Intergeneric relationships within the subfamily Rosoideae (Rosaceae) with emphasis on *Rubus* L.

Rubus chamaemorus (Cloudberry) ©Torsten Eriksson

Ardian Høgøy Abaz

Master of Science in Biology

Biodiversity, Evolution and Ecology

Department of Biology

University of Bergen

June 2018

Abstract

The genus *Rubus* belongs to the Rose family, and is currently placed in the supertribe Rosodae, of the subfamily Rosoideae. Further systematic placement of *Rubus* has not yet been clarified, although several hypotheses have been suggested. Several studies show that different datasets based on chloroplast and ribosomal nuclear genes were incongruent with respect to the placement of *Rubus*. To investigate the relationships between *Rubus* and other genera within Rosoideae, phylogenies of the chloroplast gene *matK*, the nuclear ribosomal ITS gene and four single copy genes (GAPCP1, GBSSI-1, GBSSI-2 and SbeI) were inferred on specimens from the following tribes: Colurieae, Potentilleae, Roseae, Rubeae, Sanguisorbeae and Ulmarieae. Our ITS data results are congruent with previous chloroplast data, in which Colurieae is the sister to the rest of Rosodae including *Rubus*. However, according to low copy nuclear genes analyzed here, Rubeae and Colurieae form a clade separate from rest of Rosodae. This leads us to believe that the incongruence may be caused by incomplete lineage sorting.

Keywords: incomplete lineage sorting, low copy nuclear marker, phylogenetic discordance, Rosodae

Preface

This master project is a collaboration with the Rosaceae evolution Research Group at the University Museum of Bergen. The research group currently consists of four senior researchers (Torsten Eriksson, Heidi Lie Andersen, Jenny Smedmark and Per Harald Salvesen), two graduate students (Hoda Parsian and Nannie Persson), and two master students (Ardian Høgøy Abaz and Ingrid Toresen). The research group works on the phylogeny of Rosaceae, Rosoideae and subgroups.

This thesis has been written with the intention to be published in *Journal of Systematics and Evolution*. Therefore, this thesis has been written in an article format as readily as possible with only minor edits before being submitted.

Also, some practical issues that gave no results for certain genes will be addressed in this section.

During this project, several PCR protocols were tested and run in order to yield as many results as possible. The chloroplast gene *matK* was the first gene used as a part of a lab trial, which yielded PCR products for all of the plant extractions. The protocol did not work for any of the other genes, despite changing many of the parameters, one at a time. Two different recipes were therefore used. Many primers failed to yield any products in trial PCR, such as the low copy genes DHAR, GDSL1, Leafy2int2, PEPC, TPP2 and WD. As these genes failed during amplification, they were excluded from further consideration due to time constraints. For the same reason, GAPCP1, GBSSI-1 and GBSSI-2 only yielded products for some of the samples. As none of the low copy genes gave highly supported topologies, concatenation was essentially used.

Several analyses based on multispecies coalescent model, BEAST v.1.8.0 (Drummond et al., 2012; Heled & Drummond, 2010) were also run, but ultimately failed to yield any strongly supported topologies, and the results were not included.

1. Introduction

Rubus L. is a diverse genus of ca. 250-700 species, consisting of perennial shrubs and herbs, including raspberries and brambles. The genus belongs to Rosaceae which consists of about 90 genera and 3000 species distributed worldwide with a higher concentration in the northern temperate regions (Simpson, 2010). The family is classified into three subfamilies (Potter et al., 2007); Rosoideae, which includes *Rubus*, Dryadoideae and Spireaoideae. The Rosoideae subfamily currently consists of the genus *Filipendula* (the monogeneric Ulmarieae tribe) and the supertribe Rosodae. Rosodae is further divided into *Rubus*, the tribe Colurieae, the genus *Rosa*, and the two tribes Potentilleae and Sanguisorbeae according to Potter et al. (2007). *Rosa* belongs to the monogeneric Roseae tribe (Hutchinson, 1964; Kalkman, 2004; Xiang et al., 2017; Zhang et al., 2017). The three tribes Roseae, Potentilleae and Sanguisorbeae are together referred to as the Roperculina clade (Eriksson et al., 2003).

Early molecular studies of *Rubus* classified the genus in its own monogeneric tribe called Rubeae (Hutchinson, 1964; Kalkman, 2004) which is commonly referred to in later studies as well (Zhang et al., 2017, Xiang et al., 2017). However, previously published analyses distinctly disagree on the relative positions of Rubeae and Colurieae (Potter et al., 2007, Xiang et al., 2017, Zhang et al., 2017).

Morphologically, *Rubus* and the other members of Rosoideae separates from the Spiraeoidae subfamily by having indehiscent fruits, and significantly different number of stamens, in addition to a lower base chromosome number, and a lack of flavonols and sorbitol (Robertson, 1974). The morphological traits which separates *Rubus* from the other members of Rosoideae are two ovules instead of one, and the fruit type which is an aggregation of drupelets called drupecetum (Spies & Du Plessis, 1985; Kalkman, 2004; Spjut, 1994). An early molecular study of Rosaceae resolved Rosoideae as a monophyletic group with members that have base chromosome numbers of x = 7 and 8 (Morgan et al., 1994). Chromosomal counts show that the base chromosome number for *Rubus* is 7. While many species within *Rubus* are diploid (2n = 14), polyploids are reported, up to tetradecaploid e.g. *Rubus lorentzianus* (2n = 14x = 98) (Thompson 1997).

In more recent years, the systematics of *Rubus* have been complicated on the molecular level. Based on the chloroplast gene *rbcL*, Morgan et al. (1994) resolved *Filipendula* as the sister to all the other

Rosoideae. Then, the clade now known as Colurieae is the sister to the rest, including *Rubus*, but this was weakly supported (2 substitutions, decay value = 1).

Eriksson et al. (2003) discussed *Rubus*' placement in Rosoideae as problematic, because different datasets based on chloroplast and nuclear ribosomal genes were incongruent. Potter et al. (2007) included six nuclear (18S, GBSSI-1, GBSSI-2, ITS, pgip, and ppo) and four chloroplast (*matK*, *ndhF*, *rbcL*, and *trnL-trnF*) regions, separately and in various combinations, with parsimony and likelihood-based Bayesian approaches (Fig. 1A). They resolved *Filipendula* as the sister to the rest of Rosoideae like Morgan et al. (1994). Yet, *Rubus* was suggested to be the sister group to the rest, and then Colurieae with low bootstrap support (0.42), meaning its phylogenetic relationship is still ambiguous. The relative positions of Colurieae and *Rubus* were not resolved by either of these two studies.

Later studies on Rosoideae have highlighted how incongruent the nuclear and chloroplast gene datasets can be. A whole plastome dataset by Zhang et al. (2017) hypothesize that the Colurieae tribe is the sister clade to the remaining Rosodae members (Fig. 1B). They also classify Rosoideae into six tribes: Ulmarieae, Colurieae, Rubeae, Agrimonieae (which will be called Sanguisorbeae in this study, following Potter et al., 2007), Roseae and Potentilleae. Another recent study based on a filtered set of several hundred nuclear genes by Xiang et al. (2017) resolves the first branches closest to the root to be Ulmarieae, and then Rubeae (Fig. 1C), and then Colurieae, like Potter et al. (2007). Both of these latter conflicting results are well supported.

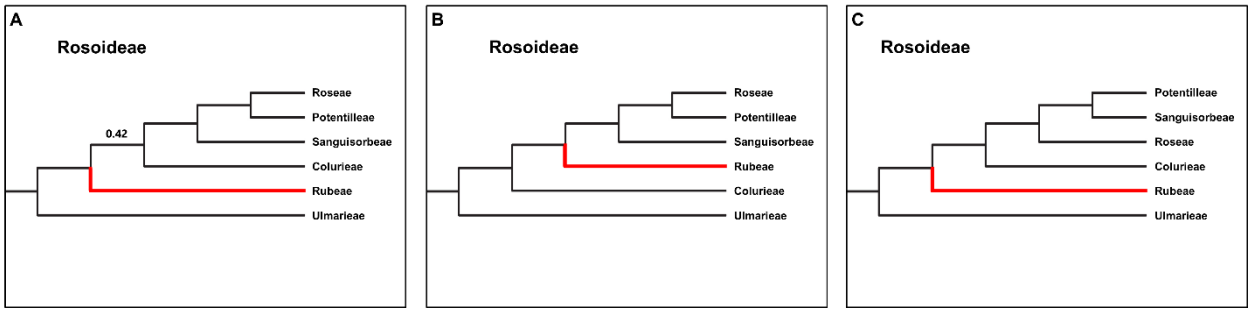

Fig. 1. Previous phylogenetic reconstructions of Rosoideae based on different datasets. The red line shows the branching of Rubeae including *Rubus*. (A) Potter et al., 2007 based on six nuclear (18S, gbssi1, gbssi2, ITS, pgip, and ppo) and four chloroplast (matK, ndhF, rbcL, and trnL-trnF) regions. (B) Zhang et al. (2017) based on whole plastome (WP) data set. (C) Xiang et al. (2017) based on coalescence analyses of up to several hundred nuclear genes.

While nuclear DNA is transmitted through meiosis and undergoes recombination, the chloroplasts are inherited uniparentally, resulting in nuclear reflecting the evolutionary history with higher accuracy than the cpDNA. It is, however, interesting to compare results of both nuclear and chloroplast data in order to elucidate the species' evolutionary history with more certainty.

Analyses of ribosomal nuclear genes may not reconstruct allopolyploid lineages properly if the DNA sequences become homogenized to either parental genotype through concerted evolution (Sang, 2002; Smedmark et al., 2003). Low copy nuclear genes, unlike ribosomal genes, are not as prone to concerted evolution, and can be used to reconstruct the phylogenetic relationships of parental lineages of hybrid species (Smedmark et al., 2003).

Single copy genes previously used on taxa within Rosaceae have a higher chance of success when amplifying Rosoideae than other families. Low copy genes have previously been utilized as valuable markers at the intergeneric level. The granule bound starch synthase (GBSSI) gene have been utilized by Evans et al. (2000), in resolving the position of the subfamily Maloideae within Rosaceae. The GBSSI gene was shown to have two loci within a diploid genome of the Rosaceae species, GBBSI-1 and GBSSI-2 (Evans et al., 2000). The same marker was used by Smedmark et al. (2003) within Colurieae. Rousseau-Gueutin et al. (2009) inferred a part of the GBSSI-2 locus on the Fragariinae subtribe within the Potentilleae tribe. The Glyceraldehyde-3-phosphate dehydrogenase (GAPCP1) has also been inferred in *Prunus* within Rosaceae and has been constructed as a valuable molecular marker for single nucleotide polymorphisms (Le Dantec et al., 2010). The Starch branching enzyme I (SbeI) gene is another single copy marker, which have been used to investigate the phylogenies in *Prunus* (Shi et al., 2013).

The objective of this study is to clarify the uncertain basal relationships between *Rubus* and other well defined genera within the Rosoideae subfamily, based on analyses of plastid, ribosomal and low copy nuclear genes.

2. Materials and Methods

2.1 Taxon sampling

Species representing the tribes within Rosoideae were selected based on two criteria; species considered to be diploid representatives were chosen as the polyploid taxa have arisen from these and will therefore be optimal for depicting the origin of the clades. In addition, species considered to be basal within the tribes were chosen to avoid Long Branch Attraction (LBA) (Bergsten, 2005). Also, the polyploid taxa require more cloning in order to find all the separate copies of a gene. Samples were collected from the wild or from wild-collected samples in the Arboretum and Botanical Gardens of Bergen, others were obtained from other botanical gardens (Appendix I). Voucher specimens was deposited in herbarium BG. For DNA extraction, leaves were dried in silica gel.

2.2 DNA Extraction

Small fragments of leaves (ca. 20 mg per species) were placed in separate Eppendorf tubes and homogenized. The DNA extraction followed the manual protocols of Qiagen DNeasy® Mini kit (Qiagen, Germantown, Maryland, USA), with a couple of deviations from the protocol. Firstly, the samples were incubated overnight at 59°C, and then incubated for 10 minutes at 65°C. Secondly, the samples were centrifuged for 2 minutes instead of 5, or until the filter was dry. An additional minute of spinning was added in the end just to ensure that no ethanol would be carried over during elution.

2.3 Molecular markers

Six DNA regions were included (Table 1): the chloroplast *matK* gene, the nuclear internal transcribed spacer of the ribosomal DNA (ITS1, 5.8S, and ITS2; ITS), the nuclear low copy Glyceraldehyde-3-phosphate dehydrogenase (GAPCP1), the nuclear low copy granule bound starch synthase I genes (GBSSI-1 and GBSSI-2) and the nuclear low copy starch branching enzyme I gene (SbeI).

Table 1. Primers used for sequencing.

DNA region	Primer name	Sequence (5'-3')	Reference
ITS	ITS1	TCC GTA GGT GAA CCT GCG G	White et al., 1990
	ITS4	TCC TCC GCT TAT TGA TAT GC	
GAPCP1	CGPPB5575 FWD	CAT GTG CTC TAT GAG GTC CA	Le Dantec et al., 2010
	CGPPB5575 REV	ATC AGG TAT GCT GCT GAT GG	
GBSSI-1	3F	TAC AAA CGA GGG GTT GAT CG	Evans et al., 2000
	7R	CCT TGG TAA GCA ATG TTG TG	
GBSSI-2	F2	TGG TCT TGG TGA TGT TCT TGG	Rousseau-Gueutin et al., 2009
	R2	GTG TAG TTG GTT GTC CTT GTA ATC C	
SbeI	SbeI F	GCT CCA CGA ATA TAT GAG GCA CAT G	Shi et al., 2013
	SbeI R	TTC CAT GAA ATT TCC TTC ATT GAC CA	
matK	*matK*472F	CCR TYC ATG GAA ATC TTG GTT C	Yu et al., 2011
	*matK*1248R	GCT RTA TAA TGA GAA AGA TTT CTG C	

2.4 Amplification and sequencing

Amplification was carried out by polymerase chain reaction (PCR), performed on a C1000 touch thermal cycler (Bio-Rad Laboratories Inc.).

The *matK* reaction mixtures contained 1 µl total DNA, Takara 1X buffer without $MgCl_2$, 1.5 mM $MgCl_2$, 1 µl 1.10 mg/mL dimethyl sulfoxide (DMSO) and 0.25 µl 10 mg/mL bovine serum albumin (BSA) enhancers, 0.8 mM dNTPs, 1 mM of each primer in the primer pair (Table 1), 2U AmpliTaq GOLD® and ddH_2O to a final volume of 25 µl.

For the nuclear genes, the reactions consisted of 2 µl total DNA, Takara 1X buffer without $MgCl_2$, 3 mM $MgCl_2$, 0.8 mM dNTPs, 0.8 mM of each primer in the primer pair (Table 1), and 1.5U TaKaRa Ex Taq® Hot Start Version, and ddH_2O to a total of 25 µl. For samples that needed cloning, this very same PCR protocol was used.

The thermal PCR cycling parameters can be found in Table 2. To ensure that no contaminations would be the cause of errors, a positive and a negative control were included for each run. For

GBSSI-1, GBSSI-2 and SbeI, the PCR programs with one annealing temperature failed to yield products in some of the extractions, and therefore touchdown PCRs were performed instead.

Table 2. PCR protocols. Note that the upper and lower annealing temperatures of the PCR programs are both shown for GBSSI-1, GBSSI-2 and SbeI.

Primer pair	Initialization temperature [°C], time [s]	Denaturation temperature [°C], time [s]	Annealing temperature [°C], time [s]	Extension temperature [°C], time [s]	Number of cycles
matK	95, 600	95, 30	52, 40	72, 60	39
ITS	94, 180	95, 30	52, 30	72, 80	35
GAPCP1	94, 180	95, 30	52, 30	72, 80	35
GBSSI-1	94, 180	94, 45	60/48, 30	72, 60	13, 36^F
GBSSI-2	94, 180	94, 45	60/48, 30	72, 60	13, 36^F
SbeI	94, 180	94, 45	62/49, 30	72, 60	15, 36^F

^{F} = Number of cycles for the final extensions.

The quality and amount of DNA of the PCR products were assessed using gel electrophoresis. One microliter 5X Ficoll loading dye was mixed with 4 µl PCR product. The PCR product was run on a 30 µl 1% agarose gel mixed with 1 µl GelRed (Biotium Inc.) staining agent. To measure the quantity of the DNA, FastRuler Middle Range DNA Ladder (ThermoFisher Scientific) markers were used. The gel was then visualized in a UV cabinet (Syngene). Snapshots and manual-band quantifications were performed using GeneSnap plugin in Syngene.

PCR products were purified using EXOSAP-IT protocol (GE Healthcare). Eight microliters PCR product was mixed with 0.1 µl Exonuclease I (EXO), 1 µl Shrimp Alkaline Phosphatase (SAP) and 0.9 ddH_2O, incubated at 37°C for 30 minutes, followed by enzyme inactivation at 85°C for 15 minutes.

To prepare for sequence reactions, the BigDye Terminator v.3.1 cycle sequencing Kit (Applied Biosystems) protocol was followed. The amount of DNA and water was determined by the DNA concentration calculated from the manual band quantifications. The samples were run in the thermal cycler starting with 96°C for 5 minutes followed by 50°C annealing temperature for 5 seconds. After the reactions, 10 µl ddH_2O was added for each sample, adding up to a total volume of 20 µl in the sample, so that the final PCR products could be sent in to the Sequencing Facility at the University of Bergen.

Some of the expected diploids or samples of unknown ploidy level showed double bands in either PCR products or DNA sequences. In order to successfully sequence these samples, molecular cloning using the StrataClone PCR Cloning Kit was performed. If the sample had weak bands from the gel electrophoresis, 2 µl cloning reaction were used instead of 1. The number of clones that were randomly picked depends on the ploidy level of the sample, as higher ploidy level requires more colonies in order to secure all of the different copies of the gene. An additional PCR program was run using the M13F and M13R-Puc(-40) primers (Macrogen) for the cloning vector. The program included 94°C initial denaturation for 10 minutes, then a cycle of denaturation at 94°C for 45 seconds, an annealing temperature of 55°C for 45 seconds, and finally an extension at 72°C for 3 minutes. This cycle was repeated 34 times.

2.5 Alignment and phylogenetic analyses

Geneious v.8.1.9 (Kearse et al., 2012) was used to edit and assemble the DNA sequences. A BLAST Search in GenBank was performed to see whether any samples were contaminated. To extend the dataset, relevant sequences were downloaded from GenBank. Accession numbers for the downloaded sequences and those newly generated sequences in this study are found in Appendix II. DNA sequences were aligned with MAFFT v7.017 (Katoh et al., 2002), using an automatic algorithm with a scoring matrix of 200PAM / k=2, and the gap open penalty set to 1.53 with an offset value of 0.123. For the individual gene trees, the polyploid clone sequences were initially aligned to check whether recombinants could be detected and excluded from further analyses. Heuristic search of the datasets were performed using PAUP* 4.0a161 (Swofford; 2002). The ITS alignment was separated into three partitions: ITS1, 5.8S and ITS2 using annotations from reference sequence in GenBank. The GAPCP1, GBSSI-1 and GBSSI-2 alignments were separated into exons and introns. The *matK* and SbeI alignments only consist of one exon.

MrAIC (Nylander, 2004), which uses PHYML as backend Maximum Likelihood program (Guindon & Gascuel, 2003), was used to infer the most suitable nucleotide substitution models for each individual gene using the AICc criterion (Sugiura, 1978; Hurvich & Tsai, 1989). MrBayes v3.3.5 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) was used to create phylogenetic trees based on Bayesian inference, under a mixed model for the individual gene trees and the models suggested by MrAIC for the concatenated trees. For the combined dataset, the

number of generations was set to 200 million with a sample frequency of every 1000th tree. Trees based on (1) each individual gene, (2) nuclear genes and (3) low copy nuclear genes (all nuclear excluding ITS) were created. The trees were rooted on the branch to *Filipendula*. When making concatenated trees, the clone sequences with the shortest branch lengths were chosen in order to avoid LBA.

Additionally, gaps were manually coded using the method "simple indel coding" from Simmons and Ochoterena (2000) and were included as separate characters using the Mk model (Lewis, 2001). When evaluating the results of the Bayesian analyses, four criteria had to be met in order to accept an analysis: the standard deviation of split frequencies should be below 0.01, the chain swap should be between 20 and 80% (McGuire et al., 2007), no trend should be seen in the overlay plot and the Potential Scale Reduction Factor (PSRF; Gelman & Rubin, 1992) values should reach 1.0 for all parameters. Tracer v.1.7 (Rambaut et al., 2018) was used to visualize and analyze the Markov Chain Monte Carlo trace files. The final trees were visualized in Mesquite v.3.4.0 (Maddison & Maddison, 2018).

3. Results

3.1 Sequence assembly and alignment

A total of 87 sequences representing 23 specimens were successfully included in the concatenated dataset of all genes. The datasets for the nuclear, low copy genes, and the total combined dataset consisted of 3526, 2829 and 4259 characters, respectively. The majority of the indels in all of the datasets but SbeI were autapomorphic. Among the low copy nuclear genes, the SbeI was the most complete dataset which included all the sampled taxa and had no gaps and ambiguities in the alignment. The number of characters for the datasets of the individual gene trees can be found in Table 3.

The alignment for *matK* revealed two overlapping indels; one indel by *Fallugia paradoxa*, *Geum aleppicum*, *Geum rivale* and *Geum urbanum* in positions 274-297 and one indel by *Fragaria vesca*, *Potentilla micrantha* and *Potentilla reptans* in positions 287-295. Sixteen of the parsimony-informative characters for ITS were indels, ranging from 1 to 13 bases. The majority of the indels for ITS were caused by *Agrimonia eupatoria*. The Roseae species in the GAPCP1 dataset shared

an indel of 18 bases and an indel of 82 bases with an exception of *Rosa majalis*. The GBSSI-1 dataset had several indels among Sanguisorbeae and Potentilleae species.

Table 3. Characteristics of the individual molecular datasets included in this study.

Gene	Number of taxa	Number of characters in total	Number of constant characters	Number of parsimony-informative characters
matK	21	734	531	136
ITS	19	697	425	172
GAPCP1	16	609	301	150
GBSSI-1	10	857	463	167
GBSSI-2	8	657	366	98
SbeI	17	706	545	77

3.2 Bayesian analyses

The Generalized time-reversible model with gamma distribution for among-site variation was selected for *matK*, ITS1, ITS2 and the GBSSI-1 introns (GTR + G; Tavaré 1986, Yang 1996). For SbeI, the same model was used but with the incorporation of the proportion of invariable sites (GTR+I+G; Shoemaker & Fitch, 1989). As for the Hasegawa-Kishino-Yano nucleotide substitution model (HKY; Hasegawa et al., 1985), HKY+I was chosen for the GBSSI-1 exons, HKY+G was chosen for the GBSSI-2 exons and introns and HKY+I+G was chosen for the GAPCP1 introns. The Kimura 2-parameter model (K2P; Kimura, 1980), K2P+I was chosen for 5.8S, and K2P+G was chosen for the GAPCP1 exons.

The individual and combined analyses resolve different groups as closest relatives to Roperculina. The topologies of the low copy gene trees generally have low support, but four 50% majority-rule consensus trees with significant branch support were generated for the following datasets: ITS, a concatenated four low-copy genes GAPCP1, GBSSI-1, GBSSI-2 and SbeI dataset, a five nuclear genes concatenated dataset and the all data concatenated dataset.

The generated trees from the Bayesian analyses with six separate clades that represent recognized tribes are shown in Figs. 2-6. In the *matK* tree, the basal node in Rosodae have three unresolved

groups of Colurieae, Rubeae and Roperculina (Roseae/Potentilleae/Sanguisorbeae) (Fig 2.). Potentilleae is the sister to Roseae and Sanguisorbeae with low support (0.84).

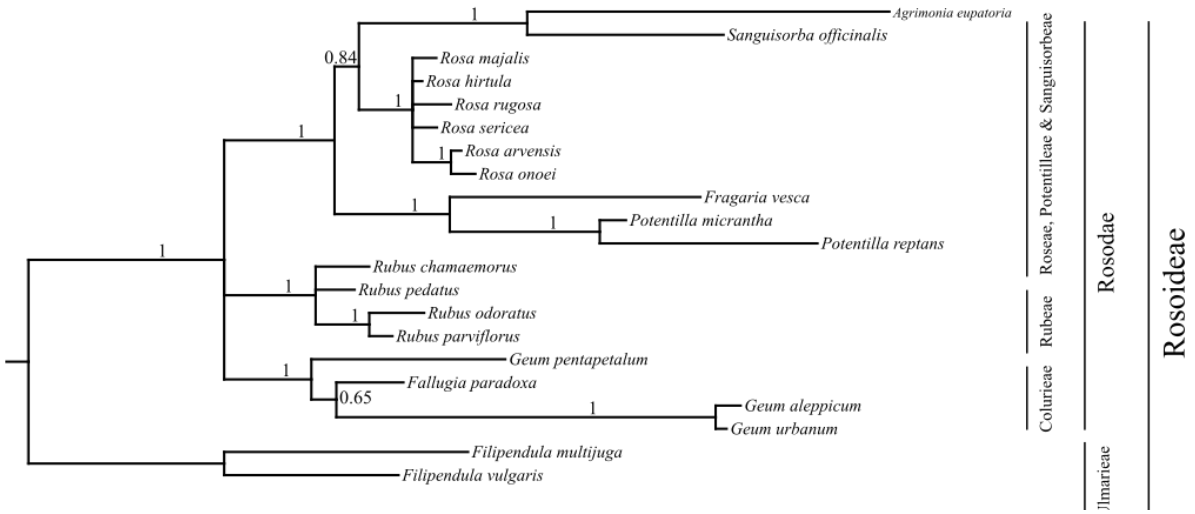

Fig. 2. Bayesian 50% majority-rule consensus tree based on analysis with a mixed model for the chloroplast *matK* region. Numbers at nodes indicate posterior probabilities of clades.

The tree resulting from ITS (Fig. 3) shows a strongly supported sister relationship between Colurieae and the rest of Rosodae. Rubeae is the following group, sister to Roperculina. Within Roperculina, Potentilleae and Sanguisorbeae are suggested to belong to a common clade whit Roseae as the sister with low support (0.8).

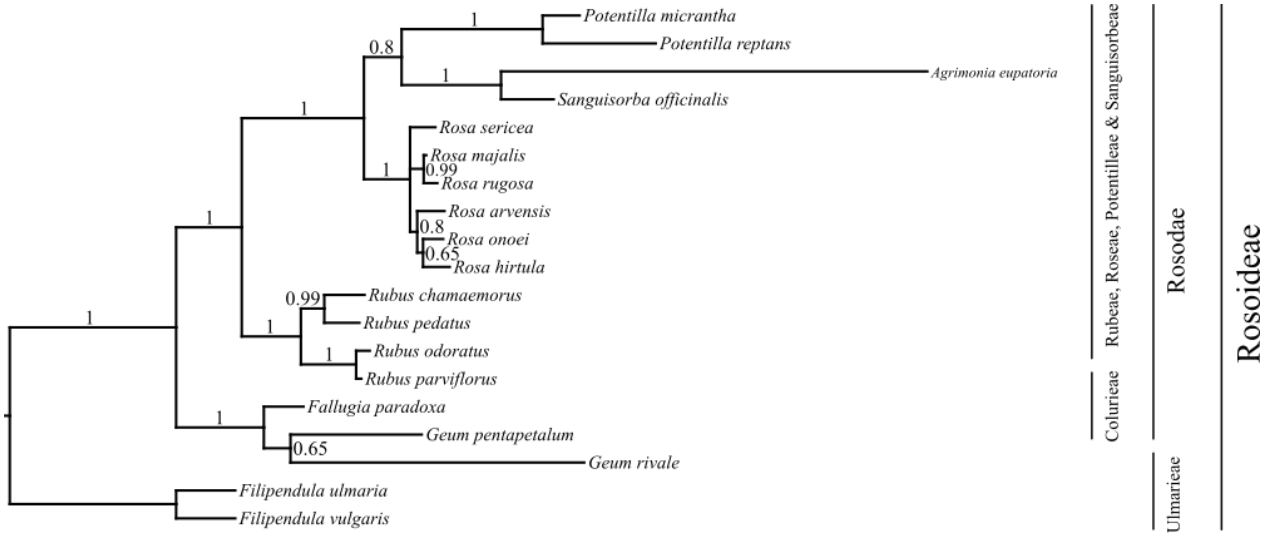

Fig. 3. Bayesian 50% majority-rule consensus tree based on analysis with a mixed model for the nuclear ribosomal ITS region. Numbers at nodes indicate posterior probabilities of clades.

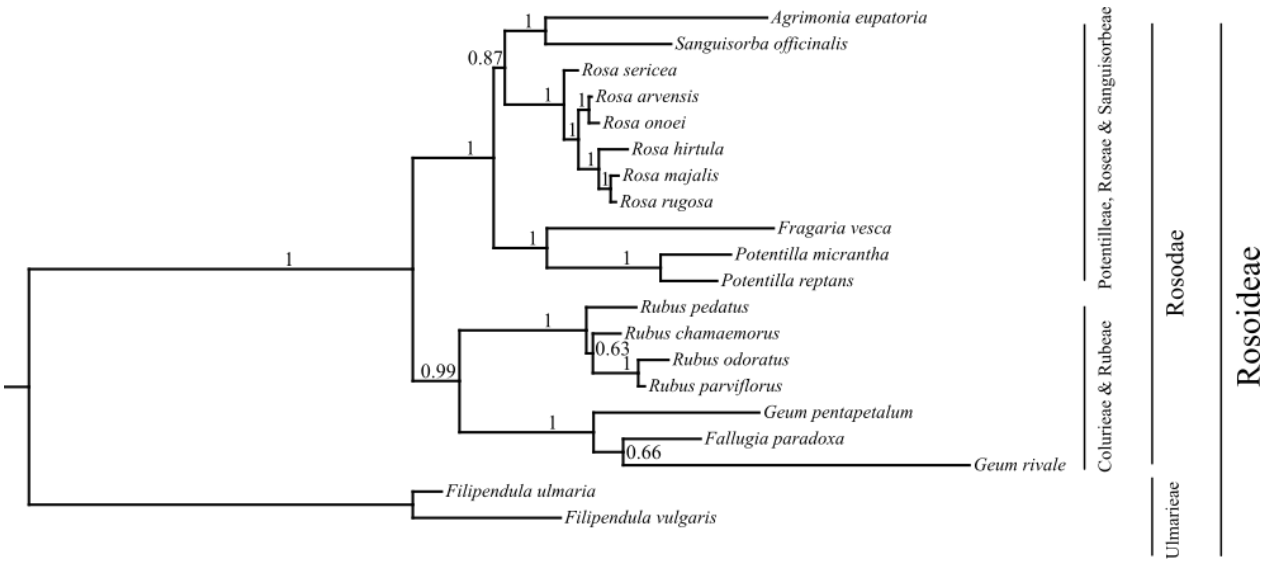

Fig. 4. Bayesian 50% majority-rule consensus tree based on analysis of the nuclear low copy GAPCP1, GBSSI-1, GBSSI-2 and SbeI regions. Numbers at nodes indicate posterior probabilities of clades.

The tree generated from the four low copy nuclear regions (Fig. 4) strongly supports one clade including the Colurieae and Rubeae as the sister group to the remainder of the Rosodae. The tree analysis of five concatenated regions including ITS shows the same core topology as the ITS with high support (Fig. 5) but for the clade of Roseae and Sanguisorbeae which has weaker support (0.67). When all six genes are combined (Fig. 6), this same pattern is shown.

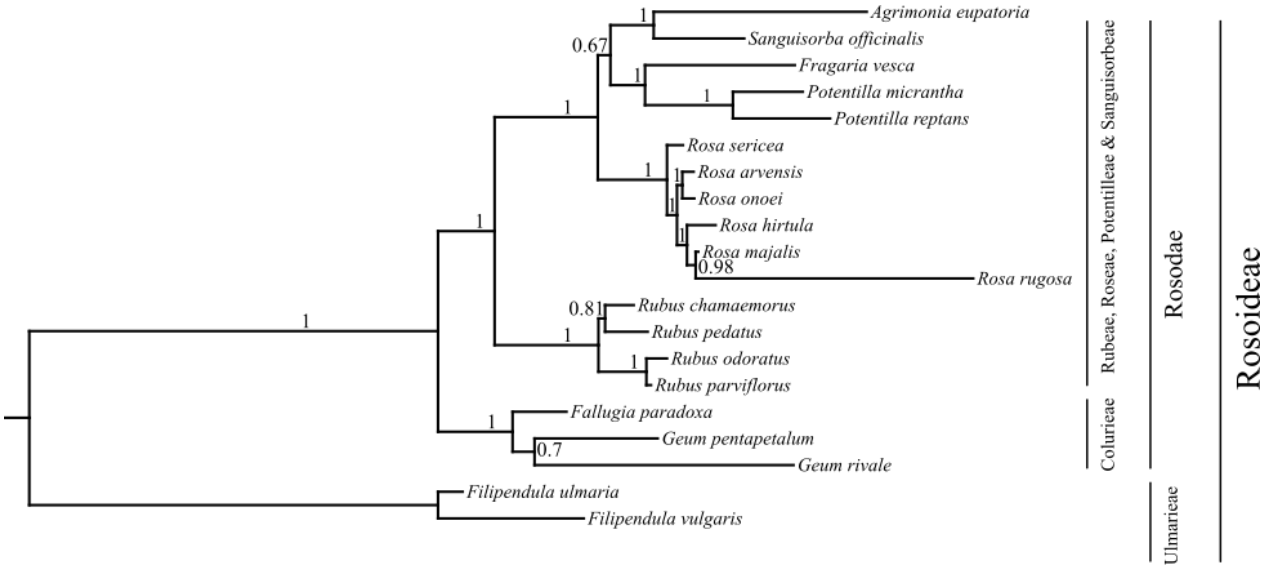

Fig. 5. Bayesian 50% majority-rule consensus tree based on analysis of the nuclear ITS, GAPCP1, GBSSI-1, GBSSI-2 and SbeI regions combined. Numbers at nodes indicate posterior probabilities of clades.

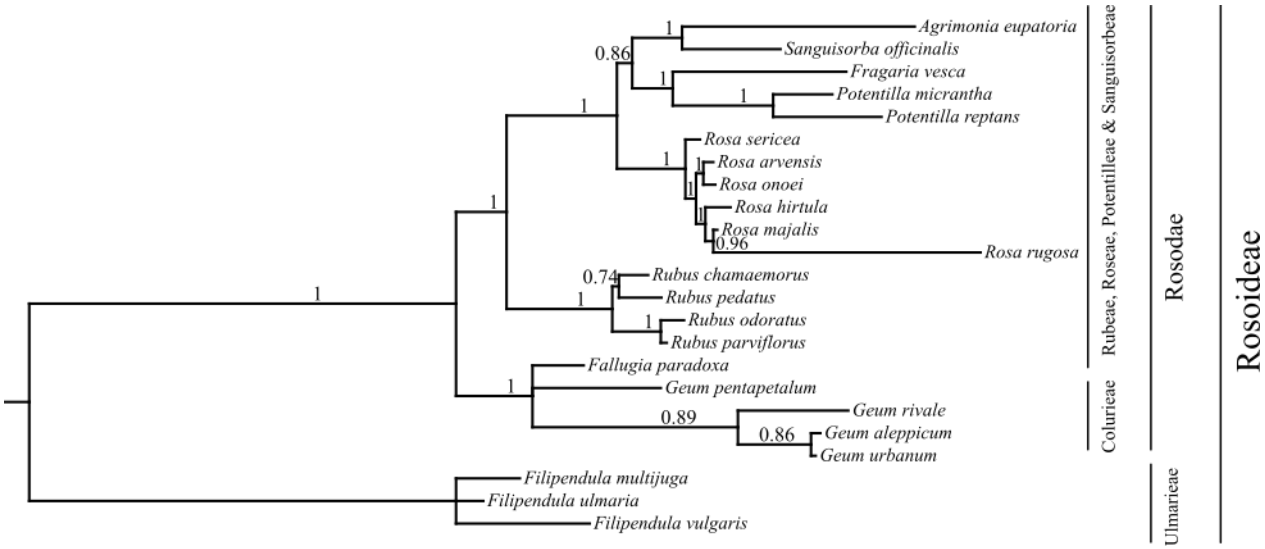

Fig. 6. Bayesian 50% majority-rule consensus tree based on analysis of the six *matK*, ITS, GAPCP1, GBSSI-1, GBSSI-2 and SbeI regions combined. Numbers at nodes indicate posterior probabilities of clades.

To summarize the results, there are two possible placements of Rubeae. It is either grouped together with Colurieae in a separate clade (low copy nuclear genes only), or sister to Roperculina alone to the exclusion of Colurieae (ITS and low copy genes plus ITS). None of these analyses suggest that Colurieae and Roperculina form a clade without Rubeae. All of the trees are in agreement with the placement of Ulmarieae, Rosodae and Roperculina.

4. Discussion

4.1 Possible placements of *Rubus*

Several phylogenetic studies have shown that Ulmarieae is included in Rosoideae but excluded from Rosodae in agreement with the results of this study (Eriksson et al., 1998; Eriksson et al., 2003; Morgan et al., 1994; Potter et al., 2007; Xiang et al., 2017; Zhang et al., 2017). The present study also further support Roperculina as a clade but the positions of Rubeae and Colurieae are ambiguous. There are three possible hypotheses regarding the relationships of the three clades Colurieae, Roperculina and Rubeae.

Hypothesis (H1): Colurieae is sister to the rest of the Rosodae including Rubus

The first hypothesis places Colurieae as the sister to the rest of Rosodae. This hypothesis was strongly supported by our ITS data, but with low support by our SbeI and GBSSI-1(0.53 and 0.55, respectively). The ITS sequences were easy to align compared to the other DNA regions in this study. The first ITS study of Rosoideae pointed out how their analysis containing distantly related genera could cause several indels (Eriksson et al., 1998). The MAFFT alignment for ITS in this study also resulted in several indels but the majority of them had a length of one base pair. Internal transcribed spacers are not feasible for distantly related genera because of difficulties with alignments due to their rapid divergence rates (Álvarez & Wendel, 2003; Sang, 2002), but it seems from this result that ITS has a divergence rate relevant to resolving the Rosoideae.

Only two studies of Rosoideae have used ITS exclusively and included *Rubus* in their datasets (Eriksson et al, 1998; Smedmark & Eriksson, 2002). The ITS study by Eriksson et al. (1998) does not support this hypothesis (see H3), but the ITS tree is congruent with Smedmark & Eriksson (2002), who focused mainly on Colurieae, and placed *Rubus* in the same clade as Roseae and Sanguisorbeae with high support. No taxa representing the Potentilleae tribe were included in that study.

This hypothesis is consistent among plastid studies (Morgan et al., 1994; Eriksson et al., 2003; Zhang et al., 2017). The main problem with chloroplast data is that these genes are inherited uniparentally, having limitations when identifying reticulate evolution.

Hypothesis (H2): Rubeae is sister to the rest of the Rosodae including Colurieae

The second hypothesis implies that Rubeae is the sister to the rest of the Rosodae, including Colurieae. None of the analyses in this study supports this hypothesis but a recent study by Xiang et al. (2017) resulted in this topology using transcriptomic and genomic dataset from hundreds of nuclear genes.

Xiang et al. (2017) selectively chose their genes based on a filtering algorithm. Every tree created from a gene should conform to their constraint tree of chosen taxa in order to be included in their concatenated tree (Fig. 7.). If any of these relationships were violated, the gene would not be included in their study. Though, they did not include a representative basal lineage within

Rosoideae, like *Filipendula*, in their algorithm. It is unknown if the results would differ if they included taxa from a basal clade (like Ulmarieae) instead of *Rubus coreanus*. If so, the overall topology for the concatenated tree might be changed, perhaps changing the positions of Rubeae and Colurieae as a consequence.

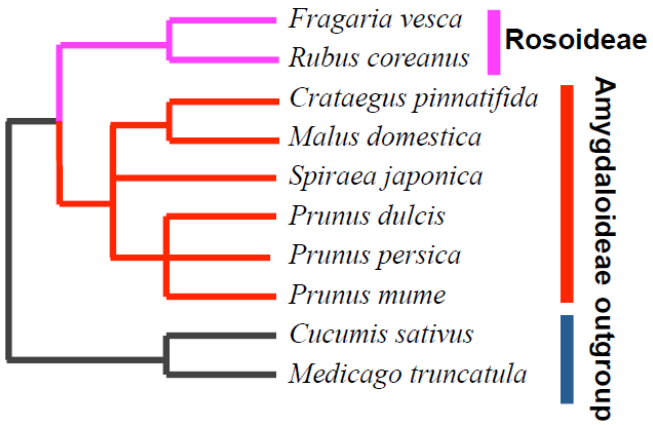

Fig. 7. Constraint tree of chosen taxa as presented in Xiang et al. (2017).

Xiang et al. (2017) combined several hundred genes. Concatenation can result in higher topological support but under the assumption that all the datasets of the genes have the same evolutionary pattern. When recombination occurs, separate loci mostly evolves independently (Cronn et al., 1999). The topology can become easily adjusted and artificial if you combine genes which tell different evolutionary histories (Bravo et al., 2018; Kubatko & Degnan, 2007; Gadagkar et al., 2005). In the combined datasets in the present study, ITS had the strongest influence over the topology of the nuclear genes. It gave the same basic topology as the ITS-only tree, supporting hypothesis (H1) when included and supporting hypothesis (H3) when excluded, which shows that combining datasets can be misleading for the species trees. However, since the support for the topologies of the individual low copy gene trees are low, one cannot directly tell if the topology of the concatenated trees in the present study are incorrect. If SbeI of all species were successfully sequenced as for the other low copy genes, the risk of phylogenetic errors would be reduced (Morales-Brionnes et al., 2018).

Hypothesis (H3): Colurieae and Rubeae belong to a clade together, sister to Roperculina

The third hypothesis suggests that Roperculina is sister to a clade consisting of Colurieae and Rubeae. This topology has only been published a few times before, and was first recovered by Eriksson et al. (1998), using parsimony analysis of ITS data. They only included one *Rubus* specimen in their dataset and the clade itself had very low support. Potter et al. (2002) similarly resolved these relationships in a parsimony analysis of the plastid *matK* and *trnL-trnF* markers, but also with low support. This was possibly due to focusing primarily on the Amygdaloideae subfamily, with only one representative of each tribe in Rosoideae except for Ulmarieae. The analysis of Eriksson et al. (2003) resulted in this topology using a parsimony analysis of ITS. Chin et al. (2014) also got this topology using the three plastid genes *matK*, *trnL-L-F* and *rbcL* but they did not include any *Filipendula* species in their sampling. The most recent study suggesting this topology was by Gaynor et al. (2018), but they do not display their branch support. Overall, there have been no well sampled previous studies which have shown this topology with strong support.

The third hypothesis was strongly supported by the combined analysis of the low copy genes in the present study. It was also supported by GAPCP1 with low support (0.58). When looking at the morphological features, the common traits among Colurieae and Rubeae, such as pinnate leaves, numerous stamens and pentamerous flowers, are synapomorphic in Rosodae. Colurieae and Rubeae differ in inflorescence, fruit type, number of ovules, and the epicalyx is present in Colurieae while absent in Rubeae. The low copy genes behind this result are more trustworthy than the previously discussed hypotheses. Therefore, to group Colurieae and Rubeae is a new way of looking into the basal phylogeny of the Rosoideae and deserve some additional attention in the future.

4.2 Possible explanations for the incongruent placement of *Rubus*

Each hypothesis is supported by either previous studies or in the present study: H1 – Zhang et al. 2017 (chloroplast genome), H2 – Xiang et al. (2017)(nuclear genes) and H3 – The low copy genes of this study. Both Xiang et al. (2017) and Zhang et al. (2017) boldly state that they have resolved the placement of *Rubus* but they are clearly incongruent. Both of their datasets include many taxa and genes but there is still a difference between the chloroplast phylogenies and nuclear

phylogenies. Does this mean there are two different evolutionary histories? Disagreements between nuclear genes and chloroplast genes are not uncommon in Rosoideae (Kerr, 2004). If the genes indeed tell different evolutionary histories, they will coalesce differently compared to the branching patterns of the species tree.

A plausible reason behind the incongruence between the gene trees include hybridization between the lineages and incomplete lineage sorting (ILS) because they share similar patterns in regard to their genetic diversity (Zhou et al., 2017). ILS can pose a problem if polymorphism occurs before speciation. Polymorphism is reported to be common in ribosomal nuclear genes such as ITS (Hershkovitz et al., 1999). ILS is also more common the more taxa and genes that are included in the dataset (Bravo et al., 2018).

Another possible explanation could be gene duplication and loss (GDL). If a gene undergoes a duplication, it will result in two copies. These two copies will evolve independently (Maddison, 1997) in different loci. If one gene at one locus goes extinct for one species, while the same gene but in the different locus goes extinct for another species, the species trees will be a result of two paralogous genes and will not concur with the gene trees.

Horizontal gene transfer (HGT), the genetic transfer between two species, is also possible among plants. This can either happen through a bacterial vector, or introgression, backcrossing between an interspecific hybrid and parent species (Maddison, 1997; Choudhuri, 2014). If an original copy of a gene is lost in similar situation as GDL for a species, and then the same species receives a copy from another species, the gene tree will discord with the species tree.

Genes can undergo many processes which may have an influence on their evolutionary history. This should be taken into account when reconstructing phylogenies, although such processes can be difficult to detect.

4.3 Future prospects

Since the individual nuclear low copy trees had low support, aside from trying to sequence the same taxa for all of the genes, a suggestion for further study is to find the single copy nuclear genes which has the appropriate evolutionary rates for the tribe taxonomic level, as different genes evolve at different rates (Clegg et al., 1997).

It would also be noteworthy to examine how much the different hypotheses were supported by the data from Xiang et al. (2017) and Zhang et al. (2017). This can be done by investigating the Bayesian tree sample for the number of trees that support the alternative topology (Lundberg et al. 2009).

Another suggestion is to create a red thread, intertwining morphological data with phylogenetic data. In other words, it would be noteworthy to see if making a matrix which includes all the morphological traits for the taxa and then comparing these morphological with already existing phylogenies will yield any matches. New methods could possibly facilitate the phylogeny. Sanger sequencing, which was conducted in this study, is a traditional method for DNA sequencing (Bleidorn et al., 2016). Third generation sequencing could facilitate the assembly of the polyploid genomes as well as heterozygosity, a problem when reconstructing the phylogeny even for diploid individuals (Jiao & Schneeberger, 2017).

4.4. Concluding remarks

This study has given an insight into a basal phylogenetic enigma within Rosoideae, which has barely been sparked upon. This study strongly support the Rubeae either forming a clade with Roperculina or Colurieae, but the latter two do not form its separate clade. ILS, GDL and HGT are suggested to be the possible causes behind the incongruence between chloroplast and nuclear phylogenies.

Acknowledgements

Many thanks to Heidi Lie Andersen for her support, guidance and feedback throughout this project. Also thanks to Torsten Eriksson for valuable information, and feedback along with Jenny Smedmark, Anne Elisabeth Bjune, Nannie Persson and Ingrid Toresen. Thanks to Per Harald Salvesen for help with collection of plant material and Louise Maria Lindblom and Solveig Thorkildsen from the Biodiversity laboratories (BDL) at UiB, as well as Hoda Parsian for their guidance in the lab during this project. Also thanks to The Arboretum and Botanical Garden in Milde, Bergius Botanical Garden, Gothenburg Botanical Garden, Jensen-Olson Arboretum and Royal Botanical Garden of Edinburgh for sending plant material.

References

Álvarez, I, Wendel, JF. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29(3): 417-434.

Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21: 163-193.

Bleidorn C. 2016. Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity* 14(1): 1-8.

Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom M, Huynh S, Jones G, Knowles LL, Lamichhaney S, Marcussen T, Morlon H, Nakhleh L, Oxelman B, Pfeil B, Schliep A, Wahlberg N, Werneck F, Wiedenhoeft J, Willows-Munro S, Edwards SV. 2018. Embracing heterogeneity: Building the Tree of Life and the future of phylogenomics. *PeerJ*. doi: 10.7287/peerj.preprints.26449v3

Bruni I, de Mattia F, Martellos S, Galimberti A, Savadori P, Casiraghi M, Nimis PL, Labra M. 2012. DNA Barcoding as an Effective Tool in Improving a Digital Plant Identification System: A Case Study for the Area of Mt. Valerio, Trieste (NE Italy). *PLOS ONE* 7(9): e43256. doi: 10.1371/journal.pone.0043256

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei M, Barrett SCH. 2011. Discriminating plant species in a local temperate flora using the *rbcL+matK* DNA barcode. *Methods in Ecology and Evolution* 2(4): 333-340.

Chin SW, Shaw J, Haberle R, Wen J, Potter D. 2014. Diversification of almonds, peaches, plums and cherries – Molecular systematics and biogeographic history of Prunus (Rosaceae). *Molecular Phylogenetics and Evolution* 76: 34-48.

Choudhuri S. 2014. Chapter 2 – Fundamentals of Molecular Evolution. Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools. 27-53.

Clegg MT, Cummings MP, Durbin ML. 1997. The evolution of plant nuclear genes. *Proceedings of the National Academy of Sciences of the United States of America* 93(15): 7791-7798.

Cronn RC, Small RL, Wendel JF. 1999. Duplicated genes evolve independently after polyploid formation in cotton. *Proceedings of the National Academy of Sciences of the United States of America* 96(25): 14406-14411.

de Vere N, Rich TCG, Ford CR, Trinder SA, Long C, Moore CM, Satterthwaite D, Davies H, Allainguillaume J, Ronca S, Tatarinova T, Garbett H, Walker K, Wilkinson MJ. 2012. Creation of a DNA database for a nation's native flowering plants and conifers. *PLOS ONE* 7(6): e37945. https://doi.org/10.1371/journal.pone.0037945

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29(8): 1969-1973.

Eriksson T, Donoghue MJ, Hibbs MS. 1998. Phylogenetic analysis of *Potentlla* using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of *Rosoideae* (*Rosaceae*). *Plant Systematic Evolution* 211(3-4): 155-179.

Eriksson T, Hibbs MS, Yoder AD, Delwiche CF, Donoghue MJ. 2003. 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. *International Journal of Plant Science* 164(2): 197-211.

Evans RC, Alice LA, Campbell CS, Kellogg EA, Dickinson TA. 2000. The Granule-Bound Starch Synthase (GBSSI) Gene in the Rosaceae: Multiple Loci and Phylogenetic Utility. *Molecular Phylogenetics and Evolution* 17(3): 388-400.

Gadagkar SR, Rosenberg MS, Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence trees versus consensus gene tree. *Journal of Experimental Zoology* 304B(1): 64-74.

Gaynor ML, Ng J, Laport RG. 2018. Phylogenetic Structure of Plant Communities: Are Polyploids Distantly Related to Co-occurring Diploids? Frontiers in Ecology and Evolution 6:52.

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4): 457-472.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52(5): 696-704.

Hasegawa M, Kishino H, Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22(2): 160-174.

Heled J, Drummond AJ. 2010. Bayesian Inference of Species Trees from Multilocus Data. Molecular Biology and Evolution 27(3): 570-580.

Hershkovitz MA, Zimmer EA, Hahn WJ. 1999. Ribosomal DNA sequences and angiosperm systematics. *Molecular systematics and plant evolution*. London, Taylor & Francis. 268-326.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754-755.

Hurvich CM, Tsai CL. 1989. Regression and time series model selection in small samples. *Biometrika* 76(2): 297-307.

Hutchinson J. 1964. *The genera of flowering plants, vol. 1, Dicotyledons*. Clarendon Press, Oxford.

IPCN, Index to plant chromosome numbers. 1979. Goldblatt P, Johnson DE. Missouri Botanical Garden, St. Louis. Available from: www.tropicos.org/Project/IPCN [Accessed 20 April 2018].

Jiao WB, Schneeberger K. 2017. The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* 36: 64-70.

Kalkman C. 2004. Rosaceae. *The families and genera of vascular plants, vol. 6, Flowering plants – Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*. Springer, Berlin. 343-386.

Katoh K, Misawa K, Kuma K, Miyata K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30(14): 3059-3066.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12): 1647-1649.

Kerr MS. 2004. A phylogenetic and biogeographic analysis of Sanguisorbeae (Rosaceae), with emphasis on the Pleistocene radiation of the high Andean genus *Polylepis*. Digital Repository at the University of Maryland. Available from: https://drum.lib.umd.edu/handle/1903/1678 [Accessed 14 May 2018].

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16(2): 111-120.

Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56(1): 17-24.

Le Dantec L, Cardinet G, Bonet J, Fouché M, Boudehri K, Monfort A, Poëssel JL, Moing A, Dirlewanger E. 2010. Development and mapping of peach candidate genes involved in fruit quality and their transferability and potential use in other Rosaceae species. *Tree Genetics & Genomes* 6(6): 995-1012.

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50(6): 913-925.

Lundberg M, Töpel M, Eriksen B, Nylander JAA, Eriksson T. 2009. Allopolyploidy in Fragariinae (Rosaceae): Comparing four DNA sequence regions, with comments on classification. *Molecular Phylogenetics and Evolution* 51(2): 269-280.

Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* 46(3): 523-536.

Maddison WP, Maddison DR. 2018. Mesquite: a modular system for evolutionary analysis. Version 3.40 http://mesquiteproject.org [Accessed 03 April 2018].

McGuire JA, Witt CC, Altshuler DL, Remsen JV Jr. 2007. Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. *Systematic Biology* 56(5): 837-856.

Mishima M, Ohmido N, Fukui K, Yahara T. 2002. Trends in site-number change of rDNA loci during polyploid evolution in Sanguisorba (Rosaceae). Chromosoma 110(8): 550-558.

Morgan DR, Soltis DE, Robertson KR. 1994. Systematic and evolutionary implications of *rbcL* sequence variation in Rosaceae. *American Journal of Botany* 81(7): 890-903.

Morales-Briones DF, Liston A, Tank DC. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* 218(4): 1668-1684.

Nylander JAA. 2004. MrAIC.pl. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Potter D, Eriksson T, Evans R, Kerr M, Robertson KR, Arsenault M, Dickinson TA, Campbell CS. 2007. Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution* 266(1): 5-43.

Potter D, Gao F, Bortiri PE, Oh SH, Baggett S. 2002. Phylogenetic relationships in Rosaceae inferred from chloroplast *matK* and *trnL-trnF* nucleotide sequence data. *Plant Systematic Evolution* 231(1-4): 77-89.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Tracer v1.7, Available from: http://beast.community/tracer [Accessed 28 March 2018].

Robertson KR. 1974. The genera of Rosaceae in the Southeastern United States. *Journal of the Arnold Arboretum* 55(2): 303-332.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572-1574.

Rousseau-Gueutin M, Gaston A, Aïnouche A, Aïnouche ML, Olbricht K, Staudt G, Richard L, Denoyes-Rothan B. 2009. Tracking the evolutionary history of polyploidy in *Fragaria* L. (strawberry): New insights from phylogenetic analyses of low-copy nuclear genes. *Molecular Phylogenetics and Evolution* 51(3): 515-530.

Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37(3): 121-147.

Shi S, Li J, Yu J, Zhou S. 2013. Phylogeny and Classification of *Prunus sensu lato* (Rosaceae). *Journal of Integrative Plant Biology* 55(11): 1069-1079.

Shoemaker JS, Fitch WM. 1989. Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. *Molecular Biology and Evolution* 6(3): 270-289.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton JM, Rees DJ, Williams KP, Holt SH, Ruiz Rojas JJ, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin SA, Troggio M, Viola R, Ashman TL, Wang H, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant DW Jr, Fox SE, Givan SA, Wilhelm LJ, Naithani S, Christoffels A, Salama DY, Carter J, Lopez Girona E, Zdepski A, Wang W, Kerstetter RA, Schwab W, Korban SS, Davik J, Monfort A, Denoyes-

Rothan B, Arus P, Mittler R, Flinn B, Aharoni A, Bennetzen JL, Salzberg SL, Dickerman AW, Velasco R, Borodovsky M, Veilleux RE, Folta KM. 2011. The genome of woodland strawberry (Fragaria vesca). *Nature Genetics* 43(2): 109-116.

Simmons MP, Ochoterena H. 2000. Gaps as Characters in Sequence-Based Phylogenetic Analyses. *Systematic Biology* 49(2): 369-381.

Simpson MG. 2010. *Plant systematics, 2nd edition*. Amsterdam, Netherlands: Elsevier/Academic Press.

Smedmark JEE, Eriksson T. 2002. Phylogenetic relationships of *Geum* (Rosaceae) and relatives inferred from the nrITS and *trnL-trnF* regions. *Systematic Botany* 27(2): 303-317.

Smedmark JEE, Eriksson T, Evans RC, Campbell CS. 2003. Ancient allopolyploid speciation in Geinae (Rosaceae): evidence from nuclear granule-bound starch synthase (GBSSI) gene sequences. *Systematic Biology* 52(3): 374-385.

Spies JJ, Du Plessis H. 1985. The genus *Rubus* in South Africa. III. The occurrence of apomixis and sexuality. *South African Journal of Botany* 52(3): 226-232.

Spjut RW. 1994. A systematic treatment of fruit types. *Memoirs of The New York Botanical Garden* 70: 1-182.

Sugiura N. 1978. Further analysis of data by Akaikes information criterion and finite corrections. *Communications in Statistics – Theory and Methods* 7(1): 13-26.

Swofford DL. 2002. *PAUP* Phylogenetic analysis using parsimony (* and other methods) Version 4*. Sinauer Associates, Sunderland, Massachusetts.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* 17(2): 57-86.

Thompson, MM. 1997. Survey of Chromosome Numbers in *Rubus* (Rosaceae: Rosoideae). Annals of the Missouri Botanical Garden 84(1): 128-164.

Toresen I. 2018. *Infrageneric phylogeny and evolutionary history of Potentilla L. (Rosaceae) based on three singlecopy genes* (Master thesis). University of Bergen.

White TJ, Bruns T, Lee SJWT, Taylor JL. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications* 18(1): 315-322.

Xiang Y, Huang CH, Hu Y, Wen J, Li S, Yi T, Chen H, Xiang J, Ma H. 2017. Evolution of Rosaceae Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. *Molecular Biology and Evolution* 34(2): 262-281.

Yang Z. 1996. Among-site variation and its impact on phylogenetic analyses. *Trends in Ecology & Evolution* 11(9): 367-371.

Yu J, Xue JH, Zhou SL. 2011. New universal *matK* primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution* 49(3): 176-181.

Zhang SD, Jin JJ, Chen SY, Chase MW, Soltis DE, Li HT, Yang JB, Li DZ, Yi TS. 2017. Diversification of Rosaceae since the Late Cretaceaous plastid phylogenetics. *New Phytologist* 214(3): 1355-1367.

Zhou Y, Duvaux L, Ren G, Zhang L, Savolainen, O, Liu J. 2017. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity (Edinburg)* 118(3): 211-220.

Appendix

Appendix I. Voucher specimens included in this study. Unknown ploidy levels are marked with a question mark. Source: Index to plant chromosome numbers (IPCN).

Tribe	Taxon	Voucher	Ploidy level
Colurieae	*Fallugia paradoxa* (D. Don) Endl.	IPEN:US:0:SBT :V:897	Diploid
	Geum pentapetalum (L.) Makino	K. Moori 2001-0710	Diploid
	Geum rivale L.	BG/S-165240	Hexaploid
Potentilleae	*Potentilla micrantha* Ramond ex DC.	RBGE	Diploid
	Potentilla reptans L.	PHS 16.54	Tetraploid
Roseae	*Rosa arvensis* Hudson	Z-1992.1462	Diploid
	Rosa hirtula (Regel) Nakai	G-1993.0773	?
	Rosa majalis J. Herrmann	PHS 16.54	Diploid
	Rosa onoei Makino	Z-1992.1516	Diploid
	Rosa rugosa Thunb.	T.E. 1059	Diploid
	Rosa sericea Lindl.	PHS2001.109	?
Rubeae	*Rubus chamaemorus* L.	BG/S-165238	Octoploid
	Rubus odoratus L.	PHS 16.59	Diploid
	Rubus parviflorus Nutt.	PHS 16.58	Diploid
	Rubus pedatus Sm.	Cult. Alaska	Diploid
Sanguisorbeae	*Agrimonia eupatoria* L.	PHS 16.51	Tetraploid
	Sanguisorba officinalis L.	BG/S-165237	Tetraploid
Ulmarieae	*Filipendula ulmaria* (L.) Maxim	BG/S-165239	Diploid
	Filipendula vulgaris Moench	PHS 16.52	Diploid

Appendix II. Extractions included in this study. Sequences marked as "New" are generated from this study and will later be published in GenBank.

Taxon	**ITS**	**GAPCP1**	**GBSSI**	**GBSSI-2**	**SbeI**	***matK***	**Reference**
Agrimonia eupatoria L.	New	New	New		New	New	This paper
Fallugia paradoxa (D. Don) Endl.	New			New	New	New	This paper
Filipendula multijuga Maxim						AB073684	Mishima et al., 2002
Filipendula ulmaria (L.) Maxim	New	New			New		This paper
Filipendula vulgaris Moench	New	New	New	New	New		This paper
Filipendula vulgaris Moench						HE966930	Bruni et al., 2012
Fragaria vesca L.			NC_020497				Shulaev et al., 2011
Geum aleppicum Jacq.						HQ593312	Burgess et al., 2012
Geum pentapetalum (L.) Makino	New	New		New	New	New	This paper
Geum rivale L.	New	New			New		This paper
Geum rivale L.			AJ534192				Smedmark et al., 2003
Geum urbanum L.						JN894110	de Vere et al., 2012
Potentilla micrantha Ramond ex DC.			New				Toresen, 2018
Potentilla reptans L.	New	New		New	New	New	Toresen, 2018
Rosa arvensis Hudson	New	New			New	New	This paper
Rosa hirtula (Regel) Nakai	New	New			New	New	This paper
Rosa majalis J. Herrmann	New	New	New		New	New	This paper
Rosa onoei Makino	New			New	New	New	This paper
Rosa rugosa Thunb	New	New	New	New	New	New	This paper
Rosa sericea Lindl.	New	New			New	New	This paper
Rubus chamaemorus L.	New	New			New	New	This paper
Rubus odoratus L.	New	New	New	New	New	New	This paper
Rubus parviflorus Nutt.	New		New		New	New	This paper
Rubus pedatus Sm.	New	New	New	New	New	New	This paper
Sanguisorba officinalis L.	New	New			New	New	This paper