genera within the Campanulaceae, further underscoring the utility of this marker as a phylogenetic tool. For several nodes of the tree of Campanula the gene tree of this non-coding region was better supported than nodes inferred by a much larger multi-gene data set. Our study therefore provides yet another example of the high potential of non-coding and rapidly evolving chloroplast DNA for resolving phylogenetic relationships. Nevertheless, the phylogenetic structure needs to be compared for different chloroplast genomic regions, especially introns and spacers, using identical taxon sets. Such comparative analyses of molecular evolution will allow us to get further insights into structural and mutational patterns that influence the accumulation of historical signal. Group II introns with a high proportion of stem-loops (i.e., the longer introns such as rpl16) appear to be particularly promising. Considering the results of this and other recent studies it will certainly be possible to select genomic regions as markers for high performance. In line with a more detailed argumentation by Borsch & Quandt (2009) a good option will be to combine high performing intron and spacer data sets. A combination of several highly performing markers will certainly be needed to generate well resolved species level phylogenies for diverse genera. Fully resolved trees are, however, needed from both the organelle and nuclear genomic compartments to detect possible incongruence as evidence for reticulations in genome evolution. Hybridization and introgression may have played an important role in Campanula, as known series of polyploidy in some lineages and also extant hybrids indicate.

Campanula s.l. certainly is one of the most extreme cases in terms of a genus that is paraplectic to a number of currently accepted other genera, with proposals to re-classify having been made in either the splitting or lumping direction (see Roquet & al. 2008). Results from phylogenetic analysis imply a highly adaptive evolution of floral morphology, rapid acquisition of features that adapt species to a habitat, both radiations of species groups in geographically confined areas and long distance migrations of ancestors leading to geographically remote ranges of extant, closely related species. As a consequence, there is little predictability in pre-cladistic classification systems to guide the design of taxon sampling, such as providing an answer to the question of which species to include for studying a given question. In addition, there are widespread, variable species that appear to be closely related to ecologically specialised endemics in certain regions.

Research in several directions appears necessary. First of all, relationships within core Campanuloideae, among the two major radiations of Campanula and the Jasione and Musschia clades, need to be clarified. This is the prerequisite for any sound revision of the classification system. Second, the taxon sampling has to be considerably increased to assess, whether there are any further isolated lineages in core Campanuloideae and to further evaluate the composition of their major clades. Many traditionally recognized species groups should be defined phylogenetically before examining the often unclear species limits. A comprehensive effort based on one or two genomic regions that are chosen for their high phylogenetic utility now appears to be feasible, with sampling extended to all the 300-400 species, in order to generate a phylogenetic framework for future evolutionary and monographic research in Campanula.

Acknowledgements

We thank the staff of the living collections department of the Botanic Garden and Botanical Museum Berlin-Dahlem, especially Thomas Dürbye, Matthias Kendzia and Michael Meyer, and of the Bonn University Botani-
cal Gardens, especially Anett Krämer, who are working for maintaining the Campanulaceae collections with great enthusiasm. We appreciate access to further material through the curator of the herbarium of the State Museum Georgia (TMG). This project benefited a lot from support during field work by Zurab Schewardnadze in Georgia, by Jan Koštúth in Slovakia and by Anneliese Borsch, Yamina und Simone Kuhlmann in various places of the Alps. Sarah Kunz helped in the lab during the 2006 annual molecular systematics course at the Nees-Institute in Bonn, and Kerstin Wilhelm during the molecular systematics course in 2008 at Oldenburg University, during which a large number of petD sequences for this study were generated. Alice Bergfeld and Kim Govers supported the labwork in Berlin. Karsten Salomo (Dresden) provided unpublished petD data for comparison from the asterids. We appreciate constant support by Wilhelm Barthlott (Bonn), also by funding the position of the second author through the Akademie der Wissenschaften und der Literatur zu Mainz. We thank Werner Greuter (Berlin) for his insightful comments to an earlier version of this paper. Helpful comments of two reviewers are also acknowledged. Support by the Deutsche Forschungsgemeinschaft for the project “Mutational dynamics of non-coding genomic regions and their potential for reconstructing evolutionary relationships in eudicots” (BO1815 /2-1 and /2-2) to the first author is acknowledged. This project provided the framework for the analysis of the petD intron molecular evolution and supported the generation of sequences for representing the asterids. Further support came through the grants program of the Verein der Freunde des Botanischen Gartens und Botanischen Museums Berlin-Dahlem e.V. (Project: “Evolution von Endemismus in der Gattung Campanula (Glockenblumen) im Euro+ Med-Raum”) that allowed to include BGBM collections into this study.

References


APG II 2003: An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. – Bot. J. Linn. Soc. 141: 399-436. [CrossRef]

Barfuss M. H. J., Samuel R., Till W. & Stuessy T. F 2002: Phylogenetic relationships in subfamily Til-landsioideae (Bromeliaceae) based on DNA sequence data from seven plastid regions. – Amer. J. Bot. 92: 337-351. [CrossRef]

Barfuss M. H. J., Samuel R., Till W. & Stuessy T. F 2005: Phylogenetic relationships in subfamily Till-landsioideae (Bromeliaceae) based on DNA sequence data from seven plastid regions. – Amer. J. Bot. 92: 337-351. [CrossRef]


Borsch T. & Quandt D. 2009: Mutational dynamics and phylogenetic utility of non-coding chloroplast DNA. – Pl. Syst. Evol. [CrossRef]


Kelchner S. A. 2002: Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. – Amer. J. Bot. 89: 1651-1669. [CrossRef]


Korotkova N., Schneider J. V., Quandt D., Worberg A., Zizka G. & Borsch T. 2009: Phylogeny of the eudicot order *Malpigiales*: analysis of a recalcitrant clade with sequences of the petD group II intron. – Pl. Syst. Evol. [CrossRef]


Lańcucka-Srodoniowa M. 1979: Macroscopic plant remains from the freshwater Miocene of the Nowy Sacz basin (West Carpathians, Poland). – Acta Palaeobot. 20: 74-75.


Lundberg J. & Bremer K. 2003: A phylogenetic study of the order Asterales using one morphological and three molecular data sets. – Int. J. Pl. Sci. 164: 553-578. [CrossRef]

Manen J. F., Natali A. & Ehrendorfer F. 1994a: Phylogeny of Rubiaceae-Rubiaceae inferred from the sequence of a cpDNA intergene region. – Pl. Syst. Evol. 190: 195-211. [CrossRef]


Posada D. & Crandall K. A. 1998: Modeltest: testing the model of DNA substitution. – Bioinformatics 14: 817-818. [CrossRef]


Simmons M. P. & Ochoterena H. 2000: Gaps as characters in sequence-based phylogenetic analyses. – Syst. Biol. 49: 369-381. [CrossRef]


Smith S. A. & Donoghue M. J. 2008: Rates of molecular evolution are linked to life history in flowering plants. – Science 322: 86-89. [CrossRef]


Tesfaye K., Borsch T., Govers K. & Bekele E. 2007: Characterization of Coffea chloroplast microstellites and evidence for the recent divergence of C. arabica and C. eugenioides chloroplast genomes. – Genome 50: 1112-1129. [CrossRef]


Toor N., Hausner G. & Zimmerly S. 2001: Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. – RNA 7: 1142-1152. [CrossRef]


Wanke S., Jaramillo M. A., Borsch T., Samain M. S., Quandt D. & Neinhuis C. 2007: Power and limita-


Toor N., Hausner G. & Zimmerly S. 2001: Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. – RNA 7: 1142-1152. [CrossRef]


Winkworth R. C., Lundberg J. & Donoghue M. J. 2008: Toward a resolution of Campanulid phylogeny, with special reference to the placement of Dipsacales. – Taxon 57: 53-65

Worberg A., Quandt D., Barniske A.-M., Löhne C., Hül K. W. & Borsch T. 2007: Phylogeny of basal eu-
dicots: insights from non-coding and rapidly evolving DNA. – Organisms Diversity Evol. 7: 55-77. [CrossRef]

Worberg A., Alford M. H., Quandt D. & Borsch T. 2009: Huerteales sister to Brassicales plus Malva-

les, and newly circumscribed to include Dipen-
todon, Gerrardina, Huertia, Perrottetia and Tapis-
cia. – Taxon 58: 468-478.

Appendix 1. Plant material used in this study. Collections made in the field first list country and locality in the case of both herbarium specimens and silica gel dried samples. Samples obtained from the living collections of the Botanic Garden Berlin-Dahlem and Bonn Botanical Gardens first list the garden accession number and then country and locality data in square brackets. Collector and collection number are given in italics, the herbarium abbreviation in parantheses. Information on the specific history of the plant individual sampled can be obtained from the label that goes with the herbarium voucher. A unique identifier is given for every DNA isolate; it follows the label that goes with the herbarium voucher. A unique identifier is given for every DNA isolate; it follows the label that goes with the herbarium voucher. A unique identifier is given for every DNA isolate; it follows the label that goes with the herbarium voucher. A unique identifier is given for every DNA isolate; it follows the label that goes with the herbarium voucher. A unique identifier is given for every DNA isolate; it follows the label that goes with the herbarium voucher.


Campanulaceae-Campanuloideae: Adenophora triphylia (Thunb.) A. DC., BG Bonn 27427 [Japan, Hokkaido Division, Ohasi, Nayoro], Korotkova 76 (BONN), CAM119, FN396977. Asyneuma campanuloides (M. Bieb. ex Sims) Bornm., Germany, Kartli, path from Kazbegi to Devdoraki, Korotkova 25 (B, TGM), CAM045, FN396978. Asyneuma canescens (Waldst. & Kit.) Griseb. & Schenk, BG Bonn 9178 [without locality data], Korotkova 89 (BONN), CAM001, FN396979. Azorina vitalii Perr., BG Bonn 830 [Aizes ex BG Berlin], Korotkova 69 (BONN), CAM019, FN396980. Campanula alliarifolia Willd., Georgia, Kazbegi, close to the Russian border, Korotkova 26 (B, TGM), CAM049, FN396987. C. alpina Jacq., Slovakia, Borsch & Košth 3903 (B), CAM076, FN397041. C. aff. annae Kolak., Georgia, Kartli, village of Kobi, Korotkova & al. 30 (B, TGM), CAM095, FN396982. C. armazica Kharadze, Georgia, Schewardnadze s.n. (BONN), CAM031, FN396988. C. barbata L., Italy, Lago Como, Monte S. Primo, Borsch 3842 (B), CAM052, FN396989; Italy, Lombardia, Monte Bazine, Borsch 3859 (B), CAM085, FN396990. C. baumgartenii J. Becker, BG Bonn 17230 [Germany, Rheinland Pfalz, Pfalz], Buttlar 181 (FR), CAM107, FN396991. C. betulifolia K. Koch, BG Bonn 26245 [without locality data], Korotkova 77 (BONN), CAM102, FN396992. C. aff. bononiensis L., Italy, Province Bozen, Borsch 912 (B), CAM038, FN396983. C. braunsi (Bolle) A. Chev., BG Bonn 13042 [Cape Verde, Fogo, Chá das Chaldeiras, between Povoação and Mte Velha], Kilian & Leyens 3326 (B, BONN), CAM026, FN396993. C. aff. carnica Mert. & W. D. J. Koch, Italy, Lombardia, Mte Tremalzo, Borsch 3847 (B), CAM054, FN396984. C. carpatha Halácsy, BG Berlin 173-54-07-10 [Greece, Karpathos, Vroukounda], Gartenherbar 46201 [orig. leg. Raus & Sipman] (B), CAM125, FN396995. C. carpathica Jacq., Slovakia, Borsch & Košth 3911 (B), CAM080, FN396996. C. cespitosa Scop., Italy, Lombardia, Mte Bazen, Borsch 3860 (B), CAM056, FN396994. C. cervicaria L., Germany, Baden-Württemberg, Hegau, Plören, Borsch s.n. (B), CAM037; FN396997; Italy, Lombardia, Mte Stino, Borsch 3845 (B), CAM082, FN396999; Slovakia, Borsch & Košth 3905 (B), CAM078, FN396998. C. choruensis Kit Tan & Sorger, BG Bonn 26246 [without locality data], Korotkova 78 (BONN), CAM105, FN397000. C. ciliata Stev., Georgia, Kartli, above Bakuriani, Zchazarho, Korotkova 37 (B), CAM047, FN397001. C. cochlearifolia Lam., Germany, Bavaria, Karwendel Mts, Borsch 3819 (B), CAM10, FN397002. C. collina Sims, Georgia, Kartli, Korotkova & al. 19 (B, TGM), CAM096, FN397003. C. darialiaca Kharadze, Georgia, Kartli, path from mountain Gergeti to Kazbegi, Korotkova & al. 23 (B), CAM098, FN397004. C. divaricata Michx., USA, Virginia, Blue Ridge Mts, Borsch & Neinhuis 3283 (FR), CAM033, FN397005. C. edulis Forsk., Ethiopia, Wondafrash s.n. (B, ETH), CAM062, FN397006. C. elatinoides Moretti, Italy, Lombardia, Lago Iseo, Borsch 3863 (B), CAM057, FN397008. C. fedorovii Kharadze, Georgia, Kartli, path from Mt Gergeti to Kazbegi, Korotkova 48 (B), CAM046, FN397009. C. fenestrella Feer, BG Bonn 25632 [Croatia, Velebit], Korotkova 79 (BONN),
Appendix 2. List of indels found in the petD region of Campanulaceae.

The petB-petD spacer

1. “TATAG” SSR, in Campanula elatines and C. portenschlagiana; duplication from the gene end of petB.
2. A 9 nt gap in all members of the Phyteuma clade, probably deletion.
3. “TA”-SSR in all members of the lobelioids.
   4. A 5 nt gap in Platycodon grandiflorum, probably deletion.
5. One-nucleotide gap in all other taxa except Legousia speculum-veneris; probably insertion of “G” in Legousia, creating a satellite-like “AG” repeat sequence.
6. One-nucleotide gap in all members of the core campanuloids and the Wahlenbergia clade, probably deletion.
7. “T” or “G” present in all members of the Campanula rotundifolia clade except Adenophora triphylla and C. turczaninovii, probably insertion.
8. One-nucleotide gap in Lobelia inflata, probably deletion.
9. One-nucleotide gap in all three species of Cyphia, probably deletion.
10. One-nucleotide gap in Campanula turczaninovii and Adenophora, probably deletion.
12. A 10 nt gap in Pratia and the two species of Isotoma, probably deletion.
13. One-nucleotide gap in Grammatopheca, probably deletion.
15. “ATATATTTTGTAT” in Cyananthus, probably insertion but origin not clear.
17. A 4 nt gap in Isotoma arillaris, probably deletion.
18. “C” SSR in Adenophora.
19. One-nucleotide gap in Lobelia nana, probably deletion.
The petD intron

20. A 2 nt gap in Campanula barbata (CAM085 and CAM052), C. hypopolia and in Edraianthus pumilio, probably deletion.
22. A 5 nt gap in both species of Canarina, probably deletion.
23. A one-nucleotide gap in Campanula patula (CAM073 and CAM077) and in C. spatulata, probably deletion.
25. A one-nucleotide gap in CAM051 Lobelia salicina.
26. A 3 nt gap in all species of Cyphia, probably deletion.
27. “A” SSR in Edraianthus tenuifolius.
30. “TGATA” SSR in all species of Cyphia.
31. A one-nucleotide gap in Musschia aurea, overlapping with 32, probably deletion.
32. A 3 nt gap in Asyneuma canescens and Edraianthus pumilio, overlapping with 31, probably deletion.
33. A 4 nt gap in Adenophora triphylla and Campanula turczaninovii, probably deletion.
34. “C” SSR in Brighamnia, Campanula patula (CAM073 and CAM077) and C. spatulata.
35. A 10 nt gap in Cyphia tysonii, probably deletion.
36. A 35 nt gap in Brighamnia, probably deletion.
37. A one-nucleotide gap in both species of Edraianthus, probably deletion.
38. “CCGAC” SSR in Codonopsis and Cyananthus.
40. A 6 nt gap in Edraianthus tenuifolius, probably deletion.
41. A 11 nt gap in Campanula trichocalycina, probably deletion, overlapping with 43.
42. A 9 nt gap in both species of Jasione, probably deletion, overlapping with 43.
43. “GTCAK” in both species of Musschia and Campanula lactiflora, origin unknown.
44. A 4 nt gap in Cyphia peteriana, probably deletion.
45. “A” SSR in Isotoma axillaris.
46. “TAAAG” SSR in CAM051 Lobelia salicina.
47. A 7 to 9 nt gap in Wahlenbercia, overlapping with 48 and 49.
48. A 2 nt gap present in all lobelioids, cyphioids, Canarina, Codonopsis, Cyananthus, Cyclodon and Platycodon, overlapping with 47 and 49.
49. A 4 nt gap in Campanula trichocalycina, overlapping with 48.
50. A 3 nt gap in Campanula ciliata, overlapping with 51 and 52.
51. A 4 nt gap in Campanula carpatha and C. latifolia, overlapping with 50.
52. A 5 nt gap in Adenophora triphylla and Campanula turczaninovii, overlapping with 50 and 51.
53. “GCKS” present in all taxa except core campanuloids, Nesocodon and Wahlenbercia, overlapping with 47.
54. Gap of approx. 51 nt in Cyphia subtubulata, probably deletion, overlapping with 55 to 65.
55. “TTACGTAAA” SSR in Cyananthus.
56. “CCATTGCCTAAAACCATTACAGTAAA” multiple SSR with partial repeats in Codonopsis.
57. A one-nucleotide gap in both species of Jasione, probably deletion, overlapping with 54.
58. “TC” SSR in Cyananthus.
59. An approx. 9 nt gap in Brighamnia, probably deletion, overlapping with 54 and 56.
60. An approx. 8 nt gap in Lobelia deckeni and L. rhynchoptela, probably deletion, overlapping with 54 and 59.
61. “TTCAGA” SSR in Canarina eminii, overlapping with 54.
62. A 9 nt gap in Campanula caesitosa and C. coehlearifolia, probably deletion, overlapping with 54.
63. A 2 nt gap in Campanula latifolia and C. sp. CAM094, probably deletion, overlapping with 54 and 62.
64. A 10 nt gap in Lobelia nana, overlapping with 54 and 65-66.
65. A 3 nt gap in Lobelia inflata, overlapping with 54 and 64.
67. A one-nucleotide gap in Phyteuma betonicifolium, probably deletion.
68. A 5 nt gap in both species of Canarina, probably deletion.
69. “TAGATA” SSR in Asyneuma, Campanula trichocalycina, Legousia, Petromarula, Physoplexis and all species of Phyteuma.
70. A one-nucleotide gap in Campanula tomentosa, probably deletion.
71. “A” in Cyphia peteriana, probably insertion, of unknown origin.
72. “G” SSR in both species of Jasione.
73. “A” SSR in Grammatotherca.
74. A 3 nt gap in Campanula bravensis, C. edulis and C. jacobaeae, probably deletion.
Appendix 3. Positions of mutational hotspots in the individual sequences of petD that were excluded from phylogenetic analysis.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Pos. hotspot 1</th>
<th>Pos. hotspot 2</th>
<th>Pos. hotspot 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outgroup taxa</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atropa bella-donna</td>
<td>371-376</td>
<td>532-561</td>
<td>609-617</td>
</tr>
<tr>
<td>Lactuca sativa</td>
<td>335-343</td>
<td>496-531</td>
<td>594-594</td>
</tr>
<tr>
<td>Nicotiana tabacum</td>
<td>371-377</td>
<td>533-562</td>
<td>610-618</td>
</tr>
<tr>
<td>Pentaphragmaacamatum</td>
<td>381-388</td>
<td>549-583</td>
<td>625-628</td>
</tr>
<tr>
<td>Sphenoecla zeylanica</td>
<td>377-389</td>
<td>519-526</td>
<td>567-568</td>
</tr>
<tr>
<td>Styliidiu majus</td>
<td>367-375</td>
<td>532-571</td>
<td>614-626</td>
</tr>
<tr>
<td><strong>Campanulaceae-Campanuloideae</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenophora triphylla</td>
<td>334-344</td>
<td>504-523</td>
<td>569-571</td>
</tr>
<tr>
<td>Asyneuma campanuloides</td>
<td>338-348</td>
<td>513-532</td>
<td>584-586</td>
</tr>
<tr>
<td>Asyneuma canescens</td>
<td>335-345</td>
<td>510-529</td>
<td>581-583</td>
</tr>
<tr>
<td>Azorina vidualii</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula alliariifolia</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula alpina</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula annae</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula armazica</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula barbata</td>
<td>336-344</td>
<td>509-528</td>
<td>574-576</td>
</tr>
<tr>
<td>Campanula barbata</td>
<td>336-344</td>
<td>509-528</td>
<td>574-576</td>
</tr>
<tr>
<td>Campanula baumgartenii</td>
<td>338-348</td>
<td>516-535</td>
<td>574-576</td>
</tr>
<tr>
<td>Campanula betulifolia</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula bracteata</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula aff. carnica</td>
<td>338-348</td>
<td>516-535</td>
<td>581-583</td>
</tr>
<tr>
<td>Campanula carpatica</td>
<td>333-343</td>
<td>508-527</td>
<td>572-574</td>
</tr>
<tr>
<td>Campanula cervicaria CAM037</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula cervicaria CAM078</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula cervicaria CAM082</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula cespitosa</td>
<td>337-347</td>
<td>512-531</td>
<td>568-570</td>
</tr>
<tr>
<td>Campanula choruhensis</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula ciliata</td>
<td>338-346</td>
<td>508-527</td>
<td>573-575</td>
</tr>
<tr>
<td>Campanula cochlearifolia</td>
<td>338-348</td>
<td>513-532</td>
<td>569-571</td>
</tr>
<tr>
<td>Campanula collina</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula dariaica</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula divaricata</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula edulis</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula elatinoides</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula fodorovii</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula fenestrellata</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula foliosa</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula fragilis</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula glorodera</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula grossheimii</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula hierapetrae</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula hohenackeri</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula hypopola</td>
<td>339-347</td>
<td>512-531</td>
<td>577-579</td>
</tr>
<tr>
<td>Campanula incureva</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula isophylla</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula jacquaea</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula kolenatiana</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula lactiflora</td>
<td>338-348</td>
<td>518-537</td>
<td>583-585</td>
</tr>
<tr>
<td>Campanula latifolia</td>
<td>338-346</td>
<td>507-526</td>
<td>570-572</td>
</tr>
<tr>
<td>Campanula makaschvili</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula medium</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula patula CAM073</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula patula CAM077</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula persicifolia CAM006</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula persicifolia CAM081</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula petrophila</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula portenschlagiana</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Taxon</td>
<td>Pos. hotspot 1</td>
<td>Pos. hotspot 2</td>
<td>Pos. hotspot 3</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Campanula pyramidalis</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula rapunculoides</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula rhomboidalis</td>
<td>338-348</td>
<td>516-535</td>
<td>581-583</td>
</tr>
<tr>
<td>Campanula rotundifolia CAM008</td>
<td>338-348</td>
<td>516-535</td>
<td>581-583</td>
</tr>
<tr>
<td>Campanula rotundifolia s.l. CAM084</td>
<td>338-348</td>
<td>516-535</td>
<td>581-583</td>
</tr>
<tr>
<td>Campanula rotundifolia s.l. CAM086</td>
<td>338-348</td>
<td>516-535</td>
<td>581-583</td>
</tr>
<tr>
<td>Campanula rumeliana</td>
<td>338-346</td>
<td>511-536</td>
<td>582-584</td>
</tr>
<tr>
<td>Campanula sarmatica</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula sartorii</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula saxifraga</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula scheuchzeri</td>
<td>338-348</td>
<td>516-535</td>
<td>581-583</td>
</tr>
<tr>
<td>Campanula sibirica subsp. divergentiformis</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula spatulata</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Campanula sp. CAM094</td>
<td>338-346</td>
<td>511-530</td>
<td>574-576</td>
</tr>
<tr>
<td>Campanula spicata</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula thersala</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula thyrsoides CAM103</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula thyrsoides CAM106</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula tomentosa CAM023</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula trachelium CAM007</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula trachelium CAM083</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula trichocalycina</td>
<td>338-348</td>
<td>498-517</td>
<td>569-571</td>
</tr>
<tr>
<td>Campanula tridentata</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Campanula turczanzinovii</td>
<td>334-342</td>
<td>502-524</td>
<td>567-569</td>
</tr>
<tr>
<td>Canarina versicolor</td>
<td>338-348</td>
<td>513-532</td>
<td>578-580</td>
</tr>
<tr>
<td>Canarina canariensis</td>
<td>333-343</td>
<td>510-535</td>
<td>576-578</td>
</tr>
<tr>
<td>Canarina eimi</td>
<td>333-343</td>
<td>510-535</td>
<td>582-584</td>
</tr>
<tr>
<td>Codonopsis lanceolata</td>
<td>338-348</td>
<td>520-546</td>
<td>619-621</td>
</tr>
<tr>
<td>Cyananthus lobatus</td>
<td>338-348</td>
<td>520-545</td>
<td>602-604</td>
</tr>
<tr>
<td>Cyclocodon lanceolatum</td>
<td>338-348</td>
<td>515-539</td>
<td>585-587</td>
</tr>
<tr>
<td>Edraianthus pumilio</td>
<td>333-341</td>
<td>505-524</td>
<td>570-572</td>
</tr>
<tr>
<td>Edraianthus tenuifolius</td>
<td>339-347</td>
<td>505-524</td>
<td>570-572</td>
</tr>
<tr>
<td>Jasione laevis</td>
<td>336-344</td>
<td>500-519</td>
<td>564-566</td>
</tr>
<tr>
<td>Jasione montana</td>
<td>336-344</td>
<td>500-519</td>
<td>564-566</td>
</tr>
<tr>
<td>Legousia speculum-veneris</td>
<td>338-346</td>
<td>511-530</td>
<td>582-584</td>
</tr>
<tr>
<td>Michauxia campanuloides</td>
<td>338-346</td>
<td>511-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Musschia aurea</td>
<td>337-347</td>
<td>517-536</td>
<td>582-584</td>
</tr>
<tr>
<td>Musschia wollastoni</td>
<td>338-348</td>
<td>518-537</td>
<td>583-585</td>
</tr>
<tr>
<td>Nesocodon mauritianus</td>
<td>338-348</td>
<td>509-528</td>
<td>574-576</td>
</tr>
<tr>
<td>Petromarula pinnata</td>
<td>338-348</td>
<td>513-533</td>
<td>586-588</td>
</tr>
<tr>
<td>Physoplexis comosa</td>
<td>338-348</td>
<td>513-532</td>
<td>584-586</td>
</tr>
<tr>
<td>Phyteuma aff. orbiculare</td>
<td>338-350</td>
<td>515-534</td>
<td>586-588</td>
</tr>
<tr>
<td>Phyteuma betonicifolium</td>
<td>338-350</td>
<td>515-534</td>
<td>585-587</td>
</tr>
<tr>
<td>Phyteuma hemisphaericum</td>
<td>338-350</td>
<td>515-534</td>
<td>586-588</td>
</tr>
<tr>
<td>Phyteuma scheuchzeri</td>
<td>338-350</td>
<td>515-534</td>
<td>586-588</td>
</tr>
<tr>
<td>Platycodon grandiflorum</td>
<td>338-348</td>
<td>515-541</td>
<td>587-589</td>
</tr>
<tr>
<td>Symphyandra hofmannii</td>
<td>338-346</td>
<td>511-526</td>
<td>572-574</td>
</tr>
<tr>
<td>Wahlenbergia saxicola</td>
<td>338-344</td>
<td>504-523</td>
<td>570-572</td>
</tr>
<tr>
<td>Campanulaceae-Cyphioideae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyphia petersoni</td>
<td>340-350</td>
<td>513-530</td>
<td>576-578</td>
</tr>
<tr>
<td>Cyphia subtilulata</td>
<td>340-351</td>
<td>518-529</td>
<td>533-535</td>
</tr>
<tr>
<td>Cyphia tsonii</td>
<td>340-350</td>
<td>507-524</td>
<td>570-572</td>
</tr>
<tr>
<td>Campanulaceae-Lobelioideae</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brighamia insignis</td>
<td>339-347</td>
<td>479-503</td>
<td>540-542</td>
</tr>
<tr>
<td>Grammatophoca bergiana</td>
<td>343-351</td>
<td>518-540</td>
<td>586-588</td>
</tr>
<tr>
<td>Isotoma azillaris</td>
<td>339-347</td>
<td>515-539</td>
<td>585-587</td>
</tr>
<tr>
<td>Isotoma flaviatilis</td>
<td>338-346</td>
<td>513-535</td>
<td>581-584</td>
</tr>
<tr>
<td>Lobelia deckeni</td>
<td>338-346</td>
<td>513-537</td>
<td>575-577</td>
</tr>
<tr>
<td>Lobelia inflata</td>
<td>338-346</td>
<td>513-537</td>
<td>583-585</td>
</tr>
<tr>
<td>Lobelia nana</td>
<td>338-346</td>
<td>513-536</td>
<td>572-574</td>
</tr>
<tr>
<td>Lobelia rhynchometalum</td>
<td>338-346</td>
<td>513-537</td>
<td>575-578</td>
</tr>
<tr>
<td>Lobelia salicina</td>
<td>337-345</td>
<td>517-541</td>
<td>587-588</td>
</tr>
<tr>
<td>Pratia nummularia</td>
<td>338-346</td>
<td>513-537</td>
<td>583-585</td>
</tr>
<tr>
<td>Siphocampylus manettiiflorus</td>
<td>338-346</td>
<td>513-537</td>
<td>583-586</td>
</tr>
</tbody>
</table>