

## THE PHYLOGENY OF ROSOIDEAE (ROSACEAE) BASED ON SEQUENCES OF THE INTERNAL TRANSCRIBED SPACERS (ITS) OF NUCLEAR RIBOSOMAL DNA AND THE *trnL/F* REGION OF CHLOROPLAST DNA

Torsten Eriksson,<sup>1,\*†</sup> Malin S. Hibbs,<sup>‡</sup> Anne D. Yoder,<sup>§</sup> Charles F. Delwiche,<sup>‡</sup> and Michael J. Donoghue<sup>§</sup>

\*Bergius Foundation, Royal Swedish Academy of Sciences, Box 50017, SE-104 05 Stockholm, Sweden; †Botaniska Institutionen, Stockholms Universitet, SE-106 91 Stockholm, Sweden; ‡Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland 20742-5815, U.S.A.; and §Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, Connecticut 06520-8106, U.S.A.

The phylogeny of Rosoideae was investigated using 44 species. Here we report new sequence data from the chloroplast *trnL/F* region as well as an increased sample of species. The analysis of these new data, along with previously used data from the nuclear ribosomal internal transcribed spacers (ITS), significantly increased resolution as well as confidence for Rosoideae phylogeny. Using both Bayesian inference and parsimony methods, we conducted analyses on the data sets separately and in combination. The resulting phylogenies are congruent with all well-supported clades of Rosoideae found in previous analyses of ITS or *rbcL* data. The support for these and other clades is improved, and we consider several clades to be supported well enough to be named. The following clades are given phylogenetic definitions: Sanguisorbeae and its subclades Agrimoniinae and Sanguisorbinae, Potentilleae and its subclades Fragariinae and *Potentilla*, Roperculina (*Rosa* + Sanguisorbeae + Potentilleae), and Sanpotina (Sanguisorbeae + Potentilleae). *Potentilla* includes the *Potentilla anserina* clade (*Argentina*) in our *trnL/F* and combined analyses, but this relationship is not resolved by ITS alone. The previously used genera *Duchesnea* (*Potentilla indica*), *Horkelia*, and *Ivesia* are strongly supported as nested within *Potentilla*. *Comarum* (*Potentilla palustris* and *Potentilla salesowianum*), *Sibbaldiopsis* (*Potentilla tridentata*), *Dasiphora* (*Potentilla fruticosa*), and *Drymocallis* (*Potentilla arguta*) join *Alchemilla*, *Aphanes*, *Sibbaldia*, *Chamaerhodos*, and *Fragaria* in the well-supported Fragariinae clade outside of *Potentilla*. The monophyly of both Potentilleae and Sanguisorbeae is well supported, and the clades correspond to previously named tribes with the exception of *Alchemilla* and its segregate *Aphanes*, which are nested within Potentilleae instead of in Sanguisorbeae. The position of *Rubus* is still not securely resolved.

**Keywords:** molecular systematics, phylogenetic analysis, phylogenetic taxonomy, chloroplast DNA introns, ribosomal DNA, Bayesian inference, Rosaceae, Rosoideae, *Potentilla*.

### Introduction

The monophyly of Rosoideae was first established by Morgan et al. (1994). They analyzed chloroplast *rbcL* sequences and found good support for a first phylogenetic split in a monophyletic Rosaceae. One of the two main branches in this first split was a Rosoideae from which a few groups had been removed as compared to previous classifications (Focke 1894; Schulze-Menz 1964; Robertson 1974). Traditional classifications of Rosaceae mainly used fruit morphology as a primary defining criterion, and the standard view of the subfamily Rosoideae was that it comprised those members of Rosaceae with achenes or drupelets. It turned out that the removal of some groups from Rosoideae actually was more in line with other data, such as chromosome base numbers and chemistry, rather than with general fruit classification (Morgan et al. 1994). Here we use the name Rosoideae for this reduced clade.

The first study to explore the phylogenetic relationships within the Rosoideae used nuclear internal transcribed spacers

(ITS) sequence data (Eriksson et al. 1998), which were correlated with some morphological data (Vretblad et al. 1996). The main goal was to investigate the monophyly of *Potentilla*, which had a complex classification history and which was suspected to be paraphyletic. In particular, the relationships of several potential segregates from *Potentilla* were of interest. These segregates displayed shared morphological similarities, some with each other and some with *Potentilla* species. For instance, *Duchesnea* and *Fragaria* share remarkably similar pseudo fruits (“strawberries”) and at the same time have shoot and flower characteristics that are distinctive for different groups of *Potentilla*. The results from that analysis showed *Potentilla*, as treated in any previous classification, to be paraphyletic or polyphyletic. Several clades of Rosoideae were well supported while others were only weakly supported and perhaps spuriously resolved (Eriksson et al. 1998).

For a more secure hypothesis of phylogeny in Rosoideae, we have expanded the data set to include more taxa and more data; specifically, we have examined new sequences of the *trnL/F* region of chloroplast DNA. Here we present the results of this expanded analysis.

Rosoideae in its wider sense was previously classified into

<sup>1</sup> Author for correspondence; e-mail torsten.eriksson@bergianska.se.

several tribes: Rubeae, Roseae, Ulmarieae, Dryadeae, Sanguisorbeae (or Poterieae), and Potentilleae (Hutchinson 1964). The first three groups are reasonably well diagnosed morphologically and are probably monophyletic. *Rubus* was shown to be monophyletic with *Dalibarda* included (Alice and Campbell 1999) or as sister to *Rubus* (chloroplast DNA; L. A. Alice, personal communication). The former case implies that *Rubus* and Rubeae may be phylogenetically synonymous, i.e., two different names for the same clade. There are indications that Roseae and Ulmarieae may be similar cases (Morgan et al. 1994; T. Eriksson, unpublished data). Dryadeae, however, was shown to be a polyphyletic assemblage, and *Dryas*—along with *Purshia*, *Cowania*, and *Cercocarpus*—at present are not considered to belong to Rosoideae (instead they form a separate clade; Morgan et al. 1994; Swensen and Mullin 1997). A recent parsimony analysis of chloroplast DNA sequences gave weak support for Dryadeae as the first clade to split off the rest of Rosaceae (Potter et al. 2002). The exact position of Dryadeae in Rosaceae is not known, however, except that it is not within the clade Rosoideae, and we do not include it in this analysis. The remnants of Dryadeae, which do belong in Rosoideae, are currently under investigation (Smedmark and Eriksson 2002; Smedmark et al. 2003). The monophyly of the last two major groups, Sanguisorbeae and Potentilleae, has so far not been well documented. On the contrary, some analyses have indicated, albeit weakly, that Sanguisorbeae as well as Potentilleae might be polyphyletic (Eriksson et al. 1998). Hence, the main focus of this article is on investigation of the possible monophyly and the relationships between Sanguisorbeae and Potentilleae.

## Material and Methods

### Selection of Taxa

We used a similar set of taxa as in Eriksson et al. (1998): a representative selection of Rosoideae with an emphasis on *Potentilla* and segregates from *Potentilla*. Some additions were made (1) in some groups, especially the Sanguisorbeae, to test their monophyly; (2) in regions of the tree where long branches might be suspected to interfere with analysis; and (3) where a too meager sampling of species might cause spurious resolution. In total, 13 species were added (table 1).

Four species were exchanged: *Filipendula ulmaria* to *vulgaris*, *Sanguisorba parviflora* to *officinalis*, *Waldsteinia fragarioides* to *geoides*, and *Fragaria ananassa* to *virginiana*. In the first three cases the switch was motivated by an intent to use type species where available, and in the last case we switched because of the availability of wild collected material. Two species were removed from all or some analyses because we failed to get good sequences from the *trnL/F* region: *Potentilla erecta* and *Hagenia abyssinica*.

The ITS region was resequenced for five species in the present analysis because those previous sequences included many uncertain base scorings (*Fragaria vesca*, *Potentilla anserina*, *Aphanes arvensis*, *Potentilla palustris*, and *Potentilla indica*). Originally, DNA had been extracted from herbarium material of these species, but we now had access to fresh material.

Five sequences were published elsewhere (Helfgott et al. 2000; Smedmark and Eriksson 2002). Previous analyses using

*rbcL* and ITS data gave good support for *Filipendula* as the sister group of the rest of Rosoideae (Morgan et al. 1994; Eriksson et al. 1998). In order to reduce problems of alignment, we removed the two non-Rosoideae species used in the analysis of Eriksson et al. (1998)—specifically, *Dryas octopetala* and *Prunus cerasifera*—and we rooted our trees on the *Filipendula* branch.

### Molecular Markers and Methods

Concerted evolution of ribosomal DNA repeats (including the ITS region) may be a problem if there are instances of allopolyploid speciation within the group (Wendel et al. 1995). Such problems might be avoided by choosing only diploid species. However, we were not able to select only diploid species, and there is a risk that some relationships among close relatives reflected in the ITS data may be erroneous. Despite this potential problem, ITS is generally considered to be of great utility for phylogenetic analysis among closely related species (Baldwin 1992; Baldwin et al. 1995), and we wanted to expand the existing sample of Rosoideae ITS sequences. Furthermore, with the addition of chloroplast DNA sequences, maternally inherited data were made available for comparison. The new ITS sequences were amplified as previously (Eriksson et al. 1998), except that the ITS-I primer of Urbatsch et al. (2000) was used instead of ITS5 (White et al. 1990).

We selected the *trnL/F* region of chloroplast DNA, which has been shown to be informative among closely related species (Wikström et al. 1999), albeit less variable than ITS. Using the c-f primer pair of Taberlet et al. (1991), we amplified a segment containing the spacer between *trnL* and *trnF* and the intron in *trnL*. The amplification primers along with the internal primers d and e were used for sequencing. Amplification of a longer fragment using primers a-f failed in Rosoideae.

Amplification products were cleaned using Qiagen spin tubes and cycle sequenced using BigDye premix and standard protocols, except that 2  $\mu$ L premix (diluted to 8  $\mu$ L) was used per reaction and run out on an ABI automated sequencer. Base scorings were proofread and assembled using the Staden package (Staden 1996) under Linux.

### Alignment

All alignments were made manually using the alignment editor Se-Al (Rambaut 1996) on Apple Macintosh computers. The previous ITS alignment was used as a basis for further additions and augmentation of the ITS alignment (Eriksson et al. 1998). Here we took a conservative approach regarding ambiguous portions of the alignment and refrained from using some parts that were clearly ambiguous. Alignment was problematic in these regions because of overlapping indels of different lengths or a lack of similarity that could be used as a basis for alignment. In one particular case at the 3' end of the ITS data (positions 1962–1997), it was possible to align the sequences more or less unambiguously in some groups of species but not over the entire selection of species. This region was therefore removed from the analyses. The following positions were excluded from all analyses: 556–566, 722–728, 1153–1189, 1357–1359, 1409–1414, 1462–1467, 1802–1809, and 1962–1997. In total, 114 positions were removed.

The aligned data set along with the trees published here have been submitted to TreeBase (<http://www.treebase.org>).

#### Treatment of Indels

Indel positions were coded “–” and treated as uncertain data in applicable analyses. Forty-one of the inferred indels were coded as separate binary characters (table 2). We used two criteria for determining what indels to use: the inferred indel should have the same length in all taxa where it was present, and the indel should not overlap ambiguously with other indels of different lengths (fig. 1). The indels were scored for presence (1) or absence (0) of the inferred gap and were added at the end of the data set: 24 from the *trnL/F* data and 17 from the ITS data. The length of the indels varied from a single up to 147 base pairs.

#### Phylogenetic Analyses

A modified version of Modeltest 3.06 (Posada and Crandall 1998) was used with PAUP\* (Swofford 2001) to select an appropriate evolutionary model for use in Bayesian inference analyses. This particular version (MrModeltest 1.0b; J. A. A. Nylander, personal communication) specifically tests the 24 models available common to PAUP\* and MrBayes 2.01 (Huelsenbeck and Ronquist 2001a). The hierarchical likelihood ratio tests selected the General Time Reversible model with gamma distribution of rates (GTR+ $\Gamma$ ) for both data sets (Rodríguez et al. 1990; Yang et al. 1994; Yang 1996). Estimations of phylogeny with Bayesian inference used the MrBayes program (Huelsenbeck and Ronquist 2001a, 2001b). The data sets were analyzed, separately and in combination, using the model suggested by MrModeltest with the following settings: use six substitution types, estimate base frequencies, allow rates of sites to follow a gamma distribution with four categories, estimate gamma shape. The combined analysis used the GTR+ $\Gamma$  model for both data set partitions with the rates = ssgamma option, which allows site-specific rates, but the rates within each partition vary as described by a separately estimated gamma distribution. Four chains (Markov Chain Monte Carlo) were run for 1 million generations, and a tree was sampled every tenth generation. The binary indel characters were not included in the MrBayes analyses. The parsimony analyses were conducted using the UNIX version of PAUP\* (Swofford 2001) with TBR branch swapping and the MULTREES setting on 1000 random addition sequence starting trees. The ACCTRAN option was used for character optimization.

We used three separate measures of node support. Majority rule consensus trees of the trees sampled in Bayesian inference analyses yielded probabilities that the clades are monophyletic (Lewis 2001). The trees from the MrBayes analyses were loaded into PAUP\*, discarding the trees sampled during the “burnin” of the chain (Huelsenbeck and Ronquist 2001b), and 95% majority rule consensus trees were made. Thus, the trees shown in the figures only contain nodes estimated to have a posterior probability of 0.95 or more. Under the parsimony criterion, we performed bootstrap (Felsenstein 1985) and decay analyses (Bremer 1988; Donoghue et al. 1992). PAUP\* was set to run 10,000 bootstrap replicates, each using TBR branch swapping on a single random addition sequence start-

ing tree, with a single tree saved per bootstrap replicate. Bootstrap results are shown as 70% bootstrap majority rule consensus trees. Decay analyses were performed using AutoDecay (Eriksson 1999) and reverse constraints in PAUP\*. Each reverse constraint run was conducted with TBR branch swapping and MULTREES on 100 random addition sequence starting trees. In this article, Bayesian inference clade probabilities are given without prefix, parsimony bootstrap values are percentages prefixed by “b,” and decay indices are prefixed by “d.” Bootstrap values, widely used as measures of support, have been shown to be conservative measures of clade accuracy probabilities (Hillis and Bull 1993). In many of their simulations, and in an experimental phylogeny, bootstrap values (proportions) of 70% corresponded to a probability of more or less 95% clade accuracy. Based on this, we have chosen to use 70% bootstrap majority rule consensus trees for comparison with the 95% Bayesian inference trees, a commonly used level for statistical significance. It is our purpose to present the well-supported clades here, and we therefore refrain from presenting the default 50% bootstrap trees, single optimal trees, or strict consensus trees, with the single exception of the tree with branch length estimates (fig. 5).

To compare the total support for the trees, we used a B70 index. This is the proportion of nodes in a fully resolved tree with a bootstrap support of 70% or higher (cf. Sanderson and Donoghue 1996). For the Bayesian inference trees, we used a Bayes95 index, calculated similarly.

## Results

The ITS data set that consisted of 44 taxa comprised 713 aligned DNA characters and 17 binary indel characters. Of these, 59 DNA characters were excluded from all analyses because of ambiguous alignment (see above), and of the remaining, 248 were informative for parsimony analysis. The length of the sequences of ITS including 5.8S varied from 589 bases in *Aremonia agrimonioides* to 651 in *Agrimonia eupatoria*. The G+C content was 60%. Of the sequenced positions, 0.7% were scored ambiguous, mainly because of low-quality DNA in certain extractions. In addition to this, 17% of the data set was coded as uncertain (indel positions “–” or unsequenced positions “?” added for alignment purposes).

The Bayesian inference tree is based on 96,802 trees sampled from chain generations 32,000–1,000,000 (fig. 2). The parsimony analysis found four most parsimonious trees of 1030 steps (consistency index [CI] 0.4942, excluding uninformative characters 0.4422; Kluge and Farris 1969). The mean CI of binary indel characters was 0.81. An analysis excluding *Hagenia*, which was lacking from the *trnL/F* data set, gave trees of 1015 steps. The topologies were the same except for the lack of the node with *Hagenia*. The support for the node below remained the same after the removal of *Hagenia*. The 70% parsimony bootstrap tree and the 95% Bayesian tree are fully congruent and differ only in that some clades are collapsed in each of the trees (cf. arrows in fig. 2). The B70 index as well as the Bayes95 index is 0.66.

The *trnL/F* data set comprised 43 taxa with 1297 aligned DNA characters and 24 binary indel characters. Of these, 55 DNA characters were excluded from all analyses, and of the remaining, 281 were informative for parsimony analysis. The

**Table 1**  
**Species Used, Accessions, Voucher Information or Publication Where Sequence Was First Used,**  
**Origin of Materials, and EMBL/Genbank Accession Numbers**

Species and DNA region	Voucher/reference	Origin of material	Accession number
<i>Acaena cylindristachya:</i>			
Both	Hibbs 167	Andes	AJ512775, AJ512780
<i>Acaena laevigata:</i>			
Both	Hibbs 6	RBGE; Falkland Islands	AJ512776, AJ512781
<i>Agrimonia eupatoria:</i>			
ITS	Eriksson et al. 1998	Sweden, Uppland	U90798
<trnl f<="" td=""> <td>Eriksson and Smedmark 41 (SBT)</td> <td>HB; Germany</td> <td>AJ512216</td> </trnl>	Eriksson and Smedmark 41 (SBT)	HB; Germany	AJ512216
<i>Alchemilla alpina:</i>			
ITS	Eriksson et al. 1998	Sweden, Torne Lappmark	U90816, U90817
<trnl f<="" td=""> <td>Eriksson 805 (SBT)</td> <td>HB; Sweden, Torne Lappmark</td> <td>AJ512217</td> </trnl>	Eriksson 805 (SBT)	HB; Sweden, Torne Lappmark	AJ512217
<i>Alchemilla mollis:</i>			
Both	Eriksson s.n. (SBT)	HB; unknown	AJ511769, AJ512218
<i>Aphanes arvensis:</i>			
Both	Eriksson s.n. (SBT)	Sweden, Uppland	AJ511770, AJ512234
<i>Aremonia agrimonioides:</i>			
ITS	Eriksson et al. 1998	LD; unknown	U90799
<trnl f<="" td=""> <td>Karlsson 94076 (LD)</td> <td>LD; unknown</td> <td>AJ512230, AJ512231</td> </trnl>	Karlsson 94076 (LD)	LD; unknown	AJ512230, AJ512231
<i>Chamaerhodos erecta:</i>			
ITS	Eriksson et al. 1998	U.S.A., Montana	U90794
<trnl f<="" td=""> <td>Norlindh and Ahti 10161A (S)</td> <td>Mongolia</td> <td>AJ512219</td> </trnl>	Norlindh and Ahti 10161A (S)	Mongolia	AJ512219
<i>Fallugia paradoxa:</i>			
ITS	Eriksson et al. 1998	U.S.A., New Mexico	U90805
<trnl f<="" td=""> <td>Smedmark and Eriksson 2002</td> <td>U.S.A., Colorado</td> <td>AJ297331</td> </trnl>	Smedmark and Eriksson 2002	U.S.A., Colorado	AJ297331
<i>Filipendula vulgaris:</i>			
Both	Eriksson 821 (SBT)	Sweden, Uppland	AJ416467, AJ416463
<i>Fragaria vesca:</i>			
Both	Eriksson and Smedmark 43 (SBT)	Sweden, Uppland	AJ511771, AJ512232
<i>Fragaria virginiana:</i>			
Both	Eriksson s.n. (SBT)	Canada, Nova Scotia	AJ511772, AJ512220
<i>Geum urbanum:</i>			
ITS	Eriksson et al. 1998	Sweden, Uppland	U90802
<trnl f<="" td=""> <td>Smedmark and Eriksson 2002</td> <td>Sweden, Uppland</td> <td>AJ297323</td> </trnl>	Smedmark and Eriksson 2002	Sweden, Uppland	AJ297323
<i>Hagenia abyssinica:</i>			
ITS	Eriksson et al. 1998	Kenya	U90800
<i>Horkelia fusca:</i>			
Both	Eriksson et al. 1998; same DNA used here	U.S.A., California	U90795, AJ512247
<i>Ivesia gordonii:</i>			
Both	Eriksson et al. 1998; same DNA used here	U.S.A., Utah	U90796, AJ512221
<i>Leucosidea sericea:</i>			
ITS	Helfgott et al. 2000		AF183547, AF183524
<trnl f<="" td=""> <td>D.M. Helfgott Ben-3 1998 (TEX)</td> <td>Unknown</td> <td>AJ512222</td> </trnl>	D.M. Helfgott Ben-3 1998 (TEX)	Unknown	AJ512222
<i>Polylepis hieronymi:</i>			
Both	Hibbs 133, 1998	Bolivia	AJ512774, AJ512779
<i>Polylepis tarapacana:</i>			
Both	Hibbs 163, 1999	Bolivia	AJ512773, AJ512778
<i>Potentilla anserina:</i>			
Both	Eriksson and Smedmark 44 (SBT)	Sweden, Uppland	AJ511773, AJ512238
<i>Potentilla arguta:</i>			
ITS	Eriksson et al. 1998	U.S.A., Washington	U90787
<trnl f<="" td=""> <td>Eriksson s.n. (SBT)</td> <td>Unknown</td> <td>AJ512223</td> </trnl>	Eriksson s.n. (SBT)	Unknown	AJ512223
<i>Potentilla bifurca:</i>			
ITS	Eriksson et al. 1998	Sweden, Uppland	U90786
<trnl f<="" td=""> <td>Eriksson 811 (SBT)</td> <td>Sweden, Uppland</td> <td>AJ512224</td> </trnl>	Eriksson 811 (SBT)	Sweden, Uppland	AJ512224
<i>Potentilla chinensis:</i>			
Both	Eriksson s.n. (SBT)	HB; China, Beijing	AJ511774, AJ512225
<i>Potentilla dickinsii:</i>			
ITS	Eriksson et al. 1998	Korea	U90785
<trnl f<="" td=""> <td>Crompton, D'Arcy &amp; Coke 139 (E)</td> <td>RBGE; Korea</td> <td>AJ512243</td> </trnl>	Crompton, D'Arcy & Coke 139 (E)	RBGE; Korea	AJ512243

**Table 1**  
(Continued)

Species and DNA region	Voucher/reference	Origin of material	Accession number
<i>Potentilla fragarioides</i> :			
ITS	Eriksson et al. 1998	Japan	U90806, U90807
<trnl f<="" td=""> <td>Eriksson s.n. (SBT)</td> <td>HB; China, Beijing</td> <td>AJ512226</td> </trnl>	Eriksson s.n. (SBT)	HB; China, Beijing	AJ512226
<i>Potentilla fruticosa</i> :			
ITS	Eriksson et al. 1998	LD; Sweden, Öland	U90808, U90809
<trnl f<="" td=""> <td>Eriksson 806 (SBT)</td> <td>HB; Sweden, Öland</td> <td>AJ512233</td> </trnl>	Eriksson 806 (SBT)	HB; Sweden, Öland	AJ512233
<i>Potentilla indica</i> :			
Both	Eriksson s.n. (GH, SBT)	HB; China, Gansu	AJ511775, AJ512242
<i>Potentilla micrantha</i> :			
ITS	Eriksson et al. 1998	LD; unknown	U90812, U90813
<trnl f<="" td=""> <td>Eriksson and Smedmark 42 (SBT)</td> <td>HB; Greece</td> <td>AJ512227</td> </trnl>	Eriksson and Smedmark 42 (SBT)	HB; Greece	AJ512227
<i>Potentilla multifida</i> :			
Both	Eriksson 705 (SBT)	Sweden, Torne Lappmark	AJ511776, AJ512245
<i>Potentilla nivea</i> :			
Both	Eriksson et al. 1998; same DNA used here	Sweden, Torne Lappmark	U90814, U90815, AJ512244
<i>Potentilla norvegica</i> :			
Both	Eriksson et al. 1998; same DNA used here	U.S.A., Massachusetts	U90790, AJ512246
<i>Potentilla palustris</i> :			
Both	Eriksson 659 (GH, S)	Sweden, Uppland	AJ511777, AJ512237
<i>Potentilla peduncularis</i> :			
Both	Eriksson and Vretblad TE758 (SBT)	GB; China, Yunnan	AJ511778, AJ512239
<i>Potentilla reptans</i> :			
ITS	Eriksson et al. 1998	Sweden, Uppland	U90784
<trnl f<="" td=""> <td>Eriksson 822 (SBT)</td> <td>Sweden, Uppland</td> <td>AJ512241</td> </trnl>	Eriksson 822 (SBT)	Sweden, Uppland	AJ512241
<i>Potentilla salesowianum</i> :			
Both	Eriksson and Vretblad TE751 (SBT)	GB; unknown	AJ511779, AJ512228
<i>Potentilla stenophylla</i> :			
Both	Eriksson and Vretblad TE763 (SBT)	GB; China, Yunnan	AJ511780, AJ512240
<i>Potentilla tridentata</i> :			
ITS	Eriksson et al. 1998	U.S.A., New Hampshire	U90791
<trnl f<="" td=""> <td>Eriksson and Smedmark 40 (SBT)</td> <td>Canada, Nova Scotia</td> <td>AJ512236</td> </trnl>	Eriksson and Smedmark 40 (SBT)	Canada, Nova Scotia	AJ512236
<i>Rosa majalis</i> :			
Both	Eriksson et al. 1998; same DNA used here	Sweden, Värmland	U90801, AJ512229
<i>Rosa persica</i> :			
Both	Eriksson and Smedmark 1 (SBT)	Uppsala Bot. Garden; Iran or Afghanistan	AJ416468, AJ416466
<i>Rubus chamaemorus</i> :			
ITS	Eriksson et al. 1998	Sweden, Torne Lappmark	U90803
<trnl f<="" td=""> <td>Eriksson 809 (SBT)</td> <td>Sweden, Västmanland</td> <td>AJ416464</td> </trnl>	Eriksson 809 (SBT)	Sweden, Västmanland	AJ416464
<i>Sanguisorba officinalis</i> :			
ITS	Helfgott et al. 2000		AF183533, AF183556
<trnl f<="" td=""> <td>Eriksson 804 (SBT)</td> <td>HB; unknown</td> <td>AJ416465</td> </trnl>	Eriksson 804 (SBT)	HB; unknown	AJ416465
<i>Sibbaldia procumbens</i> :			
ITS	Eriksson et al. 1998	Sweden, Torne Lappmark	U90820, U90821
<trnl f<="" td=""> <td>Eriksson 698 (SBT)</td> <td>Sweden, Torne Lappmark</td> <td>AJ512235</td> </trnl>	Eriksson 698 (SBT)	Sweden, Torne Lappmark	AJ512235
<i>Tetraglochin cristatum</i> :			
Both	Hibbs 150, 1999	Andes	AJ512777, AJ512782
<i>Waldsteinia geoides</i> :			
Both	Smedmark and Eriksson 2002	Stockholm University; unknown	AJ302362, AJ297348

Note. GB = Göteborg Botanical Garden; HB = Bergius Botanic Garden, Stockholm; LD = Botanic Garden Lund; RBGE = Royal Botanic Gardens Edinburgh.

**Table 2**  
**Indels Scored as Separate Binary Characters and Data on Position and Characteristics in the Combined Data Set**

Region and indel char no.	Length of indel	Position in alignment	No. of taxa sharing gap	CI <sup>a</sup>	Overlap <sup>b</sup>
<i>trnL/F</i> :					
2011	4	121–124	3	1.0	–
2012	3	226–228	30	1.0	–
2013	8	239–246	2	1.0	–
2014	1	253	39	1.0	–
2015	1	279	40	1.0	–
2016	14	294–307	30	0.5	–
2017	1	316	41	1.0	+2
2018	14	316–330	2	1.0	–3
2019	3	321–323	41	1.0	+2
2020	101	349–449	3	1.0	–3
2021	147	366–512	13	1.0	–3
2022	1	686	8	1.0	+1, 2
2023	1	687	6	1.0	+1, 2
2024	5	740–744	41	1.0	–
2025	12	823–834	2	1.0	–
2026	7	844–850	4	1.0	+1, 2
2027	4	858–861	41	1.0	–1
2028	115	871–985	2	1.0	–3
2029	6	903–908	10	1.0	+1, 2
2030	93	906–998	3	1.0	–3
2031	10	927–936	40	1.0	+2
2032	1	958	41	1.0	+2
2033	13	962–974	4	1.0	+2
2034	10	1226–1235	41	1.0	–
<i>ITS</i> :					
2035	1	1336	2	1.0	–
2036	1	1363	41	1.0	–
2037	1	1376	2	1.0	–
2038	6	1462–1467	41	1.0	+2
2039	1	1484	40	1.0	–1
2040	1	1485	2	1.0	–1
2041	3	1486–1488	38	1.0	–
2042	1	1536	35	0.5	–
2043	1	1543	35	0.25	–
2044	1	1554	17	0.5	–
2045	1	1580	27	1.0	–1
2046	1	1581	40	1.0	–1
2047	1	1826	39	0.5	–
2048	1	1845	30	1.0	–
2049	1	1856	3	1.0	–
2050	1	1922	42	0.5	–
2051	2	1951–1952	13	0.5	–

<sup>a</sup> Consistency indices (CI) relate to the most parsimonious trees of the parsimony analysis.

<sup>b</sup> 1 = gaps, where present, were of the same length but were adjacent to other gaps or variable regions; 2 = gap was scored uncertain, with a question mark, where it overlaps with other gap (fig. 1); 3 = gap overlaps with another gap, but it was not considered problematic (fig. 1).

sequences varied in length from 881 (*A. agrimonioides*) to 1122 (*Potentilla peduncularis*). The G+C content was 33%. Of the sequenced positions, 0.15% were scored ambiguous because of low-quality DNA in a few species. In total, 25%

of the data set was coded as uncertain, which was almost exclusively from indel positions.

The tree from the Bayesian inference analysis of *trnL/F* data alone is based on 94,302 trees sampled from chain generations 57,000–1,000,000 (fig. 3). The parsimony analysis found five trees of 718 steps (CI = 0.7577, 0.6842, excluding uninformative characters; fig. 3). The mean CI of binary indel characters was 0.98. The B70 index is 0.72, and the Bayes95 index is 0.75.

The Bayesian inference 95% majority rule consensus tree from the combined analysis is based on 96,802 trees sampled from chain generations 32,000–1,000,000 (figs. 4, 5). The parsimony analysis resulted in two trees of 1751 steps (CI = 0.5991, 0.5289, excluding uninformative characters), and the mean CI of binary indel characters was 0.82. These parsimony trees are 18 steps (1%) longer than the sum of lengths of the separate shortest trees. The 70% bootstrap tree from the parsimony analysis (not shown) is almost identical to the Bayesian tree in figure 4. Two nodes with bootstrap 68% and 62% show up (indicated by asterisks in fig. 4), and two are collapsed in the Bayesian tree. In terms of total support, both the B70 and the Bayes95 indices are 0.82.

## Discussion

The results from the combined analysis (fig. 4) are congruent with all of the well-supported clades found in the analysis of Eriksson et al. (1998), and the differences are not strongly supported. The addition here of more taxa and new data does much to improve our understanding of the phylogeny of Rosoideae, and several major clades that were previously weak and tentative are now well supported. A phylogenetic “backbone” of Rosoideae is emerging.

The good support for clades makes it desirable to name them, and the proposed PhyloCode (<http://phylocode.org>) shows a way to do this in a formal manner that the current International Code of Botanical Nomenclature (ICBN; Greuter et al. 2000) cannot. However, to gain stability and at the same time minimize confusion, where possible we use names for clades that are also consistent with accepted names for ranked taxa under ICBN (cf. appendix).

The close correspondence between the bootstrap trees and the Bayesian trees in both the separate and the combined analyses is remarkable, especially when we take into account that we are comparing two methods of analysis that are very different. One is based on an explicit model of evolution while the other is not. One uses a Markov Chain Monte Carlo approach and the other the Phylogenetic Bootstrap. As Bayesian inference of phylogeny is still a relatively new method, it remains to be seen if the good fit between 70% bootstrap and 95% Bayesian inference trees will be found to be the rule in future studies.

### Comparison with Previous ITS Analysis

Compared to the ITS tree published previously (Eriksson et al. 1998), the new ITS trees have similar topology but better support for basal clades (fig. 2). The B70 index increased from 0.42 to 0.66, and the Bayes95 index is also 0.66 for the ITS data. As before, parsimony resolves *Rubus* to be sister to the

ACGTACGTACGTACGTACGTACGTACGT	00
ACGTACGTACGT-----ACGTACGTACGT	01
ACGT-----ACGT	1?
ACGT-----ACGT	1?
ACGT-----ACGT	1?
ACGTACGTACGT-----ACGTACGTACGT	01
ACGTACGTACGT-----ACGTACGTACGT	00

**Fig. 1** Hypothetical example to demonstrate how certain overlapping indels were treated (cf. table 2). The horizontal gap is not problematic because it will be present regardless of the overlap. Hence, it is coded as present (1) in the first binary character. It is not possible to know if the more narrow vertical gap is present or not in the area of overlap. The corresponding taxa are therefore coded with question marks in the second binary character.

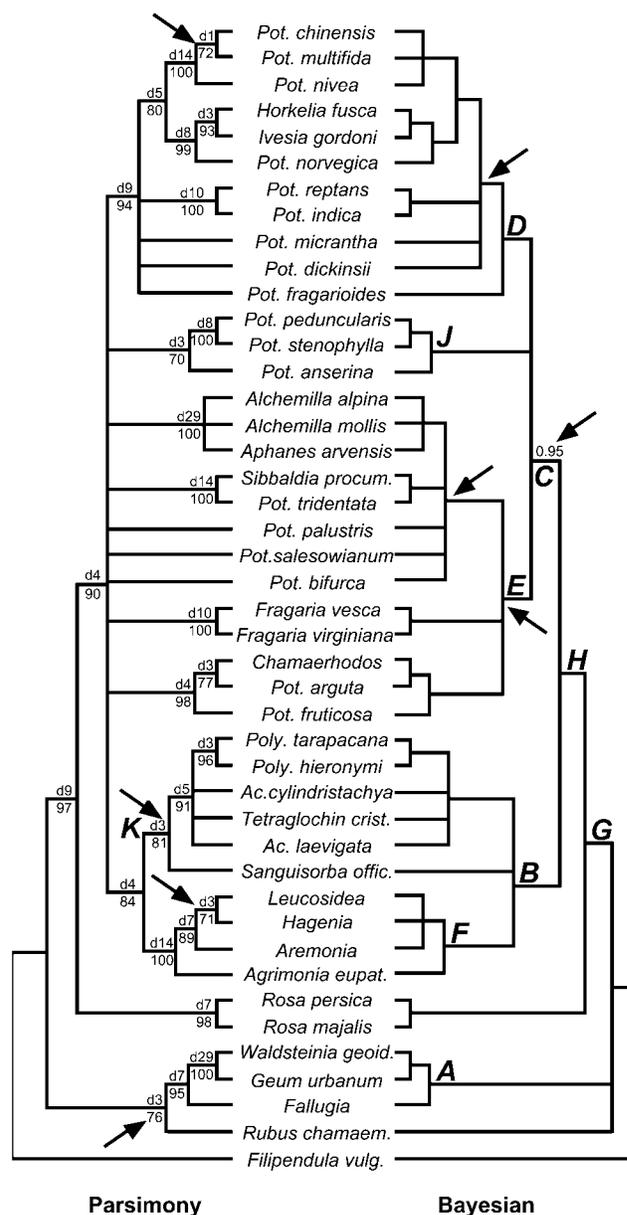
clade now called Colurieae (Smedmark and Eriksson 2002; clade A in fig. 2) with somewhat higher support. The Bayesian analysis gives this relationship low probability. Sanguisorbeae is monophyletic (clade B, 1.0, b84, d4), while it was previously polyphyletic, and Potentilleae is monophyletic as well in the Bayesian tree, but it gets only low support in the parsimony analysis (clade C, 0.95, b42, d1). The Potentilleae clade is divided into two main clades as in the previous tree (Eriksson et al. 1998). One, *Potentilla* in the strict sense, is resolved by both parsimony and Bayesian inference, and its support is increased here (clade D, 1.0, b94, d9; cf. node P3 in Eriksson et al. 1998). Its unnamed sister clade is resolved by Bayesian inference but still has low support in the parsimony analysis (clade E, 1.0, b36, d1). Clade E was resolved in the previous parsimony analysis of ITS, but with very low support. The position of *Potentilla anserina* and two close relatives is unresolved in both trees. However, in the most parsimonious trees, this group is found well nested within clade E close to *Fragaria*, but this has almost no node support (b15 or less, d1).

In the previous analysis, the species of Sanguisorbeae with petals (clade F, the "Agrimonia clade," in Eriksson et al. 1998) was found within Potentilleae but with very low support. In the present analysis, this clade is found to be sister to the rest of Sanguisorbeae (not clearly resolved by Bayesian inference). This is in agreement with previous analyses in which preliminary morphological data had been added to the ITS data (Vretblad et al. 1996).

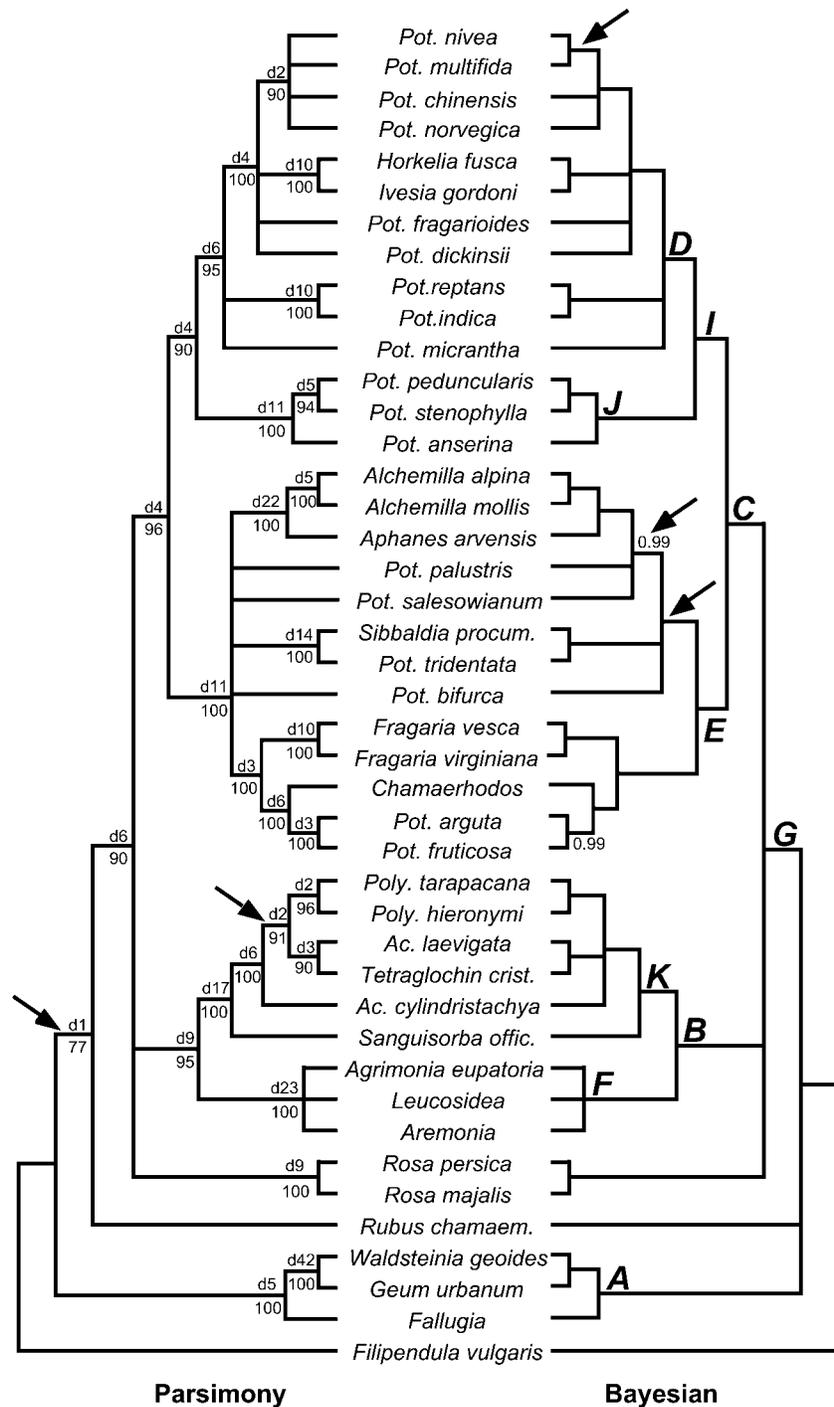
#### TrnL/F and Combined Trees

The *trnL/F* parsimony and Bayesian trees (fig. 3) are fully congruent and differ only in that some clades are collapsed in either of the trees (cf. arrows in fig. 3). They are also very

similar to the ITS trees, and the main clades are the same except that *Rubus* is supported by the parsimony analysis to be sister group to clade G (b77, d1) instead of being sister to Colurieae (clade A), as in ITS. Neither of these positions of *Rubus* gets high probability by Bayesian inference. Also, the clade consisting of Sanguisorbeae + Potentilleae, which is supported in



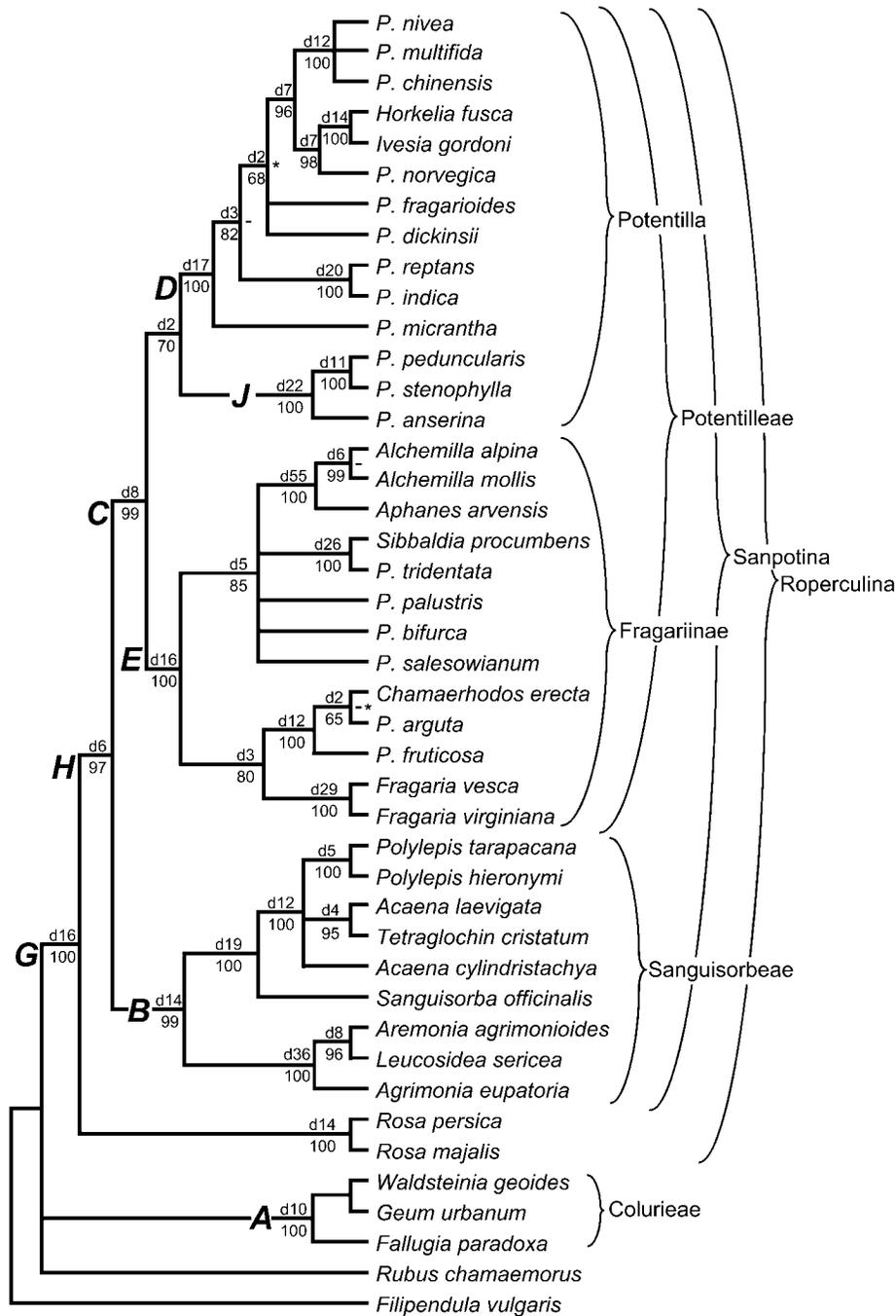
**Fig. 2** Trees from the analysis of internal transcribed spacers data alone. To the left, a 70% majority rule bootstrap consensus tree from the parsimony analysis. Decay indices (prefixed by "d") are found above branches, and bootstrap values are found below branches. To the right, a 95% majority rule consensus tree based on 96,802 trees from the Bayesian inference analysis. All clades shown have estimated probability 1.0 except where indicated. Some nodes discussed in the text are marked by capital letters. Arrows indicate resolved nodes that are collapsed in the corresponding tree. Note that these trees include *Hagenia abyssinica*, which is not present in figs. 3 and 4.



**Fig. 3** Trees from the analysis of *trnL/F* data alone. To the left, a 70% majority rule bootstrap consensus tree from the parsimony analysis. Decay indices (prefixed by “d”) are found above branches, and bootstrap values are found below branches. To the right, a 95% majority rule consensus tree based on 94,302 trees from the Bayesian inference analysis. All clades shown have estimated probability 1.0 except where indicated. Some nodes discussed in the text are marked by capital letters. Arrows indicate resolved nodes that are collapsed in the corresponding tree.

the ITS trees, is lacking in both *trnL/F* trees (cf. clade H in fig. 2). Sanguisorbeae as well as Potentilleae are monophyletic with good support (clades B and C). The Sanguisorbeae clade has an internal resolution that is similar to that of the ITS trees, and the nodes are well supported. One of the two main

clades of Potentilleae, the clade comprising *Fragaria*, *Chamaerhodos*, *Alchemilla*, *Sibbaldia*, and various *Potentilla* satellite species sometimes treated as separate genera (e.g., *Comarum*, *Sibbaldiopsis*, *Drymocallis*, and *Dasiphora* [clade E]), is well supported in both trees (1.0, b100, d11).

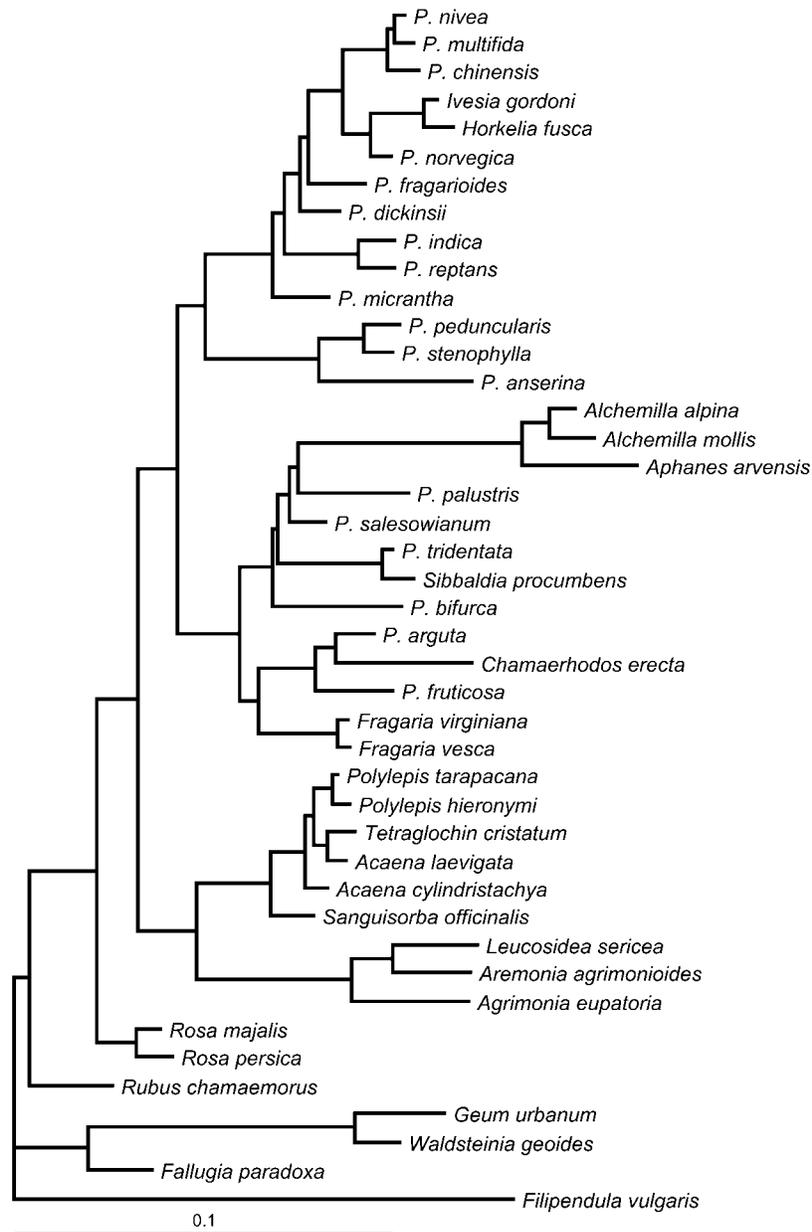


**Fig. 4** 95% majority rule consensus tree from analysis of combined data based on 96,802 trees from the Bayesian inference analysis. All clades shown have estimated probability 1.0 except for three that had 0.99 and are marked by a minus sign. Node support from the parsimony analysis is indicated in the proximity of nodes: decay indices (prefixed by “d”) are found above branches, and bootstrap values are found below branches. Some nodes discussed in the text are marked by capital letters. Asterisks denote two nodes that were collapsed in a 70% parsimony bootstrap consensus. Names of clades are shown to the right with brackets indicating the inclusiveness of the clades.

The internal structure within Potentilleae differs somewhat between the *trnL/F* and ITS trees, but the differences are in general not well supported by one or both DNA regions. See “Clades” for a few exceptions.

The tree from the combined analysis is congruent with the

well-supported clades seen in the separate analyses, and support for these nodes increases in the combined analysis (B70 as well as the Bayes95 index up to 0.82). The differences between the separate analyses are resolved in different ways in the combined analysis. The position of *Rubus*, for example, is the same as in



**Fig. 5** 50% majority rule consensus tree with all compatible clades from the Bayesian inference analysis of combined data. The branches are drawn to indicate the mean branch lengths from the 96,802 trees as calculated by the “sumt contype = allcompat” command in MrBayes (Huelsenbeck and Ronquist 2001b). Scale bar = 0.1 substitutions per site.

the ITS tree by parsimony but with lower support (b57, d1), and it is still unresolved in the Bayesian tree, while the position of *P. anserina* and relatives is as in the *trnL/F* tree, with lower support in the combined parsimony tree (b70, d2) but with a probability of 1.0 in the Bayesian tree.

#### Clades

The well-supported clade C (1.0, b99, d8) corresponds almost exactly to the tribe Potentilleae *sensu* Hutchinson (1964), except that it includes *Alchemilla*, which Hutchinson included

in Sanguisorbeae. The presence of achenes with lateral to basal styles (Wolf 1908) may be a morphological synapomorphy for the clade. Eriksson et al. (1998) discussed this clade as a potential candidate for the name *Potentilla*. However, the clade was only weakly supported in that analysis (b<50, d1), and if given formal genus rank, the current rules of nomenclature would force hundreds of species name changes. Other solutions, assuming a monophyletic taxonomy, would still encompass changes, but less dramatic ones. We therefore attach the name Potentilleae to this clade instead (fig. 4; see appendix for nomenclature).

The first of the two main clades within Potentilleae, clade E, was weakly supported in the previous analysis (Eriksson et al. 1998), apparently because of the unsettled position of the *Agrimonia* clade in that analysis (clade F). The *trnL/F* data give a strong signal for clade E, and this is further strengthened when the ITS data are added (1.0, b100, d16). This clade was quite unexpected when it was first discovered (Eriksson and Donoghue 1995; Eriksson et al. 1998) because it comprised a number of *Potentilla* species mixed in with several other groups (*Fragaria*, *Alchemilla*, *Chamaerhodos*, and *Sibbaldia*), and they had never before been suggested to form a group together. The *Potentilla* species found here have at one time or another been separated from *Potentilla*, forming, for example, *Sibbaldiopsis* (*Potentilla tridentata*), *Comarum* (*Potentilla palustris*), *Drymocallis* (*Potentilla arguta*), and *Dasiphora* (*Potentilla fruticosa*). There are also indications that additional *Potentilla* species may belong in this clade (T. Eriksson and P. Östensson, unpublished data). We use the name *Fragariinae* for this clade (fig. 4). Within *Fragariinae*, the position of *P. (Drymocallis) arguta* is notable because there are supported differences between the ITS tree (fig. 2) and the *trnL/F* tree (fig. 3). In the ITS tree, *P. arguta* and *Chamaerhodos erecta* form a clade (1.0, b77, d3), while *P. (Drymocallis) arguta* instead joins *P. (Dasiphora) fruticosa* in *trnL/F* (0.99, b100, d3). The combined analysis resolves this conflict in favor of the ITS position, but with lower support (fig. 4; 0.99, b65, d2). So far, there are no known morphological synapomorphies for this clade. The “festooned stamens” character is only found within this clade, but the distribution among the species is not yet known in detail (Rydberg 1898; Lindenhofner and Weber 2000).

The other main Potentilleae clade (I) is not especially well supported, but its major component, clade D, is very strongly supported and comprises most of the more than 300 *Potentilla* species along with the embedded segregates *Duchesnea* (*Potentilla indica*), *Horkelia*, and *Ivesia*. It appears probable that additional North American segregates not included here will also join this clade (e.g., *Purpusia*, *Horkeliella*, *Comarella*, and *Stellariopsis*). Internally, clade D is resolved somewhat differently in the separate analyses, and a few of these differences are well supported. For example, *Potentilla norvegica* is sister to *Horkelia* + *Ivesia* in the ITS tree (1.0, b99, d8) while it instead joins a clade with *Potentilla nivea* + *multifida* + *chinensis* in the *trnL/F* tree (1.0, b90, d2). Another example is *Potentilla fragarioides*, which is sister to the rest of the species in clade D in the Bayesian inference ITS tree while nested further within clade D in the *trnL/F* trees (figs. 2, 3). In this case, the clades are weakly supported in the parsimony ITS tree, but a gap in position 654–655 in the ITS data supports the ITS position of *P. fragarioides*. This gap is also present in the *P. anserina* group (clade J). The position of clade J as sister to clade D is well supported in the *trnL/F* tree (1.0, b90, d4), but it is unresolved in the ITS tree (figs. 2, 3). The combined analyses are resolved as in the *trnL/F* trees but with somewhat lower support in the parsimony analyses (b70, d2), which indicates the presence of some conflicting data. Indeed, the most parsimonious trees of the ITS data analyzed independently put the *P. anserina* group as sister to *Fragaria* with very weak support (b14, d1). The problem seems to relate to the rather divergent sequence of *P. anserina*. When it is removed, an ITS analysis yields a trichotomy of clades D, E, and J, as in the

Bayesian inference tree in figure 2. A combined parsimony analysis without *P. anserina* shows clade J as sister group to clade D with support slightly increased compared with the *trnL/F* analysis (b93, d6). Increased sampling may help to resolve the position of the *P. anserina* group more securely.

We have been using *Potentilla* in the strict sense for the species in clade D. In order to minimize species name changes, it is reasonable to apply the name *Potentilla* to a well-supported clade that includes most of the species now classified as *Potentilla* as well as the type species. The best candidate is clade D (cf. clade P3 of Eriksson et al. 1998). However, the *P. anserina* clade (J), which may include ca. 50 species (Soják 1994), has traditionally been included in *Potentilla*, and it would be preferable to avoid renaming those species in a separate genus (*Argentina*). Since clade I is not as well supported as D, we choose a phylogenetic definition of *Potentilla* that is not dependent on the placement of clade J by not giving direct reference to it (see appendix). Thereby, clade J may or may not be included in *Potentilla*, as future research will establish. In the light of the best available evidence (the combined analysis), *Potentilla* includes the *P. anserina* group, but if additional data show that it is sister to clade E, for example, it will be automatically excluded.

The Sanguisorbeae clade is well supported in all of our analyses (clade B, 1.0, b99, d14), and the tribe Sanguisorbeae has consistently been recognized for a long time (Jussieu 1789; Candolle 1825; Focke 1894; Weimarck 1934; Hutchinson 1964). There is no support in the molecular data for the inclusion of *Alchemilla* in Sanguisorbeae, which has been suggested by several authors based on the four-merous flowers and absence of a corolla as in *Sanguisorba* and relatives (but see Schulze-Menz 1964). *Alchemilla* is well nested within Potentilleae, and this is supported by the almost basal position of its styles on the ovules. Most classifications of Sanguisorbeae have included a number of species in which a normal corolla is present (clade F). In previous analyses based on ITS data alone, the connection between the petalous and apetalous groups was tenuous, and Sanguisorbeae often appeared as polyphyletic, albeit with very weak support (Eriksson et al. 1998). In those analyses, the *Agrimonia* clade (clade F), i.e., the clade of Sanguisorbeae with petals, could easily “move around” to different branches of the phylogeny without much change in tree length. This was probably caused by the rather divergent sequence of *Agrimonia*, and it resulted in very low support for many of the nodes and a spurious most parsimonious position of clade F inside the Potentilleae. The addition of just a few morphological characters made a monophyletic Sanguisorbeae more parsimonious (Vretblad et al. 1996).

In the present analysis, the increased sampling of apetalous Sanguisorbeae gave the same effect, namely, that Sanguisorbeae is monophyletic and now with good support. Forcing clade F into the position that was most parsimonious in the previous ITS analysis results in distinctly longer trees: 11 steps longer using ITS only and 37 steps longer for the combined data set. In the present analysis, the petalous *Agrimonia*-group forms a clade (F) with very good support in all analyses, and it is sister to the apetalous clade. To this clade belongs *Agrimonia* as well as the remarkable East African tree *Hagenia* (fig. 2). Further sampling (e.g., of the Asian *Spenceria*) will show if this sister-group relationship holds up or if the petalous

group is paraphyletic. The sampling among apetalous Sanguisorbeae (clade K) is limited, but it is nevertheless notable that the large and diverse *Acaena* appears to be paraphyletic. The Andean tree *Polylepis* as well as the dwarf shrub *Tetraglochin* (along with *Margyricarpus*; M. S. Hibbs, unpublished data) seem to be nested within *Acaena*. This is not entirely unexpected given that *Acaena* has a wide Austral-Antarctic distribution and is quite variable in fruit and leaf morphology as well as in habit. Apetalous Sanguisorbeae are currently under investigation with a focus on the origin of *Polylepis* (M. S. Hibbs, unpublished manuscript).

The clade of Potentilleae plus Sanguisorbeae (clade H, 1.0, b97, d6) is another clade in which support is increased. The node was present in the previous ITS analysis, but with low support (b<50, d2). Clade H is well supported by sequence data, but no morphological synapomorphies are known. The reduction in the number of stamen whorls from many to three or fewer superficially seems to mark this clade, but this is probably incorrect (Lindenhofer and Weber 2000). We use the name Sanpotina for this clade (cf. appendix). The clade one node closer to the root of the tree, comprising Sanpotina plus *Rosa*, i.e., clade G (1.0, b100, d16), is also better supported compared to previous and separate analyses. This clade, which we name Roperculina, is supported by the presence of operculate pollen (Morgan et al. 1994).

Many clades received improved support through data-set combination and by adding taxa. Some clades differ between the separate analyses, but support increases for those clades that remain when the data are combined. Such circumstances are probably the result of the presence of phylogenetic signal in one data set that is “hidden” in the optimal tree(s) (Barrett et al. 1991). There are a few such clades in our analyses, such as *Aremonia* plus *Leucosidea*, but most of them have less than 70% parsimony bootstrap in ITS, and they are collapsed in figure 2. In a few other cases, one of the clades was favored by the combined analysis, but the support was clearly reduced compared to the separate analyses. Such differences may indicate a real conflict between the data sets, such as hybridization events leading to different nuclear DNA/chloroplast DNA gene trees or lineage sorting events in closely related species. If so, this is an indication of a “local” conflict between the data sets concerning the taxa in question but not necessarily indicative of a general conflict in the data. However, sampling can be an important issue for phylogenetic analyses (Rannala et al. 1998; Rydin and Källersjö 2002), and it may be that these conflicts are simply spurious results caused by the limited sampling, especially within clade I (*Potentilla*).

The position of *Rubus* is particularly problematic. In all analyses using Bayesian inference, its position is unresolved. In parsimony analyses, however, chloroplast DNA data favor a position in which *Rubus* is sister to Roperculina (clade G in fig. 3; *rbcL* tree in Morgan et al. 1994) while ITS data favor *Rubus* as sister to Colurieae (clade A in fig. 2; Eriksson et al.

1998). In part, this may be an effect of sampling: when the sampling is increased in Colurieae, *Rubus* joins Roperculina in analyses of ITS data (Smedmark and Eriksson 2002). However, when sampling was increased in *Rubus*, the support for *Rubus* as sister to Colurieae increased (Eriksson et al. 1998). *Rubus*, *Fallugia*, *Geum*, and *Waldsteinia* share a four base pair indel in a region of *trnL/F* where many gaps overlap (position 886–889, not used as separate character). This gap is somewhat ambiguous but would support the sister group relationship of *Rubus* and Colurieae, which is contrary to the result from the *trnL/F* sequence data alone. Because the support values of these clades decrease in the combined analysis when compared to the separate analyses, it seems that there is a conflict in the data sets. The cause of that conflict is not known.

Problems also remain within Potentilleae. In *Potentilla* (clade I), a better sample of species is needed to alleviate these problems, but it is questionable if the molecular “markers” used here are variable enough to resolve details within such a closely related group. Also, the use of ITS among these groups of *Potentilla* may be particularly problematic because of the (supposedly) common hybridization (Asker 1971; Soják 1986; Yurtsev 1993) and consequences relating to ensuing concerted evolution (Wendel et al. 1995; Smedmark et al. 2003). Several groups of species in *Potentilla* are apomictic where concerted evolution might be unpredictable (Campbell et al. 1997), but this is perhaps only a local phylogenetic problem. However, hybridizations, such as those suggested by Soják (1986), of new species resulting from more distant crosses would be a problem if they are shown to be real. Distant crosses, even without the formation of new species, present hard puzzles to solve when using ITS (Alice et al. 2001). It might be more useful to use a variable nuclear low-copy DNA region, such as introns in GBSSI (*waxy*), which are less influenced by concerted evolution (Mason-Gamer 2001; Smedmark et al. 2003).

Some phylogenetic relationships resulting from separate ITS and *trnL/F* analyses clearly differ within Potentilleae in the present study, and it is possible that these discrepancies are caused by the two DNA regions actually tracking different gene histories. At present, however, our sample is not enough for a detailed investigation.

### Acknowledgments

This study was supported by grants from the Royal Swedish Academy of Sciences and the Bergius Foundation to Torsten Eriksson and NSF Doctoral Dissertation Improvement Grant DEB-9902313 to Malin S. Hibbs. We thank M. Helfgott and coworkers for two ITS sequences and leaves of *Leucosidea*. We are also grateful to Göteborg Botanical Garden and Uppsala University Botanical Garden for supplying leaves and to J. Smedmark and P. Korall for comments on an earlier version of this manuscript.

### Appendix

#### Nomenclature

Below are definitions of clade names and their corresponding ranked names (fig. 4). These phylogenetic definitions are node

based or stem based (de Queiroz and Gauthier 1992, 1994) and follow the preliminary Phylocode (<http://phylocode.org>).

References to ranked names (International Code of Botanical Nomenclature [ICBN]; Greuter et al. 2000) follow Reveal (2000).

The names Sanguisorbeae, Sanguisorbinae, Agrimoniinae, Potentilleae, Fragariinae, and Potentilla have stem-based definitions because these groups are well supported but not well sampled at this point. Any new additions to the root branch of these groups are thereby included in the clade. Because the genus name is an integral part of the species name under ICBN, subsequent changes in genus assignment of specifiers might cause future confusion regarding genera and phylogenetic names. To minimize this potential problem, we use type species as specifiers where possible because the type species will always be included in the genus.

We include a tentative list of genera that, based on current knowledge, are enclosed by the clades defined. Some of these genera have not been sampled here, and the list is therefore based on this analysis, our preliminary morphological investigations, and published sources. Synonyms are generally not included.

#### Roperculina

*Node-based definition.* Roperculina is the least-inclusive clade containing *Rosa cinnamomea*, *Sanguisorba officinalis*, *Fragaria vesca*, and *Potentilla reptans*.

This name is not ranked. Possible ranks might be subfamily or supersupertribe. We know of no usage of such ranks in Rosaceae.

The clade Roperculina comprises *Rosa* (including *Hulthemia*) along with the genera of Sanpotina (cf. below).

#### Sanpotina

*Node-based definition.* Sanpotina is the least-inclusive clade containing *S. officinalis*, *F. vesca*, and *P. reptans*.

This name is not ranked. Possible ranks might be subsubfamily or supertribe. We know of no usage of such ranks in Rosaceae.

The clade Sanpotina comprises the genera of Sanguisorbeae and Potentilleae (cf. below).

#### Sanguisorbeae

*Stem-based definition.* Sanguisorbeae is the most inclusive clade containing *S. officinalis* but not *P. reptans*, *F. vesca*, or *R. cinnamomea*.

*ICBN at tribus rank.* Sanguisorbeae DC.—Prodr. 2:588. 1825.

*Typus.* *Sanguisorba* L.

The clade Sanguisorbeae comprises the genera of Sanguisorbinae and Agrimoniinae (cf. below).

#### Sanguisorbinae

*Stem-based definition.* Sanguisorbinae is the most inclusive clade containing *S. officinalis* but not *Agrimonia eupatoria*, *F. vesca*, or *R. cinnamomea*.

*ICBN at subtribus rank.* Sanguisorbinae Torr. & A. Gray—Fl. N. Amer. 1:428. Jun 1840.

*Typus.* *Sanguisorba* L.

The clade Sanguisorbinae comprises the genera *Acaena*, *Bencomia* (Helfgott et al. 2000), *Cliffortia*, *Dendriopoterium*, *Marcetella*, *Margyricarpus*, *Polylepis*, *Poterium*, *Sanguisorba*, *Sarcopoterium*, and *Tetraglochin*.

#### Agrimoniinae

*Stem-based definition.* Agrimoniinae is the most inclusive clade containing *A. eupatoria* but not *S. officinalis*, *F. vesca*, or *R. cinnamomea*.

*ICBN at subtribus rank.* Agrimoniinae J. Presl—Wsobecny Rostl. 1:502. 1846.

*Typus.* *Agrimonia* L.

The clade Agrimoniinae comprises the genera *Agrimonia*, *Aremonia*, *Hagenia*, *Leucosidea*, and *Spenceria*.

#### Potentilleae

*Stem-based definition.* Potentilleae is the most inclusive clade containing *Potentilla reptans* but not *Sanguisorba officinalis*, *Agrimonia eupatoria*, or *Rosa cinnamomea*.

*ICBN at tribus rank.* Potentilleae Sweet—Brit. Fl. Gard. 2:124. 1825.

*Typus.* *Potentilla* L.

The clade Potentilleae comprises the genera *Potentilla* (cf. below), possibly *Comarella*, *Purpusia*, and *Stellariopsis*, and the genera of Fragariinae (cf. below).

#### Fragariinae

*Stem-based definition.* Fragariinae is the most inclusive clade containing *Fragaria vesca* but not *Potentilla anserina*, *Potentilla reptans*, or *Sanguisorba officinalis*.

*ICBN at subtribus rank.* Fragariinae Torr. & A. Gray—Fl. N. Am. 1:435. 1840.

*Typus.* *Fragaria* L.

The clade Fragariinae comprises the genera *Alchemilla* (including at least *Aphanes* but probably also *Lachemilla* and *Zygalychemilla*), *Chamaerhodos*, *Comarum*, *Dasiphora* (*Pentaphylloides*), *Drymocallis*, *Fragaria*, *Sibbaldia*, *Sibbaldianthe* (= *Schistophyllidium*; Kurtto and Eriksson 2003), and *Sibbaldiopsis*. In addition, some species currently classified as *Potentilla* will be included here (T. Eriksson and P. Östensson, unpublished data).

#### Potentilla

*Stem-based definition.* Potentilla is the most inclusive clade containing *Potentilla reptans* but not *Fragaria vesca*, *Comarum palustre*, or *Sanguisorba officinalis*.

*ICBN at genus rank.* Potentilla L.—Sp. Pl. 1753:495.

*Typus.* *P. reptans* L.

The clade *Potentilla* includes *Argentina*, *Duchesnea*, *Horkelia*, *Ivesia*, and probably *Horkeliella*. Some genera not sampled here are potential candidates for inclusion: *Comarella*, *Purpusia*, and *Stellariopsis*.

## Literature Cited

- Alice LA, CS Campbell 1999 Phylogeny of *Rubus* (Rosaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Am J Bot* 86:81–97.
- Alice LA, T Eriksson, B Eriksen, CS Campbell 2001 Hybridization and gene flow between distantly related species of *Rubus* (Rosaceae): evidence from nuclear ribosomal DNA internal transcribed spacer region sequences. *Syst Bot* 26:769–778.
- Asker S 1971 Some viewpoints on *Fragaria*-*Potentilla*-*d* intergeneric hybridization. *Hereditas* 67:181–190.
- Baldwin BG 1992 Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol Phylogenet Evol* 1:3–16.
- Baldwin BG, MJ Sanderson, MJ Porter, MF Wojciechowski, CS Campbell, MJ Donoghue 1995 The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann Mo Bot Gard* 82:247–277.
- Barrett M, MJ Donoghue, E Sober 1991 Against consensus. *Syst Zool* 40:486–493.
- Bremer K 1988 The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42:795–803.
- Campbell CS, MF Wojciechowski, BG Baldwin, LA Alice, MJ Donoghue 1997 Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Mol Biol Evol* 14:81–90.
- de Candolle AP 1825 *Prodromus systematis naturalis regni vegetabilis*. Treuttel & Würtz, Paris.
- de Jussieu AL 1789 *Genera plantarum, secundum ordines naturales disposita*. Hérissant & Barrois, Paris.
- de Queiroz K, J Gauthier 1992 Phylogenetic taxonomy. *Annu Rev Ecol Syst* 23:449–480.
- 1994 Toward a phylogenetic system of biological nomenclature. *Trends Ecol Evol* 9:27–31.
- Donoghue MJ, RG Olmstead, JF Smith, JD Palmer 1992 Phylogenetic relationships of Dipsacales based on rbcL sequences. *Ann Mo Bot Gard* 79:333–345.
- Eriksson T 1999 AutoDecay, version 4.0.2. Bergius Foundation, Royal Swedish Academy of Sciences, Stockholm.
- Eriksson T, MJ Donoghue 1995 Is *Potentilla* monophyletic? a phylogenetic analysis of the Rosoideae (Rosaceae) using nuclear ribosomal ITS sequences. *Am J Bot* 82(suppl):128 (abstr 366).
- Eriksson T, MJ Donoghue, MS Hibbs 1998 Phylogenetic analysis of *Potentilla* using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of Rosoideae (Rosaceae). *Plant Syst Evol* 211:155–179.
- Felsenstein J 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.
- Focke WO 1894 Rosaceae. Pages 1–60 in A Engler, ed. *Die natürlichen Pflanzenfamilien*. Vol 3, abt 3. Engelmann, Leipzig.
- Greuter W, J McNeill, FR Barrie, HM Burdet, V Demoulin, TS Filguieras, DH Nicolson, et al, eds 2000 *International code of botanical nomenclature*. Koeltz, Königstein.
- Helfgott DM, J Francisco-Ortega, A Santos-Guerra, RK Jansen, BB Simpson 2000 Biogeography and breeding system evolution of the woody *Bencomia* alliance (Rosaceae) in Macaronesia based on ITS sequence data. *Syst Bot* 25:82–97.
- Hillis DM, JJ Bull 1993 An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* 42:182–192.
- Huelsenbeck JP, F Ronquist 2001a MrBayes: bayesian inference of phylogenetic trees. *Biometrics* 17:754–755.
- 2001b MrBayes: a program for the Bayesian inference of phylogeny, version 2.01. Computer program distributed by the authors, Rochester, N.Y., and Uppsala.
- Hutchinson J 1964 *The genera of flowering plants*. Oxford University Press, Oxford.
- Kluge AG, JS Farris 1969 Quantitative phyletics and the evolution of Anurans. *Syst Zool* 18:1–32.
- Kurtto A, T Eriksson 2003 Generic delimitation and nomenclatural adjustments in *Potentilla* (Rosaceae). *Ann Bot Fenn* (in press).
- Lewis PO 2001 Phylogenetic systematics turns over a new leaf. *Trends Ecol Evol* 16:30–37.
- Lindenhof A, A Weber 2000 Structural and developmental diversity of the androecium of Rosoideae (Rosaceae). *Bot Jahrb Syst* 122: 63–91.
- Mason-Gamer RJ 2001 Origin of North American *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. *Syst Bot* 26:757–768.
- Morgan DR, DE Soltis, KR Robertson 1994 Systematic and evolutionary implications of rbcL sequence variation in Rosaceae. *Am J Bot* 81:890–903.
- Posada D, KA Crandall 1998 Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Potter D, F Gao, PE Bortiri, S-H Oh, S Baggett 2002 Phylogenetic relationships in Rosaceae inferred from chloroplast matK and trnL-trnF nucleotide sequence data. *Plant Syst Evol* 231:77–89.
- Rambaut A 1996 Se-Al: sequence alignment editor, version 1.0a1. Department of Zoology, Oxford University, Oxford.
- Rannala B, JP Huelsenbeck, Z Yang, R Nielsen 1998 Taxon sampling and accuracy of large phylogenies. *Syst Biol* 47:702–710.
- Reveal JL 2000 Index nominum supragenericorum plantarum vascularium. University of Maryland. [http://www.inform.umd.edu/EdRes/Colleges/LFSC/life\\_sciences/plant\\_biology/WWW/supragen.html](http://www.inform.umd.edu/EdRes/Colleges/LFSC/life_sciences/plant_biology/WWW/supragen.html).
- Robertson KR 1974 The genera of Rosaceae in the southeastern United States. *J Arnold Arbor Harv Univ* 55:303–332, 344–401, 611–662.
- Rodríguez F, JL Oliver, A Marín, JR Medina 1990 The general stochastic model of nucleotide substitution. *J Theor Biol* 142:485–501.
- Rydberg PA 1898 A monograph of the North American Potentilleae. *Mem Dep Bot Columbia College* 2:1–223, pls. 1–112.
- Rydin C, M Källersjö 2002 Taxon sampling and seed plant phylogeny. *Cladistics* 18:485–513.
- Sanderson MJ, MJ Donoghue 1996 The relationship between homoplasy and confidence in a phylogenetic tree. Pages 67–89 in MJ Sanderson, L Hufford, eds. *Homoplasy: the recurrence of similarity in evolution*. Academic Press, San Diego, Calif.
- Schulze-Menz GK 1964 Rosales. Pages 193–242 in H Melchior, ed. *A. Engler's Syllabus der Pflanzenfamilien*. Vol 2. Borntraeger, Berlin.
- Smedmark JEE, T Eriksson 2002 Phylogenetic relationships of *Geum* (Rosaceae) and relatives inferred from the nrITS and trnL-trnF regions. *Syst Bot* 27:303–317.
- Smedmark JEE, T Eriksson, RC Evans, CS Campbell 2003 Ancient allopolyploid speciation in Geinae (Rosaceae): evidence from nuclear granule-bound starch synthase (GBSS1) gene sequences. *Syst Biol* (in press).
- Soják J 1986 Notes on *Potentilla*. I. Hybridogeneous species derived from intersectional hybrids of sect. Niveae × sect. Multifidae. *Bot Jahrb Syst* 106:145–210.
- 1994 Notes on *Potentilla* (Rosaceae) X–XII. X. The section Dumosae. XI. The *P. microphylla* and *P. stenophylla* groups (sect. Pentaphylloides). XII. Key to the taxa of *P.* sect. Pentaphylloides (Anserina). *Bot Jahrb Syst* 116:11–81.
- Staden R 1996 The Staden sequence analysis package. *Mol Biotechnol* 5:233–241.
- Swensen SM, BC Mullin 1997 The impact of molecular systematics on hypotheses for the evolution of root nodule symbioses and im-

- plications for expanding symbioses to new host plant genera. *Plant Soil* 194:185–192.
- Swofford DL 2001 PAUP\*: phylogenetic analysis using parsimony and other methods, version 4.0. Sinauer, Sunderland, Mass.
- Taberlet P, L Gielly, G Pautou, J Bouvet 1991 Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol Biol* 17:1105–1109.
- Urbatsch LE, B Baldwin, MJ Donoghue 2000 Phylogeny of coneflowers and relatives (Heliantheae: Asteraceae) based on nuclear rDNA internal transcribed spacer (ITS) sequences and chloroplast DNA restriction site data. *Syst Bot* 25:539–565.
- Vretblad M, T Eriksson, MJ Donoghue 1996 A molecular and morphological phylogenetic analysis of the subfamily Rosoideae (Rosaceae), with emphasis on the monophyly of the genus *Potentilla*. Poster presented at University of Arizona, Tucson. <http://eebweb.arizona.edu/RTG/ABSTRAX2.HTM#Vretblad>.
- Weimarck H 1934 Monograph of the genus *Cliffortia*. Gleerupska Universitetsbokhandeln, Lund.
- Wendel JF, A Schnabel, T Seelanan 1995 Bi-directional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc Natl Acad Sci USA* 92:280–284.
- White TJ, T Bruns, S Lee, J Taylor 1990 Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315–322 in MA Innis, DH Gelfand, JJ Sninsky, TJ White, eds. *PCR protocols: a guide to methods and applications*. Academic Press, London.
- Wikström N, P Kenrick, MW Chase 1999 Epiphytism and terrestrialization in tropical *Huperzia* (Lycopodiaceae). *Plant Syst Evol* 218:221–243.
- Wolf T 1908 Monographie der Gattung *Potentilla*. *Bibl Bot* 16 (Heft 71):1–714.
- Yang Z 1996 Among-site variation and its impact on phylogenetic analyses. *Trends Ecol Evol* 11:367–371.
- Yang Z, N Goldman, A Friday 1994 Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Mol Biol Evol* 11:316–324.
- Yurtsev BA 1993 New taxa of the genus *Potentilla* (Rosaceae) from the Arctic Alaska. *Bot Zh* 78:78–87.