

# The taxonomy and molecular phylogeny of *Potentilla* L. (Rosaceae)

An investigation of generic delimitation and reticulate evolution, using low-copy nuclear markers

---

Nannie Linnéa Persson

Thesis for the degree of Philosophiae Doctor (PhD)  
University of Bergen, Norway  
2021

UNIVERSITY OF BERGEN



# The taxonomy and molecular phylogeny of *Potentilla* L. (Rosaceae)

An investigation of generic delimitation and reticulate evolution, using low-copy nuclear markers

Nannie Linnéa Persson



Thesis for the degree of Philosophiae Doctor (PhD)  
at the University of Bergen

Date of defense: 15.01.2021

© Copyright Nannie Linnéa Persson

The material in this publication is covered by the provisions of the Copyright Act.

Year: 2021

Title: The taxonomy and molecular phylogeny of *Potentilla* L. (Rosaceae)

Name: Nannie Linnéa Persson

Print: Skipnes Kommunikasjon / University of Bergen

## **Scientific environment**

This PhD project was carried out in the Rosaceae research group at the Department of Natural History, University Museum, University of Bergen (UiB), and the Biodiversity Lab, Department of Biological Sciences, UiB. The project was financed through a four-year PhD candidate position at the UiB. Additional funding for conference attendances, field work and lab work was provided by the Olaf Grolle Olsen and Miranda Bødtker's fund, and the University Museum at UiB.



## Acknowledgements

To start from the beginning, thank you to my aunt **Berit**, my grandfather **Arne** and my high school biology teacher **Roger**, who gave me my first book about birds, showed me the forest, and introduced me to the science of biology, which later led to my decision to study biology at Stockholm University. Thank you to my inspiring master thesis supervisor **Catarina** for teaching me about systematics, and her PhD students **Eva** and **Olle** for friendship through both my master and PhD studies.

During my time in Bergen I have met many people that have come to mean a great deal to me, both at the office and outside of the working hours. To my supervisors **Jenny** and **Torsten**, I have enjoyed our discussions about what is actually happening in a polyploid genome, how to interpret multilabelled and reticulate trees, and who should bring what to the party. Most of all, I appreciate all of your time that you have given me. To the other botanists at the museum – **Astri**, **Heidi**, **Hoda**, **Mike**, **Per Harald**, **Solfrid** and **Tor** – thank you for sharing your knowledge on anything practical and botanical, and thank you to **Ardian** and **Ingrid** for allowing me to take part in your master theses. To **Berit**, **Bjørn** and **Hilde** at the university gardens, thank you for the field trips and taking care of our seedlings.

My first office mates, **Richard** and **Trond**, thank you for putting up with my many questions when I first arrived at the museum. You gave me a very warm welcome.

A big thank you goes to the technicians at the DNA lab at BIO, especially **Louise** and **Solveig**, for help with everything from extraction to sequencing and for teaching me the art of molecular cloning. I have really appreciated the high standard of order and tidiness in the lab.

Thank you to the people I've met through ForBio for very educational courses and conferences; **Aino**, **Alison**, **Anneli**, **Emma**, **Hugo**, **José**, **Kalina**, **Miguel**, **Moa**, **Nataliya**, **William** and “the **Trondheim gang**”.

**Anna, Cessa, Justine and Luis** – I will never let you go.

Finally, thank you to my family and friends back in Sweden – my mum **Annelie**, my dad **Lennart**, my brother **Elias**, my grandparents **Inga** and **Arne** and my dear friends **Annelie, Emilie, Frida, Sandra** and **Sofia**. Your immense support means the world to me, and I am very lucky to have you. I love you.

Bergen, October 2020

Danne Linnéa Persson

## Abstract

**Premise and aims of the thesis:** Eukaryotes that have more than the two standard sets of chromosomes are called polyploids. The process of going through a whole genome duplication is called polyploidization, and this is a common mechanism in plant speciation. There are two main types of polyploidy: autopolyploidy and allopolyploidy. Autopolyploidy is the result of a genome duplication within one species, and allopolyploidy is the result of a genome duplication following a hybridization between two different species.

The genus *Potentilla* in the rose family (Rosaceae) is remarkable in that species range in ploidy level from diploid ( $2x$ ) to hexadecaploid ( $16x$ ). They are found all around the Northern Hemisphere, from lowland to mountain regions, and are generally characterized by yellow flowers and palmately compound leaves. However, taxonomists have long had problems with agreeing on which species should be included in the genus. Even today some authors exclude certain species from *Potentilla*, with the consequence of making it a non-monophyletic genus (i.e. not all of the descendants of their most recent common ancestor are included in the group).

Even though the prevalence of polyploidy in plants is well-known, it has not been reflected in phylogenetic research. This is a problem, especially concerning allopolyploids, because the understanding we get of their evolutionary history is then much simplified. Many studies have used DNA sequences that often represent only one ancestral lineage (chloroplast, nuclear ribosomal), thus omitting parts of the species' heritage. In previous phylogenetic analyses, a few major subclades were identified in *Potentilla* (informally named Alba, Anserina, Argentea, Fragarioides, Ivesioid and Reptans), but their relationships to one another differed depending on what type of DNA was studied. In addition, some species were found in different subclades in trees based on different DNA sequences. The fact that these sequences may be uniparentally inherited and that most of the species are polyploid led to an

interpretation of an evolutionary history that involves hybridization and polyploidization in *Potentilla*.

The type of DNA sequence best suited for investigating the evolutionary history of polyploids are low-copy nuclear DNA markers (LCN markers). They are present in each subgenome and inherited from both the maternal and the paternal parent. Thus, they have the potential to trace the relationships of each ancestral lineage of polyploids. LCN markers were in this thesis used for three different purposes in *Potentilla*: 1), to infer the relationships of the major subclades in the genus (Paper I); 2), to trace the putative hybrid origins of a number of North American polyploid species in the 'Rivales group' (Papers II and III); and 3), to assess the generic delimitation of *Potentilla* (Paper IV).

**Results and conclusions:** A fully resolved and supported tree showing the major subclades in *Potentilla* was obtained after excluding the *Fragarioides* species from the dataset. Two of the clades, the Ivesioid and Reptans clades, showed signs of being of autopolyploid origin. In contrast, five of the six species in the Rivales group occurring in North America were inferred to be allopolyploids with ancestral lineages in the Argentea and Ivesioid clades. Thus, hybridization and polyploidization seem to have played a larger role later in the evolution of the genus, after the major clades diverged.

Four lines of evidence – ploidy level, distribution of extant species, relationships seen in the gene trees, and a set of network analyses – indicated that precursors to three of the North American Rivales species have taken part in hybridizations that eventually formed a common ancestor for the high-ploidy Rivales species *P. intermedia* and *P. norvegica*. Parts of this population dispersed to Eurasia, while the rest remained in North America. Both lineages went through at least one more hybridization each and formed *P. intermedia* in Eurasia and *P. norvegica* in North America. Since many floras state that *P. norvegica* is of European origin, this will have implications for its assessment as native or introduced on both continents.

The gene trees inferred in Papers I, II and III showed a network of gene flow between the Alba, Argentea, *Fragarioides*, Ivesioid and Reptans clades. Thus, the generic

delimitation of *Potentilla* was set to include these clades, and excluding the Anserina clade. With this delimitation only six species, out of the ca 400 in the whole genus, had to be recombined to get new *Potentilla* names.

**Future perspectives:** The LCN markers revealed relationships that could not have been found by the traditionally used chloroplast or nuclear ribosomal markers. This points to the importance of continuing using LCN markers when investigating the evolutionary history of polyploids. Additional markers are, however, needed to resolve some relationships, especially the putatively diploid *Fragarioides* species destabilizing the backbone phylogeny, and some species in the Rivaes group of which we could not find all putative ancestral lineages. The High-Throughput Sequencing technique Target Capture could potentially generate enough data to solve these problems.

Software programs that analyze reticulate evolution still struggle with species of high ploidy levels, and a good deal of manual preparation of analyses and interpretation of the results are still needed. In addition, a discussion is needed concerning criteria for species delimitation of allopolyploids. If the ancestral lineages are distantly related, this could have implications at even higher taxonomical levels, such as genera and families.



## List of publications

### PAPER I

Persson NL, Toresen I, Andersen HL, Smedmark JEE, Eriksson T. 2020. Detecting destabilizing species in the phylogenetic backbone of *Potentilla* L. (Rosaceae) using low-copy nuclear markers. *Annals of Botany Plants* 12(3):plaa017  
Open access; DOI: <https://doi.org/10.1093/aobpla/plaa017>

### PAPER II

Persson NL, Eriksson T, Smedmark JEE. 2020. Complex patterns of reticulate evolution in opportunistic weeds (*Potentilla* L., Rosaceae), as revealed by low-copy nuclear markers. *BMC Evolutionary Biology* 20:38  
Open access; DOI: <https://doi.org/10.1186/s12862-020-1597-7>

### Paper III

Persson NL, Eriksson T, Smedmark JEE. Native or introduced? Tracing the origins of high-level allopolyploids in North America (*Potentilla* L., Rosaceae)  
Manuscript

### Paper IV

Eriksson T, Persson NL, Smedmark JEE. What is *Potentilla*? A phylogeny-based taxonomy for Potentillinae (Rosaceae)  
Manuscript



## Table of contents

<b>Acknowledgements</b>	<b>5</b>
<b>Abstract</b>	<b>7</b>
<b>List of publications</b>	<b>11</b>
<b>1. Introduction</b>	<b>15</b>
<b>1.1 Polyploidization – a driving force in plant evolution</b>	<b>15</b>
1.1.1 Polyploidy in plant research	18
<b>1.2 <i>Potentilla</i> L. of the Rose family (Rosaceae)</b>	<b>20</b>
1.2.1 Intrageneric relationships in <i>Potentilla</i>	21
1.2.2 The Rivales group	24
<b>2. Aims of the thesis</b>	<b>26</b>
<b>3. Materials and Methods</b>	<b>28</b>
<b>3.1 Material collection and taxon selection</b>	<b>28</b>
<b>3.2 Morphological methods</b>	<b>29</b>
<b>3.3 Molecular methods</b>	<b>29</b>
3.3.1 Molecular markers and primer design	29
3.3.2 Extraction and production of DNA sequences	29
<b>3.4 Sequence treatment and alignment</b>	<b>30</b>
<b>3.5 Phylogenetic analyses</b>	<b>30</b>
3.5.1 Bayesian Inference and Maximum Likelihood	30
3.5.2 Multispecies Coalescent analyses	31
3.5.3 Estimation of divergence times	31
3.5.4 Network analyses	32
<b>4. Results and Discussion</b>	<b>33</b>
<b>4.1 Intrageneric relationships of the major clades in <i>Potentilla</i></b>	<b>33</b>
4.1.1 Gene trees	33
4.1.2 Species trees	34
4.1.3 The Ivesioid clade and its taxonomical treatment	34
<b>4.2 The evolutionary history of the <i>Potentilla</i> Rivales group</b>	<b>34</b>
4.2.1 <i>Potentilla norvegica</i> and its two subspecies	34
4.2.2 <i>Potentilla norvegica</i> and <i>P. intermedia</i> have a common evolutionary history	35
4.2.3 The reticulate relationships of the North American Rivales species	35
<b>4.3 A suggestion for a monophyletic generic delimitation of <i>Potentilla</i></b>	<b>36</b>

<b>5. Future perspectives</b>	<b>38</b>
5.1 The putatively autoployploid origin of the Ivesioid species and the Reptans clade	38
5.2 The species-rich Argentea clade	38
5.3 Additional allopolyploid species in need of investigation	39
5.4 Other molecular methods and the need for software programs producing reticulate trees	39
5.5 Evaluation of the taxonomic treatment of polyploids	40
<b>6. References</b>	<b>42</b>

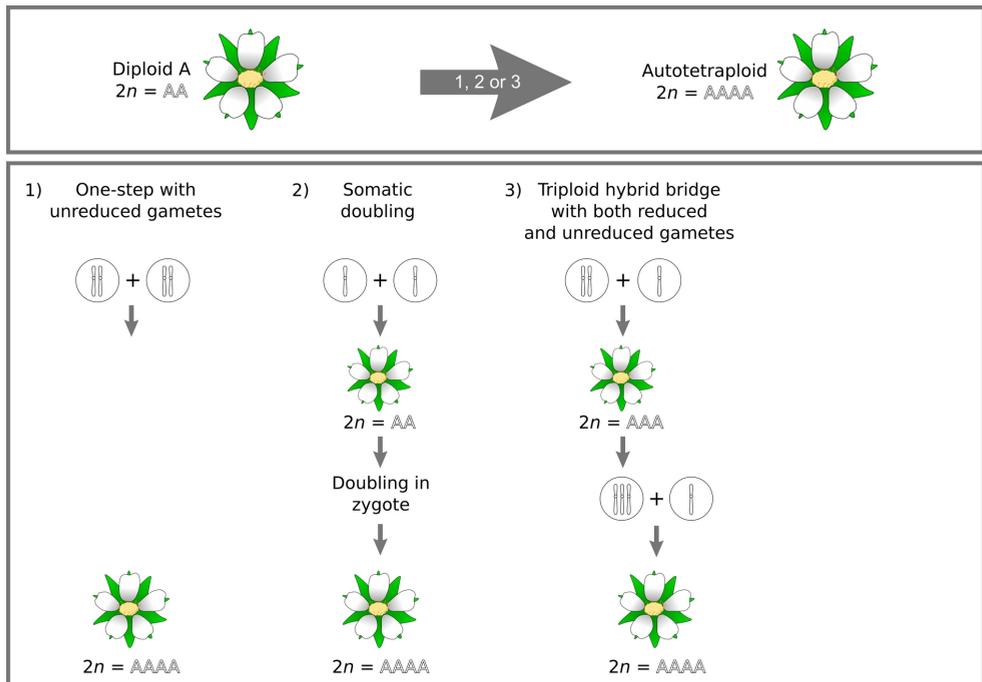
# 1. Introduction

## 1.1 Polyploidization – a driving force in plant evolution

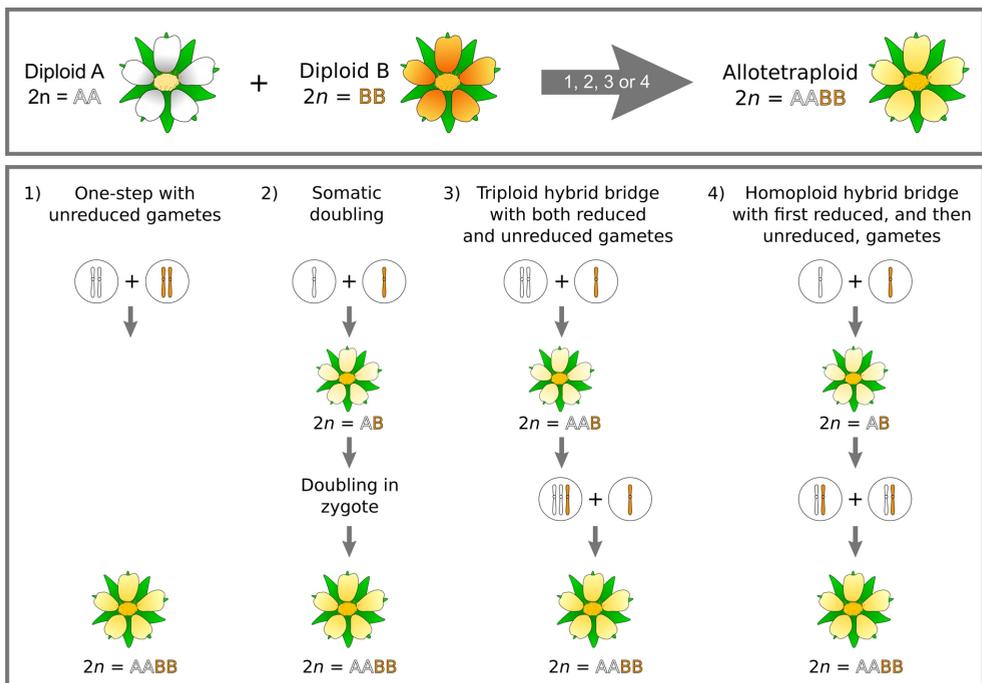
Many philosophers and naturalists have throughout history developed hypotheses on why there are so many different organisms on Earth (Uddenberg 2003). Ever since Darwin and Wallace, the theory of evolution has met both praise and heavy resistance, and still does. Within the scientific community the theory of evolution is, however, not just what everyday speech defines as a theory, but a well-established fact. Evolution through spontaneous mutations and natural selection is generally a slow process, working over thousands or millions of years, but sometimes the process is extremely quick, establishing a population of a new species over only a few generations. This rapid speciation can be explained by polyploidization – a doubling of all chromosomes in a genome – which may lead to a reproductive barrier towards individuals in the original population that have not gone through polyploidization. Polyploid individuals can then be regarded to constitute a new lineage that evolves independently from the rest of the population, and therefore be considered a new species (de Queiroz 2005).

Polyploidy differs from the standard eukaryotic (sporophytic) genomic state of diploidy ( $2x$ ) in there being three or more chromosome sets in each cell. It exists in both plants, animals and fungi (Albertin and Marullo 2012), but, as far as known, it is most common in plants. The majority of all flowering plants may in fact be paleopolyploids, stemming from a whole genome duplication event early in their evolutionary history (Cui et al. 2006). Since then, lineages have reverted back to functioning like diploids (diploidization) (Leitch and Bennett 2004; Bento et al. 2011; Mandáková et al. 2017). Nevertheless, polyploidization has continued to be an important process for speciation, and humans have taken advantage of this when developing new cultivars for crops and ornamental flowers (Mason and Batley 2015, Manzoor et al. 2019).

There are two main types of polyploidy: autopolyploidy, where a duplication has occurred within one species, and allopolyploidy, where a genome duplication has followed a hybridization between two different species (Kihara and Ono 1926). There are a number of different pathways in which a plant may reach a polyploid state, involving, for instance, unreduced gametes or post-zygotic doubling of the genome. Figures 1 and 2 show how tetraploid ( $4x$ ) auto- and allopolyploids may arise from diploid progenitors. The shortest way from diploid to polyploid is when two unreduced gametes fuse and form a tetraploid zygote (Figures 1.1 and 2.1). Another way is somatic doubling, which is caused by mitotic errors in a newly formed zygote (Figures 1.2 and 2.2). The third pathway that both auto- and allopolyploidization theoretically may go through includes a triploid bridge, where one reduced and one unreduced gamete fuse and form a triploid zygote (Figures 1.3 and 2.3). If this triploid is able to develop into an adult plant and produce gametes, any unreduced gametes it produces may then fuse with reduced gametes of a diploid and form tetraploid zygotes. An allopolyploid can also form via a homoploid hybrid bridge, where a homoploid hybrid first is formed via reduced gametes from two different species, and this hybrid is then fertilized by the unreduced gametes of another homoploid hybrid with the same parental species (Figure 2.4). There are additional variations to the pathways depicted, as, for instance, the unreduced gametes from a diploid may fuse with the reduced gametes of an autotetraploid of a different species and give rise to the same type of allotetraploid as in Figure 2. This latter example could also be defined as autoallopolyploidy, since a part of the hybrid genome comes from an autopolyploidization event. It is, however, difficult to determine which road to polyploidy is, or has been, the most common in plants. Matsuoka et al. (2013) showed that unreduced gametes are important in wheat, and Ramsey and Schemske (1998) argue that the triploid bridge could statistically play a large role in the doubling of the genome.



**Figure 1.** Pathways from diploid to autotetraploid. Adapted from Mason and Pires (2015).



**Figure 2.** Pathways from diploid to allotetraploid. Adapted from Mason and Pires (2015).

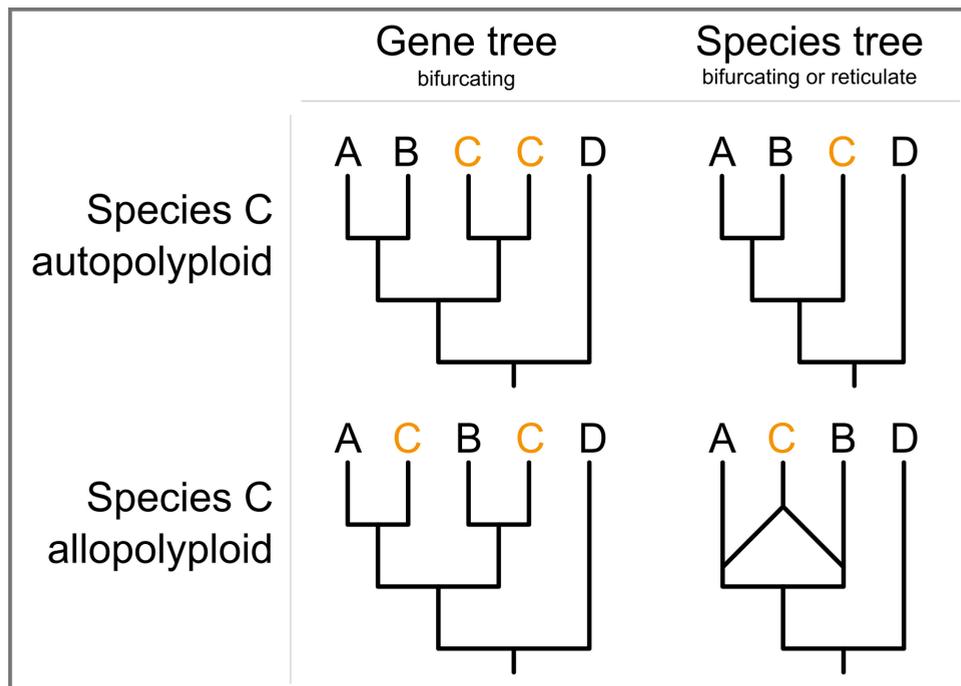
The reason for polyploidy being common among plants has been debated over the years. Newly formed polyploids are likely going through a period of genetic instability, where epigenetic forces work to both activate and suppress gene expression (Comai et al. 2000, Ramsey and Schemske 2002). During this process, individuals may have reduced fertility, but a population may be maintained through apomixis or self-fertilization (Sax 1954, Guggisberg et al. 2006). Once stability has been reached, a polyploid genome may provide heterosis and gene redundancy. Heterosis, or hybrid vigour, is sometimes referred to as the opposite of inbreeding depression, since the offspring show higher vigour than their parents due to increased heterozygosity at their loci (Timberlake 2013). Gene redundancy comes from there being multiple genes with the same function, which means that deleterious mutations at one locus may be masked by the other, still functioning, copies (Ascencio and LeLuna 2013). Therefore, the result of the mutation has little or no effect on the organism's overall fitness. On the opposite, mutations may in rare instances lead to a new function of one of the copies (neofunctionalization), while the other copies retain the original function (Prince and Pickett 2002). These traits may lead to a higher degree of durability and adaptability, and thus aid in colonization of other habitats than the parental species are normally found in (Brochmann et al. 2004).

### **1.1.1 Polyploidy in plant research**

Even though polyploidy has been known for over 100 years (Lutz 1907), tracing polyploid evolutionary history is still not an easy task. Different methods have been used, from comparative morphology and cytological studies, to molecular methods such as restriction fragment analysis and the use of DNA sequence data. The two types of molecular markers mostly used today in plant systematics research – chloroplast DNA (cpDNA) and nuclear ribosomal DNA (nrDNA) – have been instrumental in resolving both ancient and more recent nodes (Embley et al. 1994, Bouetard et al. 2010). Their advantage is that they are relatively easy to amplify because of their high number of copies in each cell. However, they are not very well suited for differentiating between autopolyploidy and allopolyploidy since cpDNA is usually inherited on the maternal side and nrDNA is often subjected to concerted

evolution (homogenization towards one of the parents' copy) (Small et al. 2004). Lundberg et al. (2009) relied on concerted evolution towards the paternal parent to detect allopolyploidy in the strawberry subtribe *Fragariinae* (Rosaceae). By comparing trees based on cpDNA and nrDNA, they looked for incongruent relationships that could indicate hybridization. However, with this method, hybridizations followed by concerted evolution towards the maternal parent would go undetected.

For that reason, low-copy nuclear (LCN) markers are better suited to trace polyploidizations since they are inherited biparentally, less subjected to concerted evolution and, at least initially after a polyploidization event, present in each subgenome (Small et al. 2004). Also, instead of the need for two trees, polyploidization can with this type of marker be detected in a single phylogenetic gene tree, since the gene copies of an autopolyploid (paralogues) would be each other's sisters, while the gene copies of an allopolyploid (homoeologues) would be sisters to their respective ancestral lineage (Figure 3). Nevertheless, different LCN



**Figure 3.** The different patterns that autopolyploid and allopolyploid species display in gene trees and in species trees.

markers may have different evolutionary histories due to horizontal gene transfer, deep coalescence and incomplete lineage sorting (Maddison 1997). In addition, LCN markers can not be concatenated into larger datasets since it is not possible to know before analysis which copies belong to homologous subgenomes. If the resulting gene trees of the different markers show different topologies, concatenation is still not possible. It is therefore important to investigate several LCN markers in order to find the species tree. In the case of allopolyploids, their evolutionary history cannot be illustrated with a traditional bifurcating tree, but with a reticulate tree (a network) where some lineages merge (Figure 3).

The traditional, well-tested method for extracting and separating the many copies of a LCN marker in a polyploid is molecular cloning followed by Sanger sequencing. This can either be done *in vitro*, where the PCR product is diluted to such low concentration that a second PCR only has a single template molecule to initiate the amplification from (Marcussen et al. 2012), or *in vivo*, by inserting the PCR products into vectors that are taken up by bacteria. The bacteria are then diluted and spread across a growth medium, and the resulting colony will thus contain the clones from a single bacterium and one specific PCR product (Ford and Gottlieb 1999, Brassac et al. 2012). The modern high-throughput method of Target Capture followed by next generation sequencing has the advantage of being able to produce a hundred-fold many more markers than cloning in the same amount of time, but the bioinformatic process to sort the data post-sequencing is also much more extensive (Eriksson et al. 2017, Kamneva et al. 2017). Therefore, this new technique has so far mostly been applied to tetraploids, and very rarely hexaploids (6x) or octoploids (8x).

## 1.2 *Potentilla* L. of the Rose family (Rosaceae)

The Rose family (Rosaceae) is well-known for its economically important crops, such as cherries (*Prunus*), apples (*Malus*), and strawberries (*Fragaria*), and ornamentals such as roses (*Rosa*), whitebeams (*Sorbus*) and cotoneasters (*Cotoneaster*). The family is also known for its many polyploid taxa (Vamosi and Dickinson 2006). Most genera have a base chromosome number of 7, 8 or 9, while some genera in the subfamily

Amygdaloideae have a base chromosome number of 17, believed to have arisen through an ancient polyploidization event (Evans and Campbell 2002). Early molecular studies of the family revealed the base chromosome number to be a better indicator of relationships than the traditionally used character fruit type (Morgan et al. 1994). The current classification of subgroups within Rosaceae include three subfamilies (Dryadoideae, Rosoideae and Amygdaloideae) and a number of tribes and subtribes within these (Potter et al. 2007, Xiang et al. 2017, Zhang et al. 2017).

The genus *Potentilla* L. is currently classified in subfamily Rosoideae, tribe Potentilleae and subtribe Potentillinae (Eriksson et al. 1998, Dobeš and Paule 2010, Töpel et al. 2011). All species in *Potentilla* are herbs characterized by white, red or yellow petals, an epicalyx, subterminal styles, pinnate or palmately compound leaves and lateral stipular auricles (Soják 2008, Soják 2010). They are mostly pollinated by bees and flies, and their fruit type is an achene (McIver and Erickson 2012, Ertter et al. 2014). There are species native to all continents on the Northern Hemisphere, and a few species are native to South America (POWO). The genus is remarkable in that the ploidy levels range from diploid to hexadecaploid (16x), and single species can be of multiple ploidy levels (Kalkman 2004, Rice et al. 2014). In addition to reproducing sexually, some species are facultative apomicts (Asker 1970, Eriksen 1996).

### 1.2.1 Intrageneric relationships in *Potentilla*

Even before Linnaeus (1753) and later, taxonomists have classified the species of *Potentilla* in a number of different genera and sections (e.g. Séguier 1754, Rydberg 1898, Soják 2010, Kechaykin and Shmakov 2016). The most comprehensive monograph so far written is that of Theodor Wolf (1908), who assigned just over 300 species to the genus. Eriksson et al. (1998) performed the first molecular study of *Potentilla* and discovered that the circumscription at the time was polyphyletic. This caused some species described as *Potentilla* by Linnaeus (1753) to be reclassified in other genera, e.g. to *Drymocallis* L. and *Dasiphora* Raf. In contrast, some species of other genera were incorporated in *Potentilla*, e.g. from *Duchesnea* Sm. and *Sibbaldia* L. (Eriksson et al. 2003, 2015). Subsequent molecular studies have found a number of

supported subclades within the genus, informally named the Alba, Anserina, Argentea, Fragarioides, Ivesioid, and Reptans clades (Töpel et al. 2011) (Figure 4). The Alba clade consists of many of the genus' white-flowered members and they are mainly found in mountainous regions in Europe and Asia. The Anserina clade is today classified as the genus *Argentina* Hill. (Soják 2010, Roskov et al. 2019, POWO), and most species are found in Asia. The Argentea clade is the most species-rich and its members are found all around the Northern Hemisphere. In contrast, the Fragarioides clade consists of four species that are found only in East Asia. The species of the Ivesioid clade are restricted to western North America and have, due to some species' differing morphology, often been assigned to other genera (Ertter and Reveal 2014). Finally, the Reptans clade is a small clade with a circumpolar distribution that includes the type species of *Potentilla*, *P. reptans* L. The name of a type species is taxonomically connected to the genus it is classified in, and therefore the name *Potentilla* cannot be applied to a group of species that does not include *P. reptans*.

Even though the subclades are well-supported in several markers, there are a few noticeable incongruences when comparing cpDNA and nrDNA trees (Dobeš and Paule 2010, Töpel et al. 2011, Feng et al. 2017). These incongruences include the position of the Anserina clade in relation to Potentillinae and its sister subtribe Fragariinae, and the position of the Reptans clade in relation to the Fragarioides, Argentea and Ivesioid clades. In addition to incongruences between the clades, the internal relationships of the clades are in many cases also uncertain. For example, the large Argentea clade is at present almost completely unresolved, and a few species switch position between the Argentea and Ivesioid clades. Nevertheless, there are also some congruencies in the trees. For instance, the Argentea and Ivesioid clades seem to be closely related, the Anserina clade is always positioned outside of the other subclades and the majority of the species are found in the same subclades in all trees.



**Figure 4.** Representatives of some of the subgroups in the Rosaceae subtribe Potentillinae; a) *Potentilla aurea* in the Argentea clade, b) *P. tilingii* in the Ivesioid clade, c) *P. reptans* in the Reptans clade and type species of *Potentilla*, d) *P. clusiana* in the Alba clade, e) *Argentina anserina* in the Anserina clade, f) *P. norvegica* in the Rivales group. Photos by Nannie L. Persson, 2017 & 2019.

### 1.2.2 The Rivales group

All the species that have been seen to switch positions between the Argentea and Ivesioid clades were classified by Wolf (1908) in his “GreX” Rivales and occur in North America (POWO). He defined larger groups based on style shape and its position on the ovary, and Rivales is further characterized by that the species are relatively short-lived, have long and narrow styles, and have an affinity for moist soil. In addition to North America, the group consists of species native to Europe, Asia and South America. The species that has been included in most phylogenetic analyses is the circumpolar *Potentilla norvegica* L., which is in the Argentea clade in analyses based on cpDNA (Eriksson et al. 2003, Dobeš and Paule 2010, Töpel et al. 2011), and in the Ivesioid clade in analyses based on nrDNA (Eriksson et al. 1998, Töpel et al. 2011). In the combined cpDNA and nrDNA analysis by Koski and Ashman (2016), *P. intermedia* L., *P. newberryi* A.Gray, and *P. rivalis* Nutt. resolved together with *P. norvegica* in the Ivesioid clade. In contrast, *P. biennis* Greene is always resolved with the Ivesioid clade, and *P. supina* L. is always resolved with the Argentea clade (Dobeš and Paule 2010, Töpel et al. 2011, Koski and Ashman 2016).

Available chromosome counting data reveal *P. intermedia*, *P. norvegica*, *P. rivalis* and *P. supina* to be polyploid, while no such data exist for *P. biennis* and *P. newberryi* (Rice et al. 2014). For *P. intermedia*, there are reports on tetra-, hexa- and octoploid individuals, for *P. norvegica* there are octo- and decaploid (10x) individuals, for *P. rivalis* there are decaploid individuals (and tetraploid for *P. pentandra* Engelm., at present considered to be synonymous with *P. rivalis*) and for *P. supina* there are tetra- and hexaploid individuals (Rice et al. 2014). In light of several species being polyploid, Töpel et al. (2011) suggested that the incongruences seen between their trees based on cpDNA and nrDNA could be explained by allopolyploidization.

*Potentilla norvegica* was described as two species by Linnaeus (1753) – *P. norvegica* and *P. monspeliensis* L. – but today *P. monspeliensis* is usually classified as a subspecies to *P. norvegica*; *P. norvegica* ssp. *hirsuta* (Michx.) Hyl. In general, *P. norvegica* ssp. *hirsuta* is most common in North America, and is therefore sometimes

referred to as the American form. Similarly, the autonym *P. norvegica* ssp. *norvegica* is sometimes referred to as the European form, but there are numerous findings of *P. norvegica* ssp. *hirsuta* in Europe and vice versa. Most floras state that the species originated in eastern Europe, and that *P. norvegica* ssp. *hirsuta* later dispersed to Europe from North America (Tutin et al. 1968, Hultén 1971, Kurtto et al. 2004, Lid and Lid 2013).

However, no molecular phylogenetic studies have previously been performed in order to test whether allopolyploidization is the cause of the incongruences seen in the trees, if the Rivales species share hybridization events, and if geographical origin and morphology are consistent with the intraspecies phylogeny of *P. norvegica*.

## 2. Aims of the thesis

The understanding we have today of the evolutionary history of plants is not incorrect, but much simplified given that polyploidy exists on the majority of the branches. Markers of cpDNA and nrDNA have been of tremendous importance in building the foundation of the tree, but they have not been able to resolve all the nodes even after more taxa and more markers have been added to the datasets. In addition, the relationships they present are usually uniparental. Therefore, another type of data is needed, and few previous studies have attempted to resolve the evolutionary history of high-level polyploids using biparentally inherited low-copy nuclear (LCN) markers. Trees based on LCN markers in polyploids are multi-labelled and may be difficult to interpret, but they also bring us closer to the true evolutionary history of organisms.

The genus *Potentilla* is of particular systematic interest due to its wide range in ploidy levels ( $2x - 16x$ ), wide range in distribution and wide range in types of habitat. While the type of habitat for each species is usually restricted, some species show a circumpolar distribution and some species have multiple ploidy levels. The different relationships seen for the subclades and for certain species indicate different histories of the genes analyzed. Considering the many polyploid taxa in the genus, it is likely that some of the incongruences seen are caused by hybridization in combination with polyploidization (allopolyploidization).

Most classifications of *Potentilla* were made before the first molecular analyses of the genus and its closest relatives (e.g. Linnaeus 1753, Rydberg 1898, Wolf 1908), and these circumscriptions turned out to be non-monophyletic (Eriksson et al. 1998). After several molecular studies have been published, some more recent authors still suggest classifications that are non-monophyletic (Mabberley 2002, Ertter et al. 2014, Kechaykin and Shmakov 2016).

Thus, the main objectives of this thesis were to resolve the uncertain nodes in *Potentilla* concerning the relationships among the major clades (Paper I) and the origin of a few species in the Rivales group (Wolf 1908) suspected to be closely related. First

focusing on *P. norvegica* (Paper II), the dataset was later extended to include all Rivaless species occurring in North America (Paper III).

On the basis of the new phylogenies produced in Papers I, II and III, as well as those published in previous studies, a suggestion for a monophyletic generic delimitation of *Potentilla* was made (Paper IV).

### 3. Materials and Methods

Unless specified otherwise, all methods described below concern Papers I, II and III.

#### 3.1 Material collection and taxon selection

Material for DNA extraction and morphological studies was obtained from herbaria (BG, E, GB, JEPS, MARY, O, S, UPS, W and WU) and botanical gardens (Bergius Botanic Garden Stockholm, Bonn University Botanic Gardens, The Linnéan Gardens of Uppsala The Museum Garden in Bergen and Royal Botanic Garden Edinburgh). Two collection trips were also made, one to Austria in July 2017 and one to California, USA, in August 2019.

To resolve the phylogenetic position of the *Potentilla* subclades (Paper I), taxa were chosen to represent the six major clades (Alba, Anserina, Argentea, Fragarioides, Ivesioid and Reptans) identified by Dobeš and Paule (2010), Töpel et al. (2011) and Feng et al. (2017). Taxa that have recently been classified in the genera *Horkelia*, *Horkeliella* and *Ivesia* of the Ivesioid clade (Ertter and Reveal, 2014), *Duchesnea* of the Reptans clade (Chaoluan et al. 2003; Ertter and Reveal 2014) and *Argentina* and *Tylosperma* of the Anserina clade were also included. Species of lower ploidy levels were prioritized over those of high ploidy levels when selecting representative species.

To evaluate the taxonomic status and history of the two *P. norvegica* subspecies – *P. norvegica* ssp. *hirsuta* and *P. norvegica* ssp. *norvegica* – (Paper II), herbarium material of one morphologically typical individual of each subspecies were selected from Scandinavia and central Europe, as well as two North American and one eastern Russian specimen of ssp. *hirsuta*.

To investigate the cause of the Rivales species showing different clade relationships (Papers II and III), all Rivales species occurring in North America were included (*P. biennis*, *P. intermedia*, *P. newberryi*, *P. norvegica*, *P. rivalis* and *P. supina*), as were two East Asian Rivales species (*P. centigrana* and *P. cryptotaeniae*).

---

## 3.2 Morphological methods

A number of *Potentilla norvegica* herbarium specimens were investigated to study the defining characters of its two subspecies (*P. norvegica* ssp. *norvegica* and *P. norvegica* ssp. *hirsuta*); leaflet form, leaflet dentation and stipule dentation (Ascherson and Graebner 1904, Lid and Lid 2013, Mossberg and Stenberg 2014) (Paper II).

## 3.3 Molecular methods

### 3.3.1 Molecular markers and primer design

One nrDNA marker, four cpDNA markers and six low-copy nuclear (LCN) markers were used to produce gene trees and species trees; the nuclear ribosomal internal transcribed spacer (ITS), the chloroplast gene maturase K (*matK*) and spacers *trnLF*, *trnC* and *trnSG*, and the LCN genes *DHAR2* (dehydroascorbate reductase 2) (two regions), *GAPCP1* (glyceraldehyde-3-phosphate dehydrogenase), *GBSSI-1* (granule-bound starch synthase I-1), *GBSSI-2* (granule-bound starch synthase I-2) and *SbeI* (starch branching enzyme I). Four new primer pairs for amplification and sequencing were designed for the LCN markers *DHAR2*, with primer sites in exons 1 and 4 and in exons 4 and 5, *GAPCP1*, with primer sites in exons 11 and 14, and *GBSSI-1*, with primer sites in exons 1 and 4. Suitable primer placements were found through alignment of an unpublished genome of *Potentilla argentea* with other *Potentilla* and *Fragaria* sequences available at GenBank.

### 3.3.2 Extraction and production of DNA sequences

DNA extraction from silica gel-dried or herbarium leaf material were performed using the Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) and standard protocols for Polymerase Chain Reactions (PCR) and Sanger sequencing.

To separate the different LCN gene copies that had been amplified in the PCR, molecular cloning was performed on the PCR products from polyploids and specimens failing direct sequencing, using the StrataClone PCR Cloning kit (Agilent

Technologies, Santa Clara, CA, USA). To ensure a 95% probability of finding all gene copies, the number of clones sequenced for each specimen was at least 6 clones for tetraploids, 11 for hexaploids, 16 for octoploids and 21 for decaploids (Lundberg et al., unpublished).

## 3.4 Sequence treatment and alignment

Forward and reverse reads of each species or clone were assembled and proofread using the Staden Package (Staden 1996) and aligned in AliView version 1.18 (Larsson 2014). Identical sequences within species were removed, but those across different species were not.

The alignments of cloned specimens were analyzed in SplitsTree version 4.14.6 (Huson and Bryant 2006) to identify PCR-induced inter-homoeolog recombinants (Marcussen et al. 2015). Those identified as recombinants were removed from the alignments.

## 3.5 Phylogenetic analyses

### 3.5.1 Bayesian Inference and Maximum Likelihood

Codon positions and introns of each marker were coded according to the evolutionary models and partitioning schemes suggested by PartitionFinder2 (Lanfear et al. 2016), while the Mk model (Lewis 2001) was applied for indels.

To infer gene phylogenies, Bayesian Inference (BI) analyses were run in MrBayes version 3.2.6 (Huelsenbeck and Ronquist 2001, Ronquist et al. 2012) under the Metropolis Coupled Markov Chain Monte Carlo algorithm (Yang and Rannala 1997, Altekar et al. 2004), with one cold chain and three heated chains for each of two runs. Sampling from the chain was done every 1000<sup>th</sup> generation and run until the chains had converged (according to the criteria listed in Persson and Rydin 2016).

---

Maximum Likelihood (ML) analyses (Felsenstein 1981) were performed in RAxML version 7.2.8 (Stamatakis 2006) and run for 1000 rapid bootstrap replicates (Felsenstein 1985) (Papers II and III). The GTR+G model (Tavaré 1986, Yang 1993) was applied for the nucleotides and the Mk model (Lewis 2001) was applied for the indels.

In order to test which species were destabilizing the relationships between the major clades (Paper I), a selective taxon removal approach was taken. Five different datasets were tested for each marker; all taxa included, removal of the *Fragarioides* species *P. dickinsii* Franch. & Sav., removal of the *Fragarioides* species *P. fragarioides* L., removal of both *Fragarioides* species, and removal of all species of the Reptans clade.

A clade was considered strongly supported if its Bayesian posterior probability was 0.95 or higher, or if the bootstrap support was 75 or higher.

### 3.5.2 Multispecies Coalescent analyses

To account for incongruences seen in the gene trees, species phylogenies were inferred under the Multispecies Coalescent (MSC) model in \*BEAST (Heled and Drummond 2010), as implemented in BEAST version 1.8.0 (Drummond et al. 2012) (Papers I and II). This type of analysis can account for incomplete lineage sorting (ILS) (Maddison 1997) and different histories of loci. Two different clock models and tree priors were tested using path sampling and stepping stone sampling (Baele et al. 2012, 2013). Two independent analyses with the models and priors best fit to the data were run, as was an additional run with sampling only from the prior to ensure that the data, rather than the priors, were driving the results. The tree files were combined in LogCombiner with a burnin of 20% of each run.

### 3.5.3 Estimation of divergence times

In order to determine the mean ages of the clades in *Potentilla*, an estimation of divergence times was performed on a combined dataset of the three cpDNA markers trnC, trnLF and trnSG (Paper III). Sequences from species of Fragariinae, *Aremonia*

Neck. ex Nestl. and *Rosa* L. were obtained from Genbank. Four fossil calibration points were used, three of which had been used in previous studies. The fourth calibration point represented the stem node of the Ivesioid species and had, as far as known, not been used previously when dating divergence times within Rosaceae (Becker 1961). The analysis procedure was the same as that for the MSC analyses in BEAST, with model and prior testing and sampling from only the prior.

### **3.5.4 Network analyses**

To test whether the Rivales species share ancestral lineages and hybridization events (Paper III), network analyses using the parsimony criterion were performed in PhyloNet version 3.7.3 (Wen et al. 2018), based on the gene tree topologies of DHAR2 and GAPCP1. The number of distinct clades seen for each Rivales species in the trees defined the number of reticulations (hybridizations) that they could theoretically have gone through. Since the analyses were computationally heavy, especially when adding more reticulation events, each species had to be analyzed separately. The analyses were run for 100 iterations and the three most parsimonious trees were returned. To choose the optimal network, the networks were compared to the BI and dated trees.

---

## 4. Results and Discussion

### 4.1 Intrageneric relationships of the major clades in *Potentilla*

#### 4.1.1 Gene trees

Paper I concerned the relationships of the major clades in *Potentilla*, and different relationships were retrieved from each analysis of the complete datasets of the six markers included; nrITS, chloroplast matK and the low-copy nuclear (LCN) markers DHAR2, GAPCP1, GBSSI-2 and SbeI. The clades themselves were supported in a majority of the markers, but the two species included from the Fragarioides clade, *P. dickinsii* and *P. fragarioides*, did not form a monophyletic group in trees based on any marker. The differing relationships were also seen for the datasets where either *P. dickinsii*, *P. fragarioides* or the Reptans clade were removed. However, when excluding both *P. dickinsii* and *P. fragarioides*, the Alba clade was sister to the rest of *Potentilla*, and then the Reptans species resolved as sisters to the Argentea plus Ivesioid clade. Töpel et al. (2011) suggested allopolyploidy as a possible explanation for the differing phylogenies, but *P. dickinsii* and *P. fragarioides* are diploid according to chromosome counts (Rice et al. 2014). Homoploid hybridization could explain their number of chromosomes, but both species had more than two supported positions in the trees. Incomplete lineage sorting (ILS) could therefore be an additional process behind the incongruences.

In the LCN markers, relationships indicating autopolyploid origins of the Ivesioid and Reptans clades were seen. This was especially evident in the GAPCP1 tree, where there were two subclades in both clades and sequences from all Ivesioid and Reptans species, respectively, were resolved in both subclades.

### 4.1.2 Species trees

Two species trees were inferred by Multispecies Coalescent (MSC) analyses in Paper I, one based on the complete datasets and one excluding the *Fragarioides* species. The tree including all taxa was not completely resolved; the nodes where *P. dickinsii* and *P. fragarioides* branched off were not supported. However, when excluding the *Fragarioides* species, the backbone was strongly supported. The internal relationships of the clades were unresolved, and in the case of the Ivesioid and Reptans clades that was probably due to their autopolyploid origins, as indicated in the gene trees.

### 4.1.3 The Ivesioid clade and its taxonomical treatment

Most species of the Ivesioid clade are assigned to the genera *Horkelia*, *Horkeliella* and *Ivesia* in the latest edition of Flora of North America (Ertter and Reveal 2014), but all molecular studies based on cpDNA and nrDNA data that have been performed on *Potentilla* have shown these species to be nested within the genus (Eriksson et al. 1998, Eriksson et al. 2003, Dobeš and Paule 2010, Töpel et al. 2011, and Feng et al. 2017). All molecular markers in Papers I, II and III show the same pattern, and the least cumbersome solution to achieve monophyly would be to include these species in *Potentilla*. In that case, only a handful of species would have to be assigned new names, since most of the species in the clade already have synonyms in *Potentilla* after other authors' classifications, e.g. that of Greene (1887). The alternative would be to divide the about 400 species currently named *Potentilla* into multiple genera. Since the type species is found in the small Reptans clade, *P. reptans*, only these species could retain the name *Potentilla* without having to conserve a new type. The generic delimitation is discussed further below (Paper IV).

## 4.2 The evolutionary history of the *Potentilla* Rivaless group

### 4.2.1 *Potentilla norvegica* and its two subspecies

The herbarium specimens studied for the characters defining the two *P. norvegica* subspecies, *P. norvegica* ssp. *norvegica* and *P. norvegica* ssp. *hirsuta*, were mostly

intermediate in morphology (Paper II). Neither was there any clade seen in the gene trees that was specific to, or excluding, any subspecies or geographic origin of the seven specimens included in the molecular analyses. Thus, there was no support for neither species nor subspecies differentiation, but rather extensive intraspecies morphological variation.

#### **4.2.2 *Potentilla norvegica* and *P. intermedia* have a common evolutionary history**

Between the three gene trees in Paper II, homoeologues of the Rivales species *P. intermedia* and *P. norvegica* were resolved as sisters in four distinct clades. Three of these were found in the Argentea clade, and one in the Ivesioid clade. This pattern suggests that the two species are allopolyploid and share some hybridization events in their evolutionary history. This is the first time that LCN markers have resolved the evolutionary history of species of high ploidy level in *Potentilla*. The markers revealed a much more complicated history than cpDNA or nrDNA markers could ever do, and this points to the importance of using LCN markers in future studies of the evolutionary history of polyploid species in general, and *Potentilla* in particular.

#### **4.2.3 The reticulate relationships of the North American Rivales species**

In Paper III, the dataset in Paper II was extended to include all six species of the Rivales group that occur in North America; *Potentilla biennis*, *P. intermedia*, *P. newberryi*, *P. norvegica*, *P. rivalis* and *P. supina* (Wolf 1908, POWO). In addition to cpDNA and nrDNA data, three LCN markers were analyzed – one of which was also used in Papers I and II (GAPCP1), and two which were also used in Paper I (DHAR2 exons 1-4 and GBSSI-2). The tree of GBSSI-2 was largely unresolved, but the trees of the other two markers were used as input for the network analyses. The close relationship of *P. intermedia* and *P. norvegica* was confirmed with the two additional markers, and some of their homoeologues were also in clades with the other four Rivales species analyzed. No chromosome counting data exist for *P. biennis* and *P. newberryi*, but *P. biennis* could be sequenced without first cloning the LCN markers

and presumably it is therefore diploid or a homozygous polyploid. *Potentilla newberryi* was found in three distinct clades and it is therefore most likely an allopolyploid, and possibly hexaploid.

The network analyses did not indicate that *P. newberryi*, *P. rivalis* or *P. supina* shared ancestral lineages. Thus, none of them gave rise to one another as the extant species we know of today. However, it could be concluded from the gene trees and the network analyses that their precursors of lower ploidy levels may have been involved in the formation of *P. intermedia* and *P. norvegica*. Thus, based on the relationships seen and the current distributions of the species, the most recent common ancestor (MRCA) of *P. intermedia* and *P. norvegica* was most likely formed in North America. Each of *P. intermedia* and *P. norvegica* had homoeologues in two additional distinct clades in which the other species were not present. Hybridization with these lineages presumably occurred in Eurasia for *P. intermedia* and in North America for *P. norvegica*. This goes against the common statement in many floras, that *P. norvegica* is native to eastern Europe and not likely native to North America (Hitchcock and Cronquist 1961, Tutin et al. 1968, Hultén 1971, Kurtto et al. 2004, Lid and Lid 2013).

The results of Papers I, II and III may on a smaller scale have implications for the conservation status of *P. norvegica* as native or introduced in Europe and North America. Some authors have mentioned *P. norvegica* as a weed (Werner and Soule 1976, Mossberg et al. 1992, Lid and Lid 2013), which could influence this assessment. On a larger scale, the polyphyletic origin of established, independently evolving allopolyploids should be taken into account when reassessing generic delimitations in the future.

### 4.3 A suggestion for a monophyletic generic delimitation of *Potentilla*

Based on the relationships inferred in Papers I, II and III, as well as those in previous studies, we suggested the generic delimitation of *Potentilla* to include the Alba, Argentea, Fragarioides, Ivesioid and Reptans clades, but excluding the Anserina clade. The strongest argument for this delimitation, and not a narrower one, is the apparent

---

network of gene flow between some of the clades. The Rivaless species *P. intermedia*, *P. newberryi* and *P. norvegica* were shown to have ancestral lineages in both the Argentea clade and in the Ivesioid clade, while the Fragarioides species *P. dickinsii* and *P. fragarioides* were found in the Reptans and Alba clades in trees based on different markers. *Potentilla fragarioides* was even resolved as sister to the rest of *Potentilla* (excluding the Anserina clade) in one marker.

An exclusion of the Anserina clade was motivated by its uncertain phylogenetic position. With cpDNA, the clade is strongly supported as sister to *Potentilla*, but with nrDNA, the clade is instead weakly supported as sister to the subtribe Fragariinae (Eriksson et al. 2003; Töpel et al. 2011; Feng et al. 2017). In addition, there are morphological characters that separate the species of this clade from the rest of *Potentilla* (Soják 2010). *Argentina* has also become a widely used genus name for the majority of the species of this clade, adopted by recent floras and databases (Krok and Almquist 2012; Lid and Lid 2013, Roskov et al. 2019, POWO).

As mentioned under section 4.1.3, this delimitation also meant that very few species had to be given new names in order for the name *Potentilla* to refer to a monophyletic group. Only six species in the Ivesioid clade, out of the ca 400 in the entire genus, had to be recombined.

## 5. Future perspectives

### 5.1 The putatively autopolyploid origin of the Ivesioid species and the Reptans clade

In the gene trees in Paper I, the internal relationships of the Ivesioid clade (the Rivaless species excluded) and the Reptans clade indicated autopolyploid origins of the clades. This was most evident in the GAPCP1 tree, where both clades were divided into two subclades with one gene copy from each species in both subclades. However, with the same marker but a different taxon set, the two Ivesioid subclades were not resolved in the GAPCP1 tree in Paper III. In that tree, two subclades were supported, but there were also unresolved species outside of these clades. In the other LCN trees, the resolution was too low to be able to say anything for certain about an autopolyploid origin. The only known diploid species of the Reptans clade, *P. flagellaris* D.F.K.Schltdl. (Wolf 1908), has, to our knowledge, never been part of any molecular analysis but could potentially help resolve the origin of this clade. Also in the Reptans clade is *P. indica* (Jacks.) Th.Wolf, which according to chromosome counts has a higher ploidy level than the rest of the species in the clade (6x and 12x instead of 4x), and it is possible that this species has gone through additional rounds of autopolyploidization. In order to resolve the clades' origins, additional LCN markers and taxa are needed, especially for the Ivesioid clade.

### 5.2 The species-rich Argentea clade

Previous phylogenetic analyses based on cpDNA and nrDNA, as well as those in Papers I and III, have been largely unsuccessful in resolving the relationships within the Argentea clade, the most species-rich clade in *Potentilla*. Looking at the datasets presented in the papers in this thesis, there is very little variation within the chloroplast markers. The rapid speciation in this clade could potentially be explained by polyploidization. The analyses based on LCN markers have indeed been more successful in resolving relationships since nuclear markers have a higher evolutionary

---

rate and can detect polyploidization events in a single gene tree. To resolve the clade, but to reduce the complexity of the trees, species of higher ploidy levels should initially be omitted from phylogenetic analyses with LCN markers. Extensive sampling of diploid and tetraploid species could be used to build a backbone phylogeny, in which species with higher ploidy levels could be added later.

### 5.3 Additional allopolyploid species in need of investigation

There seems to be very few *Potentilla* species with only diploid populations, and in addition to the species focused on in this thesis, several other species have also shown indications of being allopolyploid; e.g. *P. aurea*, *P. heptaphylla*, *P. incana* and *P. pennsylvanica* (Papers II, III). The tetraploid North American species *P. pennsylvanica* showed a close relationship to the diploid East Asian species *P. chinensis*. This relationship could be of particular interest, since authors have disagreed on whether Asian *P. pennsylvanica* is a different species than the one occurring in North America (reviewed in Soják 2009). Both *P. pennsylvanica* and *P. chinensis* have pinnate leaves, and an overlapping distribution in East Asia.

### 5.4 Other molecular methods and the need for software programs producing reticulate trees

Molecular cloning followed by Sanger sequencing is the traditional method of separating the different gene copies of polyploids, and this is the method that has been used in all molecular papers in this thesis (Papers I, II, III). Molecular cloning has the advantage of being able to handle sequences from PCR products of several thousand base pairs (although internal primers would have to be used for sequencing), but the process is much more money and time consuming per marker compared to High-Throughput Sequencing (HTS) techniques. Thousands of markers could potentially be targeted per sequencing round with this newer sequencing method. Target Capture is an alternative method to PCR where probes, which would be equivalent to PCR primers, hybridize with target sequences in the genome and separate them from the unwanted sequences. The difficulties lie in assembling the resulting 150-200 base pair

long sequenced fragments, and then to separate them into orthologues and homoeologues. Naturally, this assembly and separation gets harder the higher the ploidy level of the species. To our knowledge it has, so far, rarely been done for polyploids, and it is an open question whether it is a method that can be applied with high polyploids, such as those studied here. However, in order to resolve the complex relationships seen during the work with this thesis, more information is key to be able to answer the remaining phylogenetic questions in *Potentilla*, but also to resolve the evolutionary histories of polyploids in general.

The software used in Paper III to produce reticulate trees, PhyloNet (Wen et al. 2018), constitutes a user-friendly program that takes gene trees as input and infers networks for the hybrid species included. Automated processes are always preferable to manual processes in terms of repeatability, but this program required meticulous manual consideration of the trees when defining the homoeologues in the input file, as well as careful studies of the results and if the orders of hybridization events suggested were plausible. As for now, it is doubtful that analysis of allopolyploids will ever be fully automated, but with more and more people using HTS methods, it will be easier to obtain a good amount of underlying data for this type of analysis. And with that, hopefully the programs for network analysis will be developed further.

## 5.5 Evaluation of the taxonomic treatment of polyploids

When Linnaeus introduced the binomial nomenclature system (Linnaeus 1753), he defined species based on morphological similarity. To him, species were distinct and constant units that had been created by a deity (Uddenberg 2003). Later, species were thought of as dynamic entities, and perhaps the best known species concept today is the biological species concept (Mayr 1942), where “species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups”. Even though all life on Earth stem from the same origin and share basic traits, extant organisms are fundamentally different in their physiology, morphology and ecology. Different organisms are affected by different genetic, biotic and abiotic processes, and what drives speciation in one group may have little effect in

other groups. Thus, as noted by Mayr himself, the biological species concept is not as well-suited for plants as it is for most animals, since plants can reproduce asexually (Stebbins 1950, Benson and Hartnett 2006), and fertile hybrids occur frequently (Ownbey 1950, Grant and Wilken 1988, Arnold 1994).

The International Code of Nomenclature for algae, fungi and plants (Turland et al. 2018) provides guidelines on how to name species. However, it is first up to the researchers to decide through phylogenetic, ecological and morphological studies whether what they have found is a new species or not. The unified species concept by de Queiroz (2005) could be better suited, since it focuses on lineages and shared evolutionary history rather than only reproductive abilities. This concept could theoretically be applied to species with a hybrid origin, since lineages (i.e. species) may fuse. In a wider sense, established, independently evolving hybrids (including both homoploids and allopolyploids) could also affect where delimitations for genera, and even higher ranks, are drawn under a hierarchical naming system, depending on how distantly related their subgenomes are.

## 6. References

- Albertin W, Marullo P. 2012. Polyploidy in fungi: evolution after whole genome duplication. *Proceedings of the Royal Society B* 279:2497-2509
- Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20(3):407-415
- Arnold ML. 1994. Natural hybridization and Louisiana irises. *Bioscience* 44(3):141-147
- Ascencio D, DeLuna A. 2013. Genetic Redundancy In: Dubitzky W, Wolkenhauser O, Cho KH, Yokota H. (Eds.) *Encyclopedia of Systems Biology*. New York: Springer
- Ascherson P, Graebner P. 1904. *Synopsis der Mitteleuropäischen Flora*. Vol. 6 pt.1. Leipzig: Engelmann
- Asker S. 1970. Apomictic biotypes in *Potentilla intermedia* and *P. norvegica*. *Hereditas* 66:101-108
- Brassac J, Jakob SS, Blattner FR. 2012. Progenitor-derivative relationships of *Hordeum* polyploids (Poaceae, Triticeae) inferred from sequences of TOPO6, a nuclear low-copy gene gerion. *PLoS One* 7(3):e33808
- Brochmann C, Brysting AK, Alsos IG, Borgen L, Geundt HH, Scheen AC, Elven R. 2004. Polyploidy in Arctic plants. *Biological Journal of the Linnean Society* 82:521-536
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* 29:2157-2167
- Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution* 30:239-243
- Becker HF. 1961. *Oligocene plants from the upper Ruby River Basin, southwestern Montana*. New York: Geological Society of America 82:112-113
- Bento M, Gustafsson JP, Viegas W, Silva M. 2011. Size matters in Triticeae polyploids: larger genomes have higher remodeling. *Genome* 54:175-183
- Bouetard A, Lefeuvre P, Gigant R, Bory S, Pignal M, Besse P, Grisoni M. 2010. Evidence of transoceanic dispersion of the genus *Vanilla* based on plastid DNA phylogenetic analysis. *Molecular Phylogenetics and Evolution* 55:621-630
- Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. 2000. Phenotypic instability and rapid silencing in newly formed *Arabidopsis* allotetraploids. *The Plant Cell* 12:1551-1567
- Cui L, Kerr Wall P, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumunganathan K, Bakarat A, Albert VA, Ma H, de Pamphilis CW. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16:738-749

- Dobeš C, Paule J. 2010. A comprehensive chloroplast DNA-based phylogeny of the genus *Potentilla* (Rosaceae): Implications for its geographic origin, phylogeography and generic circumscription. *Molecular Phylogenetics and Evolution* 56:156-175
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUTi and the BEAST 1.7. *Molecular Biology and Evolution* 29:1969-1973
- Embley TM, Hirt RP, Williams DM. 1994. Biodiversity at the molecular level: The domains, kingdoms and phyla of life. *Philosophical Transactions: Biological Sciences* 345(1311):21-33
- Eriksen B. 1996. Mating systems in two species of *Potentilla* from Alaska. *Folia Geobotanica & Phytotaxonomica* 31(3):333-344
- Eriksson JS, Blanco-Pastor JL, Sousa F, Bertrand YJK, Pfeil BE. 2017. A cryptic species produced by autopolyploidy and subsequent introgression involving *Medicago prostrata* (Fabaceae). *Molecular Phylogenetics and Evolution* 107:367-381
- Eriksson T, Donoghue MJ, Hibbs MS. 1998. Phylogenetic analysis of *Potentilla* using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of Rosoideae (Rosaceae). *Plant Systematics and 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 Sciences* 164(2):197-211
- Eriksson T, Lundberg M, Töpel M, Östensson P, Smedmark JEE. 2015. *Sibbaldia*: a molecular phylogenetic study of a remarkably polyphyletic genus in Rosaceae. *Plant Systematics and Evolution* 301:171-184
- Ertter B, Elven R, Reveal JL, Murray DF. 2014. *Potentilla*. In: Flora of North America Editorial Committee (Eds.) *Flora of North America North of Mexico*. Vol. 9. New York: Oxford University Press; pp. 121-218
- Ertter B, Reveal JL. 2014. *Ivesia, Horkelia & Horkeliella*. In: Flora of North America Editorial Committee (Eds.) *Flora of North America North of Mexico*. Vol. 9. New York: Oxford University Press; pp. 219-272
- Evans RC, Campbell CS. 2002. The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *American Journal of Botany* 89(9):1478-1484
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17:368-376
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39(4):783-791
- Feng T, Moore MJ, Yan MH, Sun YX, Zhang HJ, Meng AP, Li XD, Jian SG, Li JQ, Wang HC. 2017. Phylogenetic study of the tribe Potentilleae (Rosaceae), with further insight into the disintegration of *Sibbaldia*. *Journal of Systematics and Evolution* 55(3):177-191
- Ford VS, Gottlieb LD. 1999. Molecular characterization of PgiC in a tetraploid plant and its diploid relatives. *Evolution* 53(4):1060-1067
- Grant V. 1971. *Plant speciation*, 1<sup>st</sup> ed. New York: Columbia University Press

- Grant V, Wilken DH. 1988. Natural hybridization between *Ipomopsis aggregata* and *I. tenuituba* (Polemoniaceae). *Botanical Gazette* 149(2):213-221
- Greene EL. West American phases of the genus *Potentilla*. In: *Pittonia. A series of papers relating to botany and botanists*. Vol. 1. Berkeley: Doxey & Co.; pp. 95-106
- Guggisberg A, Mansion G, Kelso S, Conti E. 2006. Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid-polyploid species complex of *Primula*. *New Phytologist* 171(3):617-632
- Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27(3):570-580
- Hitchcock CL, Cronquist A. 1961. *Vascular plants of the Pacific northwest. Part 3: Saxifragaceae to Ericaceae*. Seattle: University of Washington Press
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755
- Hultén E. 1971 *The circumpolar plants. II – Dicotyledons. Kungliga Vetenskapsakademiens Handlingar*. Fjärde serien, band 13, nr. 1. Stockholm: Almqvist & Wiksell
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2):254-267
- Kalkman C. 2004. *Potentilla*. In: Kubitzki K, Ed. *Flowering plants – Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*. Berlin: Springer; p. 366
- Kamneva OK, Syring J, Liston A, Rosenberg NA. 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evolutionary Biology* 17:180
- Kechaykin AA, Shmakov AI. 2016. A system of subtribe Potentillinae J. Presl (Rosaceae Juss.). *Turczaninowia* 19(4):114-128
- Kihara H, Ono T. 1926. Chromosomenzahlen und systematische Gruppierung der Rumex-Arten. *Zeitschrift für Zellforschung und Mikroskopische Anatomie* 4:475-481
- Krok TOBN, Almquist S. 2012. *Svensk flora. Fanerogamer och kärlkryptogamer*. 29<sup>th</sup> ed. Jonsell L, Jonsell B (Eds.) Stockholm: Liber
- Kurtto A, Lampinen R, Junikka L. 2004. *Atlas Florae Europaeae. Distribution of vascular plants in Europe*. Vol. 13. Helsinki: Societas Biologica Fennica Vanamo
- Lanfear R, Frandsen, PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34(3):772-773
- Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22):3276-3278
- Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploids. *Biological Journal of the Linnean Society* 82:651-663
- Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50(6):913-925
- Lid J, Lid DT. 2013. *Potentilla* L. In: Elven R (Ed.) *Norsk flora*, 7th ed. Oslo: Det Norske Samlaget; pp. 413-423

- Linnaeus C. 1753. *Species plantarum*. Vol. 1. Stockholm: Salvius
- Lundberg M, Töpel M, Eriksen B, Nylander JAA, Eriksson T. Allopolyploidy in Fragariinae (Rosaceae): Comparing four DNA sequence regions, with comments on classification. *Molecular Phylogenetics and Evolution* 51:269-280
- Lutz AM. 1907. A preliminary note on the chromosomes of *Oenothera lamarckiana* and one of its mutants, *O. gigas*. *Science* 26(657):151-152
- Mabberley DJ. 2002. *Potentilla* and *Fragaria* (Rosaceae) reunited. *Telopea* 9(4):793-801
- Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* 46(3):523-536
- Mandáková T, Pouch M, Harmanová K, Zhan SH, Mayrose I, Lysak MA. 2017. Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Molecular Ecology* 26:6445-6462
- Manzoor A, Ahmad T, Bashir MA, Hafiz IH, Silvestri C. 2019. Studies on colchicine induced chromosome doubling for enhancement of quality traits in ornamental plants. *Plants* 8(7):194
- Marcussen T, Heier L, Brysting AK, Oxelman B, Jakobsen KS. 2015. From gene trees to a dated allopolyploid network: Insights from the Angiosperm genus *Viola* (Violaceae). *Systematic Biology* 64(1):84-101
- Marcussen T, Jakobsen KS, Danihelka J, Ballard HE, Blaxland K, Brysting AK, Oxelman B. 2012. Inferring species networks from gene trees in high-polyploid North American and Hawaiian violets (*Viola*, Violaceae). *Systematic Biology* 61(1):107-126
- Mason AS, Batley J. 2015. Creating new interspecific hybrid and polyploid crops. *Trends in Biotechnology* 33(8):436-441
- Mason AS, Pires JC. 2015. Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends in Genetics* 31(1):5-10
- Matsuoka Y, Nasuda S, Ashida Y, Nitta M, Tsujimoto H, Takumi S, Kawahara T. 2013. Genetic basis for spontaneous hybrid genome doubling during allopolyploid speciation of common wheat. Shown by natural variation analyses of the paternal species *PloS ONE* 8(8):e68310
- Mayr E. 1942. *Systematics and the origin of species*. New York: Columbia University Press
- McIver J, Erickson K. 2012. Pollination biology of *Potentilla recta* (Sulfur Cinquefoil) and its cooccurring native congener *Potentilla gracilis* in northeastern Oregon. *Psyche* 281732
- 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
- Mossberg B, Stenberg L, Ericsson S. 1992. *Den nordiska floran*. Stockholm: Wahlström & Widstrand
- Nixon KC, Wheeler QD. 1990. An amplification of the Phylogenetic Species Concept. *Cladistics* 6:211-223
- Ownbey M. 1950. Natural hybridization and amphiploidy in the genus *Tragopogon*. *American Journal of Botany* 37(7):487-499

- Persson NL, Rydin C. 2016. Phylogenetic relationships of the 'Briza complex' to other members of the subfamily Pooideae (Poaceae). *Plant Ecology and Evolution* 149(2):216-227
- Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA, Campbell CS. 2007. Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution* 266(1-2):5-43
- POWO. 2020. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; <http://www.plantsoftheworldonline.org/> Retrieved 27 May 2020
- Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics* 3:827-837
- de Queiroz K. 2005. A unified concept of species and its consequences for the future of taxonomy. *Proceedings of the California Academy of Science* 56(18):196-215
- Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology, Evolution, and Systematics* 29:467-501
- Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annual Review of Ecology, Evolution, and Systematics* 33:589-639
- Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. 2014. The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. *New Phytologist* 206(1):19-26
- Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* 317(5840):910-914
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61(3):539-542
- Roskov Y, Ower G, Orrell T, Nicolson D, Bailly N, Kirk PM, Bourgoin T, DeWalt RE, Decock W, van Nieukerken E, Zarucchi J, Penev L (Eds.). 2019. *Species 2000 & ITIS Catalogue of Life, 2019 Annual Checklist*. Digital resource at [www.catalogueoflife.org/annual-checklist/2019](http://www.catalogueoflife.org/annual-checklist/2019). Species 2000: Naturalis, Leiden, the Netherlands
- Rydberg KA. 1898. *A monograph of the North American Potentilleae*. Vol. 2. Lancaster: The New Era Printing Company
- Séguier, JF. 1754. *Plantarum quae in agro Veronensi reperiuntur*. Vol. 3, supplementum. Verona: Andreonius
- Small RL, Cronn RC, Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17:145-170
- Soják J. 2008. Notes on *Potentilla* XXI. A new division of the tribe Potentilleae (Rosaceae) and notes on generic delimitations. *Botanische Jahrbücher für Systematik* 127(3):349-358
- Soják J. 2009. *Potentilla* L. (Rosaceae) in the former USSR; second part: comments; Notes on *Potentilla* XXIV. *Feddes Repertorium* 120(3-4):185-217
- Soják J. 2010. *Argentina* Hill., a genus distinct from *Potentilla* (Rosaceae). *Thaiszia – Journal of Botany* 20:91-97

- 
- Staden R. 1996. The Staden sequence analysis package. *Molecular Biotechnology* 5:233
- Stebbins GL. 1950. *Variation and evolution in plants*. New York: Columbia University Press
- 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
- Lectures on Mathematics in the Life Sciences 17(2): 57–86. Timberlake WE. 2013. Heterosis. In: Maloy S, Hughes K (Eds.) *Brenner's Encyclopedia of Genetics*, 2<sup>nd</sup> ed. San Diego: Academic Press; pp. 451-453
- Turland NJ, Wiersema JH, Barrie FR, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Kusber W-H, Li D-Z, Marhold K, May TW, McNeill J, Monro AM, Prado J, Price MJ, Smith GF (Eds.). 2018: *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. Regnum Vegetabile 159. Glashütten: Koeltz Botanical Books
- Tutin TG, Heywood VH, Burges NA, Valentine DH, Walters SM, Webb DA. 1968. *Flora Europaea, Rosaceae to Umbelliferae*. Vol. 2. London: Cambridge University Press
- Töpel M, Lundberg M, Eriksson T, Eriksen B. 2011. Molecular data and ploidal levels indicate several putative allopolyploidization events in the genus *Potentilla* (Rosaceae). *PLoS currents* 3
- Uddenberg N. 2003. *Idéer om livet, Band I. En levande materia, Mångfald och utveckling*. Stockholm: Natur och Kultur
- Vamosi JC, Dickinson TA. 2006. Polyploidy and diversification: a phylogenetic investigation in Rosaceae. *International Journal of Plant Sciences* 167(2):349-358
- Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. *Systematic Biology* 67:735-740
- Werner PA, Soule DJ. 1976. The biology of Canadian weeds. 18. *Potentilla recta* L., *P. norvegica* L., *P. argenta* L. *Canadian Journal of Plant Science* 56:591-603
- Wolf T. 1908. *Monographie der Gattung Potentilla*. Stuttgart: *Bibliotheca Botanica*, 16(Heft 71): 1-714
- Xiang Y, Huang C-H, 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* 34:262-281
- Yang Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* 10:1396-1401
- Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular Biology and Evolution* 14:717-724.
- Zhang S-D, Jin J-J, Chen S-Y, Chase MW, Soltis DE, Li H-T, Yang J-B, Yi T-S. 2017. Diversification of Rosaceae since the late Cretaceous based on plastid phylogenomics. *New Phytologist* 214:1355-1367



---

# Paper I





## STUDIES

# Detecting destabilizing species in the phylogenetic backbone of *Potentilla* (Rosaceae) using low-copy nuclear markers

Nannie L. Persson, Ingrid Toresen, Heidi Lie Andersen, Jenny E. E. Smedmark and Torsten Eriksson\*

Department of Natural History, University Museum, University of Bergen, Postboks 7800, NO-5020 Bergen, Norway

\*Corresponding author's e-mail address: [Torsten.Eriksson@uib.no](mailto:Torsten.Eriksson@uib.no)

Associate Editor: Jeremy Beaulieu

Citation: Persson NL, Toresen I, Andersen HL, Smedmark JEE, Eriksson T. 2020. Detecting destabilizing species in the phylogenetic backbone of *Potentilla* (Rosaceae) using low-copy nuclear markers. *AoB PLANTS* 12: plaa017; doi: 10.1093/aobpla/plaa017

## Abstract

The genus *Potentilla* (Rosaceae) has been subjected to several phylogenetic studies, but resolving its evolutionary history has proven challenging. Previous analyses recovered six, informally named, groups: the Argentea, Ivesioid, Fragarioides, Reptans, Alba and Anserina clades, but the relationships among some of these clades differ between data sets. The Reptans clade, which includes the type species of *Potentilla*, has been noticed to shift position between plastid and nuclear ribosomal data sets. We studied this incongruence by analysing four low-copy nuclear markers, in addition to chloroplast and nuclear ribosomal data, with a set of Bayesian phylogenetic and Multispecies Coalescent (MSC) analyses. A selective taxon removal strategy demonstrated that the included representatives from the Fragarioides clade, *P. dickinsii* and *P. fragarioides*, were the main sources of the instability seen in the trees. The Fragarioides species showed different relationships in each gene tree, and were only supported as a monophyletic group in a single marker when the Reptans clade was excluded from the analysis. The incongruences could not be explained by allopolyploidy, but rather by homoploid hybridization, incomplete lineage sorting or taxon sampling effects. When *P. dickinsii* and *P. fragarioides* were removed from the data set, a fully resolved, supported backbone phylogeny of *Potentilla* was obtained in the MSC analysis. Additionally, indications of autopolyploid origins of the Reptans and Ivesioid clades were discovered in the low-copy gene trees.

**Keywords:** Autopolyploidy; Fragarioides; incomplete lineage sorting; Multispecies Coalescent; Potentilleae.

## Introduction

Polyploidy is a well-known and common phenomenon in plants, defined as having three or more complete sets of chromosomes. All extant species of flowering plants may in fact be paleopolyploids, as a result of whole-genome duplications early in the history of the angiosperms (Cui *et al.* 2006). However, through a number of different processes resulting in genomic reorganizations, many species with polyploidy in their ancestry

now function as diploids (Leitch and Bennett 2004; Bento *et al.* 2011; Mandáková *et al.* 2017). The genus *Potentilla* (Rosaceae) consists of ~400 species which are mainly yellow-flowered, herbaceous perennials from the Northern Hemisphere. There are diploid as well as polyploid species (Index to Plant Chromosome Numbers, IPCN 1979; Kurtto *et al.* 2004), with ploidy levels of up to hexadecaploid (16x) (Kalkman 2004), and a base chromosome

Received: 13 December 2019; Editorial decision: 13 March 2020; Accepted: 4 May 2020

© The Author(s) 2020. Published by Oxford University Press on behalf of the Annals of Botany Company.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

number of 7. Polyploidization as well as hybridization are considered important processes in the evolution of *Potentilla* (Potter et al. 2007; Dobeš and Paule 2010; Paule et al. 2011, 2012).

In the latest monograph of *Potentilla*, Wolf (1908) identified just over 300 species and divided them into six subsections based on style shape and its position on the ovary. Even though the first molecular studies of *Potentilla* showed that the genus was not monophyletic as circumscribed by Wolf (Eriksson et al. 1998, 2003), recent classifications maintain a non-monophyletic *Potentilla* by recognizing the genera *Horkelia*, *Horkeliella*, *Ivesia* and *Duchesnea* (Chaoluan et al. 2003; Ertter and Reveal 2014a, b; Kechaykin and Shmakov 2016). Although certain aspects of their morphology differ from most other *Potentilla* species, molecular studies have consistently shown that these genera are nested within the *Potentilla* clade (Eriksson et al. 1998, 2003; Dobeš and Paule 2010; Töpel et al. 2011; Feng et al. 2017; Xiang et al. 2017; Persson et al. 2020).

The phylogenetic study of *Potentilla* by Töpel et al. (2011), based on chloroplast and nuclear ribosomal data, identified six major clades that were informally named the Argentea, Ivesioid, Fragarioides, Reptans, Alba and Anserina clades. They found the style type character used by Wolf (1908) to be informative, largely corresponding to the different clades. Using the same type of molecular data, Dobeš and Paule (2010) and Feng et al. (2017) also recovered these clades. However, not all of the clades are well-supported, nor are the relationships between them certain. One of the most prominent incongruences concerns the Reptans clade and its position in relation to the Fragarioides clade. The Reptans clade includes the type species of *Potentilla*, *P. reptans*, and corresponds to 'Grex' Tormentillae in the monograph by Wolf (1908). It comprises eight species that are found in Europe, Asia and North America (Global Biodiversity Information Facility, GBIF Secretariat 2019), characterized by having long pedicels (Wolf 1908). All species but one are polyploid (IPCN) and they form a clade in previous phylogenetic analyses (Eriksson et al. 1998, 2003; Dobeš and Paule 2010; Töpel et al. 2011; Feng et al. 2017). Grex Fragarioides comprises, according to Wolf (1908), two species; *P. fragarioides* and *P. freyniana*, characterized by pinnate leaves where the three terminal leaflets are much larger than the proximal leaflets. Töpel et al. (2011) associated two additional species with this clade; *P. dickinsii* in Grex Eriocarpaceae, characterized by the indumentum of the fruits (Wolf 1908) and *P. stolonifera* (Grex Fragarioides, as *P. fragarioides* var. *stolonifera*). These four species are found in East Asia (GBIF) and are diploid according to published chromosome counts (IPCN).

Reconstructing species phylogenies with chloroplast DNA can be problematic with polyploids (and allopolyploids in particular), since chloroplast DNA is uniparentally inherited and therefore not able to recover polyploid signals. Similarly, nuclear ribosomal DNA is typically subject to concerted evolution with homogenization towards either the maternal or the paternal lineage (Wendel 2000). In certain cases, discrepancies seen between chloroplast and nuclear ribosomal phylogenies may be explained by hybridization and diversification of fertile hybrids or by allopolyploidization (Lundberg et al. 2009; Töpel et al. 2011). Low-copy nuclear (LCN) markers are better candidates for resolving relationships where the species are known to be polyploid. This is because subgenome-specific copies are, at least initially after a polyploidization event, present in each subgenome, inherited biparentally and less influenced by concerted evolution (Small et al. 2004). Several studies have used LCN markers to resolve phylogenetic relationships, and to trace polyploidization and hybridization events, at different taxonomic levels within Rosaceae, such as the Maloideae subfamily (Evans and Campbell

2002), subtribe Geinae (Smedmark et al. 2005), *Prunus* (Shi et al. 2013) and *Potentilla* (Persson et al. 2020). However, LCN markers have so far not been used to resolve the phylogenetic backbone structure of *Potentilla*. A robust backbone is of great benefit to future studies within *Potentilla*, as a basis for studies of historical biogeography or for classification. It can also be used to select proper outgroups when investigating internal relationships of the subclades. Lastly, certain flower and leaf characteristics have been used in classifications of tribe Potentilleae, and we need this backbone in order to more securely trace the evolution of such characteristics on the branches of the phylogeny.

The aim of this study is to (i) infer the backbone phylogeny of *Potentilla* and (ii) to identify underlying sources of incongruence between conflicting topologies. We present four gene trees based on LCN markers and compare our results with chloroplast and nuclear ribosomal phylogenies. In addition, two species trees are presented, showing a supported backbone after the sources of incongruence are removed.

## Materials and Methods

### Plant material

Twenty-four specimens from 19 species (including subspecies) were selected to represent the six major clades identified in recent studies of *Potentilla* (Dobeš and Paule 2010; Töpel et al. 2011; Feng et al. 2017), including species that have been classified in the genera *Horkelia*, *Horkeliella* and *Ivesia* of the Ivesioid clade (Ertter and Reveal 2014a), *Duchesnea* of the Reptans clade (Chaoluan et al. 2003; Ertter and Reveal 2014b) as *P. indica* in this study and *Argentina* and *Tylosperma* of the Anserina clade (Table 1). Plant material for DNA extraction was obtained from botanical gardens (Bergius Botanic Garden Stockholm, Bonn University Botanic Gardens, The Linnéan Gardens of Uppsala and Royal Botanic Garden Edinburgh) and herbaria (BG, E, GB, MARY, O, S and UPS).

### DNA extraction

DNA was extracted from 20 mg of dried leaves using the DNeasy Plant Mini Kit (Qiagen). In order to increase the amount of extracted DNA, the samples were left to lyse at 59 °C overnight before increasing the temperature to 65 °C.

### Genetic markers and DNA amplification

One chloroplast and five nuclear markers were analysed in this study; the chloroplast gene *maturase K* (*matK*), the nuclear ribosomal internal transcribed spacer (ITS) and the LCN genes dehydroascorbate reductase 2 (*DHAR2*), glyceraldehyde-3-phosphate dehydrogenase (*GAPCP1*), granule-bound starch synthase 1-2 (*GBSSI-2*) and starch-branching enzyme 1 (*Sbe1*). The forward and reverse strands of the genomes of *Fragaria vesca* (Shulaev et al. 2011) and *P. micrantha* (Buti et al. 2018) were searched through for the LCN primer sequences [see Supporting Information—Table S1]. Primer specificity was assessed by using the Search for Motifs option in Geneious version 10.2.3 (Markowitz et al. 2012), allowing for up to three mismatches.

DNA was amplified in a mixture of 1–20 ng total DNA, 1× Ex Taq Buffer, 0.2 mM of each dNTP, 0.4 μM of each primer, 0.75 U TaKaRa Ex Taq Hot Start Version and dH<sub>2</sub>O to a total volume of 25 μL. The PCR thermal cycling was run on a C1000 Touch thermal cycler (Bio-Rad Laboratories). Amplification of *matK*, ITS, *GAPCP1*, *GBSSI-2* and *Sbe1* was performed using a touchdown PCR procedure, starting with a 3 min initial denaturation at 94 °C. Then, 11 cycles of 45 s denaturation at 94 °C, 30 s of successively



Table 1. Continued

Taxon	Clade	Ploidy level	Voucher	Accession number			DHA2	GAPCP1	GBSSI-2	Sbel
				matK	ITS	trnT-trnL				
<i>Potentilla erecta</i>	Reptans	4x <sup>a,b</sup>	Eriksson 1060 (BG)	MT134119	MT112956	MT134194(B) MT134195(C)	MT134154(3) MT134155(4) MT134156(16) MT134157(1-7)	MT134229(2) MT134230(3) MT134231(9) MT134238(1-1) MT134239(1-9) MT134240(1-11) MT134241(11-3)	MT134136	
<i>Potentilla indica</i>	Reptans	10x, 12x <sup>a,b</sup>	Cult Bergen Museum Garden; Eriksson 1092 (BG)	-	MT112960	-	-	MT134239(1-9) MT134240(1-11) MT134241(11-3) MT134242(11-5) MT134243(11-8)	MT134140	
<i>Potentilla indica</i>	Reptans	10x, 12x <sup>a,b</sup>	Eriksson 1061 (BG)	MT134124	MT112961	MT134205(11-W) MT134221(11-Bb) MT134206(11-Cb) MT134207(11-Kb) MT134222(T) MT134223(Y)	MT134158(11-2) MT134159(11-13) MT134160(11-19) MT134161(11-20) MT134162(11-23)	MT134240(1-11) MT134241(11-3) MT134242(11-5) MT134243(11-8)	MT134141	
<i>Potentilla reptans</i>	Reptans	4x <sup>a,b</sup>	Salvesen 16.45 (BG)	MT134128	MT112967	MT134215(G) MT134216(H) MT134217(I) MT134218(K) MT134219(O) MT134220(P)	MN946885(F) MT134175(1)	MT134254(11) MT134255(12) MT134256(823)	MT134146	
<i>Potentilla simplex</i>	Reptans	-	Eriksson 797 (S)	MT134129	MT112968	MT134218(K) MT134219(O) MT134220(P)	MT134176(2) MT134177(4) MT134178(10) MT134179(11) MT134150	MT134257(5) MT134258(7) MT134259(10) MT134260(11) MT134224	-	
<i>Potentilla biflora</i> <i>Potentilla biflora</i>	Alba Alba	2x <sup>a</sup> 2x <sup>a</sup>	Viereck 5042 (S) Gabrielsen & Jørgensen (O)	- -	- -	MT134115	MT134115	MT134257(5) MT134258(7) MT134259(10) MT134260(11) MT134224	MT134132	
<i>Potentilla micrantha</i>	Alba	2x <sup>a,b</sup>	Cult The Linnaean Gardens of Uppsala 1972- 1035; Kårehed 432 (UPS)	MT134126	MT112964	MT134212	MT134168	MT134250	MT134143	
<i>Potentilla sterilis</i>	Alba	4x <sup>a,b</sup>	Eriksson 734 (S)	MT134130	MT112969	-	MT134180(2) MT134181(7)	MT134261(3) MT134262(7)	MT134147	
<i>Argentina anserina</i>	Anserina (outgroup)	4x, 5x, 6x <sup>a,b</sup>	H. Andersen BG/S- 165236 (BG)	MT134115	MT112950	MT134182(C) MT134183(C) MT134184(F)	MT134149(6) MT134148(9)	-	MT134131	
<i>Tylosperma lignosa</i>	Anserina (outgroup)	-	Mats Andersson 132 (GB)	MT134125	MT112963	MT134210(F) MT134211(O)	MT134166(5) MT134167(11)	MT134247(1) MT134248(6) MT134249(12)	-	

decreasing annealing temperatures starting at 55 °C with 0.5 °C decrement per cycle and 1 min extension at 72 °C. This was followed by 36 cycles of 45 s denaturation at 94 °C, 30 s annealing at 49 °C and 1 min extension at 72 °C, and a 7 min final extension at 72 °C. Amplification of DHAR2 was performed at higher annealing temperatures, starting with a 3 min initial denaturation at 94 °C. Then, 16 cycles of 45 s denaturation at 94 °C, 30 s of successively decreasing annealing temperatures starting at 65 °C with 0.5 °C decrement per cycle and 1 min extension at 72 °C. This was followed by 31 cycles of 45 s denaturation at 94 °C, 30 s annealing at 55 °C and 1 min extension at 72 °C, and a 7 min final extension at 72 °C. The primers used for the different markers are given in [Supporting Information—Table S1](#).

### Cloning

The amplified fragments of matK and ITS displayed no or little intra-species variation and did not need cloning. This was also true for the LCN marker SbeI, and since the other three LCN markers did not show any indications of hybridization between the major clades (see Bayesian inference section), SbeI was not cloned.

PCR products from DHAR2, GAPCP1 and GBSSI-2 of species known to be polyploid or failing direct sequencing were cloned using the StrataClone PCR Cloning Kit (Agilent) following the manufacturer's instructions. Cloned DNA was amplified in a second PCR in the same mixture as described above, only replacing DNA extract with transformed cells. The universal primers M13 forward and M13 reverse were used to amplify the cloning vector, with a 10 min initial denaturation at 94 °C, 35 cycles of 45 s denaturation at 94 °C, 45 s annealing at 55 °C and 3 min extension at 72 °C, and a 10 min final extension at 72 °C.

### Purification and sequencing

All PCR products were purified using Exosap-IT (GE Healthcare), following the manufacturer's instructions. The number of clones sequenced for each specimen was at least 6 for tetraploids, 11 for hexaploids and 21 for decaploids, corresponding to 95 % probability of finding all gene copies (Lundberg et al. manuscript). The amplification primers were also used for sequencing. Sequencing reactions were performed using the BigDye Terminator Cycle Sequencing kit (Applied Biosystems) according to the manufacturer's instructions. DNA was sequenced using an ABI Prism 3730XL DNA analyser (Applied Biosystems). All labwork was performed in the Biodiversity Lab and Sequencing Lab at the University of Bergen, Norway.

### Sequence treatments

The Staden Package (Staden 1996) and AliView v. 1.18 (Larsson 2014) were used for sequence proof reading, assembly and alignment. Scoring of uncertain or polymorphic sites was done with standard IUPAC codes. All sequences were first aligned automatically using MUSCLE (Edgar 2004), followed by manual adjustments. To identify PCR-induced inter-homoeolog recombinants (Marcussen et al. 2015), the sequences of cloned specimens were analysed in SplitsTree v. 4.14.6 (Huson and Bryant 2006). Those identified were removed from the alignments. All sequences have been submitted to GenBank (Table 1) and alignments have been submitted to Dataverse NO (<https://doi.org/10.18710/XRQEKH>).

### Model testing and Bayesian inference

Phylogenies for the individual markers were reconstructed by Bayesian inference (BI; Yang and Rannala 1997) with MrBayes v. 3.2.6 (Huelsenbeck and Ronquist 2001; Ronquist et al. 2012)

using the MC<sup>3</sup> algorithm (Altekar et al. 2004). The alignments of matK, DHAR2, GAPCP1, GBSSI-2 and SbeI were divided in up to five character sets each, corresponding to codon positions (3), introns (1) and indels (1). Boundaries for exons and introns were found by alignment with annotated *Fragaria* sequences from GenBank (Shulavev et al. 2011) and indels were coded according to the simple indel coding method of Simmons and Ochoterena (2000). Partitioning schemes and their models were based on the results from PartitonFinder2 (Lanfear et al. 2016) under the AICc criterion for models available in MrBayes. The Mk model (Lewis 2001) was used for the coded indels. Analyses were investigated for chain stationarity and accepted if the following criteria were fulfilled: the standard deviation of split frequencies was below 0.01, the chain swap was between 20 and 80 % (McGuire et al. 2007), there was no trend seen in the overlay plot and the Potential Scale Reduction Factor values (Gelman and Rubin 1992) had reached 1.0 for all parameters. The analyses were run for 5 million generations, every 1000th generation was sampled and burn-in was set to 25 or 30 %. Additional analyses were run using the same methods, taking a selective taxon removal approach by excluding either *P. dickinsii*, *P. fragarioides*, both *P. dickinsii* and *P. fragarioides* (of the *Fragarioides* clade), or the species of the Reptans clade, to test how this would affect the phylogeny. The trees were rooted on the Anserina clade, since it has been shown to be an outgroup to *Potentilla* (Eriksson et al. 2003; Töpel et al. 2011; Feng et al. 2017).

### Multispecies Coalescent analyses

Species phylogenies were inferred under the Multispecies Coalescent (MSC) model to account for ancestral polymorphisms and conflicts seen in the gene trees. The MSC model can take incomplete lineage sorting (ILS) into account, but not reticulations or gene duplication and loss (GDL) (Bravo et al. 2019). One ortholog is expected per set of chromosomes, and therefore we expected a single amplified fragment per chromosome set (if minor allelic variation is disregarded). Thus, for each species, the number of gene variants should be less or equal to their ploidy level (Table 1). There were no indications of reticulations in our gene trees, nor any indication of paralogs, since the expected number of gene variants was not exceeded in any species (see Bayesian inference section). Thus, we assumed that our sample did not violate the MSC model. The MSC analyses were run in \*BEAST (Heled and Drummond 2010), as implemented in BEAST v. 1.8.0 (Drummond et al. 2012) using the same alignments as in the BI analyses. Two data sets were analysed, one including *P. dickinsii* and *P. fragarioides*, and one excluding them. The data sets comprised 19 and 17 species, respectively, in which *P. dickinsii* and *P. ancistrifolia* var. *dickinsii* were designated as the same species (Takeda 1911), as were *Ivesia kingii* and *Ivesia kingii* var. *eremica* (Ertter 1989). The substitution model for each marker was selected using PartitionFinder2 (Lanfear et al. 2016) under the AICc criterion for models available in BEAST. For each data set, two clock models were tested; strict and relaxed uncorrelated log normal (Drummond et al. 2006). For each clock model, two tree priors were tested; a birth-death process (Kendall 1948) and a birth process (Yule 1924). The analyses were run for 150 million generations, with sampling from the chain every 1000th generation, and rooted on the Anserina clade. To test the fit of the models to the data, path sampling and stepping-stone sampling (Baele et al. 2012, 2013) were performed with 150 steps with a length of 1 million iterations each. Log marginal likelihood differences larger than three were considered significant (Kass and Raftery 1995). Two independent analyses were run using the best-fitting models, and the results were inspected using Tracer

v. 1.7.1.1 (Rambaut et al. 2018). To test that the prior did not have stronger influence over the results than the data, an additional run with sampling from prior only was performed. The tree files from the independent runs of each data set were combined using LogCombiner of the BEAST package with a burn-in of 25 % of each run. PartitionFinder2, MrBayes and BEAST were run at the CIPRES Science Gateway (Miller et al. 2010).

## Results

### Genetic markers

The search for the primer sites in the published genomes of *F. vesca* (Shulaev et al. 2011) and *Potentilla micrantha* (Buti et al. 2018) generated only one hit in each genome for DHAR2, GAPCP1, GBSSI-2 and SbeI, confirming their specificity.

### Bayesian inference

Models and partitioning schemes for the BI analyses are found in Supporting Information—Table S2. Supported clades are defined as having a posterior probability (pp) of  $\geq 0.95$ .

The matK tree with all species included (Fig. 1A) recovers the Argentea, Ivesioid and Reptans clades (all pp 1.0). The Alba species are in unresolved positions to the rest of the ingroup (pp 0.94), in which the Reptans clade is sister to a clade (pp 1.0) that consists of *P. dickinsii*, *P. fragariooides*, the Argentea clade and the Ivesioid clade. *Potentilla fragariooides*, Argentea and the Ivesioids are in a trichotomy (pp 1.0). Excluding only *P. dickinsii* [see Supporting Information—Fig. S1] reduces the posterior probability for the clade of Reptans, *P. fragariooides*, Argentea and

the Ivesioids from 0.94 to 0.51. When *P. fragariooides* is excluded [see Supporting Information—Fig. S2], there are only small changes in the posterior probabilities of the tree, and the same is true in the tree in which the Reptans clade is excluded [see Supporting Information—Fig. S4]. Exclusion of both *P. fragariooides* and *P. dickinsii* [see Supporting Information—Fig. S3] collapses the clade of Argentea, the Ivesioids and Reptans.

The ITS tree with all species included (Fig. 1B) recovers the Argentea, Ivesioid and Reptans clades (all pp 1.0). Apart from the Argentea and Ivesioid clades being sisters (pp 1.0), there is no other supported resolution among the clades. *Potentilla dickinsii* and *P. fragariooides* are, however, associated with the Alba species in all trees resulting from the removal analyses [see Supporting Information—Figs S5–S8]. This connection is weakly supported, except when the Reptans clade is removed [see Supporting Information—Fig. S8]. In that tree, the Alba species are in a clade (pp 1.0) with both *P. dickinsii* and *P. fragariooides* nested inside.

The DHAR2 tree with all species included (Fig. 2A) recovers the Argentea, Ivesioid and Alba clades (pp 1.0, 1.0 and 0.98, respectively), as well as a clade comprising Argentea and the Ivesioids (pp 1.0). In this tree, the Reptans species are divided into two clades where one ('Reptans I'; pp 1.0) is sister (pp 1.0) to *P. dickinsii*, and the other ('Reptans II'; pp 1.0) is sister to *P. fragariooides* with low support (pp 0.85). The clade of Reptans I plus *P. dickinsii* is sister (pp 1.0) to a clade (pp 1.0) that consists of the Reptans II plus *P. fragariooides* clade, and the clade of Argentea and the Ivesioids. There is some evidence of duplicated patterns of relationships in the Reptans II clade (*P. reptans* and *P. erecta* are sisters in both subclades; pp 1.0), as well as in the Ivesioid clade where *Horkelia bolanderi*, *H. californica* and *Ivesia multifoliolata*

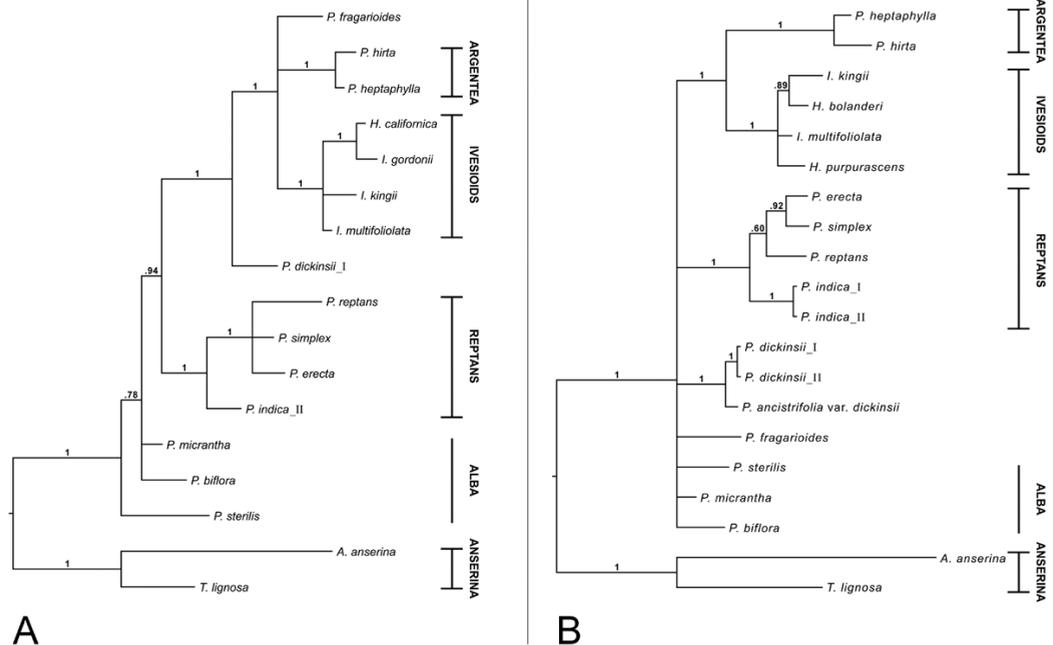


Figure 1. Fifty per cent majority rule consensus tree from the BI analyses of the chloroplast matK gene (A) and nuclear ribosomal ITS (B). Posterior probabilities are shown on the branch above the corresponding nodes. Specific individuals are indicated by Roman numerals. Clade affiliations of species are given to the right, where horizontal lines indicate that the clade is supported (cf. Table 1).

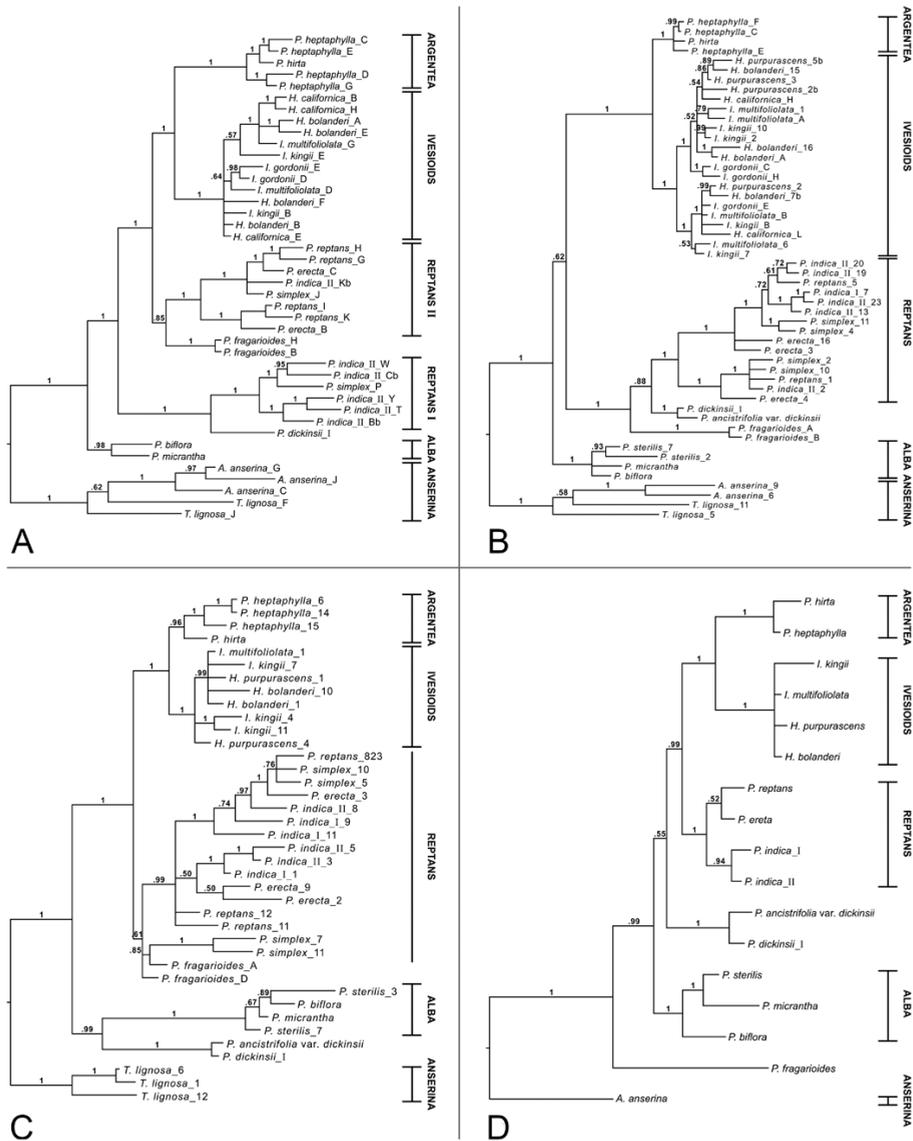


Figure 2. Fifty per cent majority rule consensus tree from the BI analyses of the nuclear low-copy genes DHAR2 (A), GAPCP1 (B), GBSSI-2 (C) and SbeI (D). Posterior probabilities are shown on the branch above the corresponding nodes. Specific individuals are indicated by Roman numerals. Clade affiliations of species are given within vertical lines to the right, where horizontal lines indicate that the clade is supported (cf. Table 1).

constitute one subclade (pp 1.0) while the other sequences of the same species are in unresolved positions outside of this subclade. None of the removal analyses [see Supporting Information—Figs S9–S12] change the topology of the trees, and there are only small changes in the posterior probabilities of the clades.

The GAPCP1 tree with all species included (Fig. 2B) recovers the Argentea, Ivesioid, Reptans and Alba clades (all pp 1.0), as well as the clade comprising Argentea and the Ivesioids (pp 1.0). A clade including all species but the Alba clade is very weakly supported (pp 0.62). Both *P. dickinsii* and *P. fragarioides* are in a clade (pp 1.0) with the Reptans clade, but the posterior probability for *P. dickinsii*

being the immediate sister to Reptans is low (pp 0.88). Within the Reptans clade there are two subclades (both pp 1.0), each including gene copies of the same species, and with *P. erecta* as sister to the rest (pp 1.0). The Ivesioid clade is also divided into two subclades (both pp 1.0) with gene copies of all included Ivesioid species in each subclade, but there is no further supported pattern. When removing *P. dickinsii* there are only small changes in the posterior probabilities in the tree [see Supporting Information—Fig. S13], but when removing *P. fragariooides* [see Supporting Information—Fig. S14] and both *P. dickinsii* and *P. fragariooides* [see Supporting Information—Fig. S15], there is support for the clade including all species but the Alba clade (pp 1.0 instead of pp 0.62 or lower). Removal of the Reptans clade does not change the topology of the tree, and shows *P. dickinsii* and *P. fragariooides* as sisters (pp 1.0) [see Supporting Information—Fig. S16].

The GBSSI-2 tree with all species included (Fig. 2C) recovers the Argentea, Ivesioid and Alba clades (pp 0.96, 1.0 and 1.0, respectively), as well as the clade comprising Argentea and the Ivesioids (pp 1.0). *Potentilla dickinsii* is sister (pp 0.99) to the Alba clade and this clade is sister (pp 1.0) to the rest of the ingroup (pp 1.0), which contains the Reptans species, *P. fragariooides* and the Argentea plus Ivesioid clade. There is some evidence of duplicated patterns of relationships in the Reptans clade, where sequences from the four included Reptans species form one subclade (pp 1.0), while the other sequences of the same species are in unresolved positions outside of this subclade. Removal of *P. dickinsii*, *P. fragariooides* or both of them does not change the topology of the trees [see Supporting Information—Figs S17–S19]. A notable change in the analysis excluding the Reptans clade [see Supporting Information—Fig. S20] is the drop in posterior probability for the Argentea clade (from pp 0.96 to pp 0.62).

The SbeI tree with all species included (Fig. 2D) recovers the Argentea, Ivesioid, Reptans and Alba clades (all pp 1.0). The Argentea and Ivesioid clades are sisters (pp 1.0), and the Reptans clade is in turn their sister (pp 0.99). *Potentilla dickinsii* is

the sister of these three clades with very low support (pp 0.55), while *P. fragariooides* is supported as sister (pp 1.0) to the rest of the ingroup (pp 0.99). The removal analyses [see Supporting Information—Figs S21–S24] result in no changes in the topology.

### MSC analyses

Models for the markers in the MSC analyses are found in Supporting Information—Table S3. For both data sets, a relaxed log-normal clock model and a birth-death process as tree prior were best fit to the data [see Supporting Information—Table S4]. The two MSC analyses recover the Argentea, Ivesioid, Reptans and Alba clades (all pp 1.0) (Fig. 3). In the analysis including *P. dickinsii* and *P. fragariooides* (Fig. 3A), the former is sister with low support (pp 0.90) to a very weakly supported clade (pp 0.44) constituting Argentea, the Ivesioids, *P. fragariooides* and Reptans, and the latter is sister with very low support (pp 0.49) to the clade (pp 0.98) of Argentea and the Ivesioids. The MSC analysis excluding *P. dickinsii* and *P. fragariooides* (Fig. 3B) shows a fully resolved tree of the major clades, where the Alba clade is sister (pp 1.0) to the rest of the ingroup (pp 0.94), in which the Reptans clade is sister to Argentea and the Ivesioids (pp 1.0).

### Discussion

This study resolves the backbone phylogeny of *Potentilla* using LCN markers. Our gene trees revealed patterns that could not have been discovered by chloroplast or nuclear ribosomal data, which makes it clear that LCN markers are crucial to the study of the evolutionary history of polyplipoids. Except for the *Fragariooides* clade, the clades found by Töpel et al. (2011) are supported in the majority of our gene trees.

### The *Fragariooides* species

In our gene trees, the *Fragariooides* species *P. dickinsii* and *P. fragariooides* did not constitute a clade on their own (Figs 1

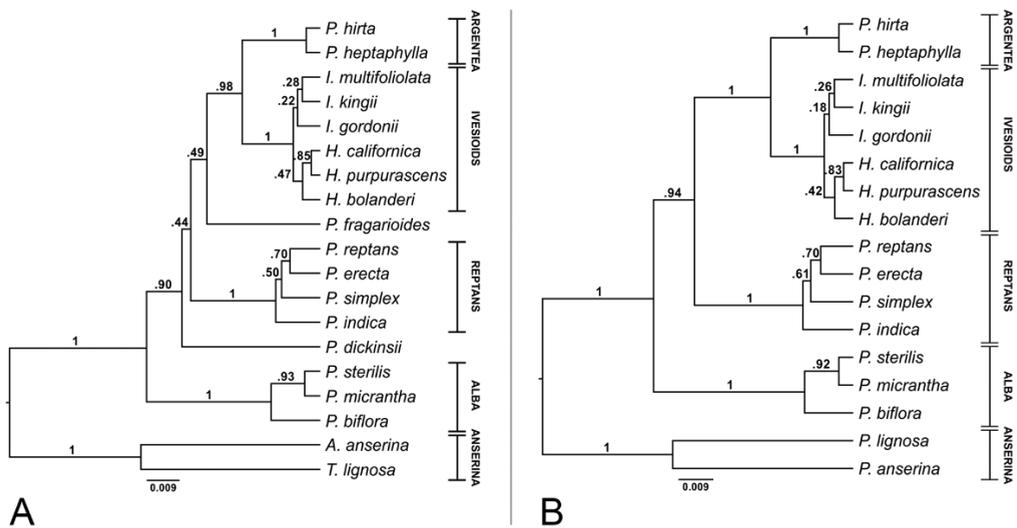


Figure 3. Bayesian consensus tree from the MSC analyses including *P. dickinsii* and *P. fragariooides* (A) and excluding *P. dickinsii* and *P. fragariooides* (B). Posterior probabilities are shown on the branch above the corresponding nodes. Clade affiliations of species are given within vertical lines to the right, where horizontal lines indicate that the clade is supported.

and 2), except in GAPCP1 only when the Reptans species were excluded [see [Supporting Information—Fig. S16](#)]. The *Fragarioides* species not being resolved as a monophyletic group is in agreement with most other previous analyses, where *P. fragarioides* is resolved as sister to *P. freyniana* or *P. stolonifera* to the exclusion of *P. dickinsii* (Dobeš and Paule 2010; Töpel et al. 2011, chloroplast tree; Feng et al. 2017). The only exception seems to be in the nuclear ribosomal tree by Töpel et al. (2011), where *P. dickinsii* is supported as sister to *P. fragarioides* and *P. stolonifera*. We therefore suggest that *P. dickinsii* should not be treated in the same infrageneric taxon as the other *Fragarioides* species.

Both *P. dickinsii* and *P. fragarioides* showed several different relationships in our gene trees; *P. dickinsii* was either sister to a clade consisting of *P. fragarioides*, *Argentea* and the *Ivesioids* (matK; Fig. 1A), in an unresolved ingroup consisting of the Reptans clade, *P. fragarioides*, the *Alba* species and a clade with *Argentea* plus the *Ivesioids* (ITS; Fig. 1B), sister to Reptans I (DHAR2; Fig. 2A), unresolved with *P. fragarioides* and Reptans (GAPCP1; Fig. 2B), sister to *Alba* (GBSSI-2; Fig. 2C) or unresolved with *Alba* and a clade consisting of Reptans and *Argentea* plus the *Ivesioids* (SbeI; Fig. 2D). The position of *P. fragarioides* was either in an unresolved clade with *Argentea* and the *Ivesioids* (matK; Fig. 1A), in an unresolved ingroup consisting of the Reptans clade, *P. dickinsii*, the *Alba* species and a clade consisting of *Argentea* plus the *Ivesioids* (ITS; Fig. 1B), unresolved with Reptans II and a clade consisting of *Argentea* plus the *Ivesioids* (DHAR2; Fig. 2A), unresolved with *P. dickinsii* and Reptans (GAPCP1; Fig. 2B), unresolved with the Reptans species and *Argentea* plus the *Ivesioids* (GBSSI-2; Fig. 2C) or sister to the rest of the ingroup (SbeI; Fig. 2D). Except in a few cases, the relationships seen in the low-copy markers were not seen in our or previous chloroplast and ribosomal DNA analyses; *P. fragarioides* was sister to the rest of *Potentilla* in the ribosomal tree of Eriksson et al. (1998), as in our SbeI tree. In the same tree, *P. dickinsii* was sister to *Alba*, which is a relationship seen in our GBSSI-2 tree and in our nuclear ribosomal tree when excluding the Reptans clade [see [Supporting Information—Figs S5 and S8](#)].

Exclusion of one or the other of *P. dickinsii* or *P. fragarioides* did not reduce incongruence among the gene trees [see [Supporting Information—Figs S1, S2, S5, S6, S9, S10, S13, S14, S17 and S18](#)]. However, when both *P. dickinsii* and *P. fragarioides* were excluded, the LCN markers showed the Reptans clade as sister to *Argentea* plus the *Ivesioids* (GAPCP1 and SbeI; see [Supporting Information—Figs S15 and S23](#)), or as a grade below the *Argentea* plus *Ivesioid* clade (DHAR2 and GBSSI-2; see [Supporting Information—Figs S11 and S19](#)). This topology was not contradicted by the chloroplast or ribosomal trees [see [Supporting Information—Figs S3 and S7](#)], although neither resolved these relationships with support. With this stable phylogenetic position of the Reptans clade in the backbone of the trees, we interpret *P. dickinsii* and *P. fragarioides* to be the main sources of conflicts seen in the gene phylogenies of *Potentilla*, and not the Reptans clade as initially thought.

### The Reptans clade

The Reptans clade has been monophyletic in previous phylogenetic analyses (Eriksson et al. 1998, 2003; Dobeš and Paule 2010; Töpel et al. 2011; Feng et al. 2017) and this was also true in most of our markers, the exceptions being DHAR2 and GBSSI-2 (Fig. 2A and C). In DHAR2, the clade was split into two clades, 'I' and 'II', where clade I was sister to *P. dickinsii* and clade II was sister with low support to *P. fragarioides*. In GBSSI-2, the clade was unresolved. The division of the Reptans clade into subclades in the DHAR2, GAPCP1 and GBSSI-2 trees (Fig. 2A–C),

and all but one species being polyploid (IPCN), suggests an early genome duplication event (autopolyploidization) in this clade. This is particularly evident in the GAPCP1 tree, where there are two supported subclades, and each species is represented in both. Of the Reptans species included in our study, *P. erecta* and *P. reptans* are tetraploids, *P. indica* is deca- and dodecaploid (10x, 12x), while the ploidy level of *P. simplex* is not known (IPCN; Kurtto et al. 2004). We found two and three different gene variants in *P. simplex*, that were placed in different subclades, which suggests that it may also be at least tetraploid. However, it is not possible to know based on our sample if the addition of unsampled species that belong to the Reptans clade would change these patterns, and therefore additional data are required to confirm an autopolyploid origin. *Potentilla flagellaris* included in the Reptans group by Wolf (1908; in *Grex Tormentillae*) is reported to be diploid (Sokolovskaya et al. 1985), but has never been part of a phylogenetic analysis. Inclusion of this species in future analyses might shed more light on the evolutionary history of the Reptans clade.

The Reptans species *P. indica* was recently classified in the genus *Duchesnea* (Chaoluan et al. 2003; Erterter and Reveal 2014b; Kechaykin and Shmakov 2016), but recognition of this genus renders *Potentilla* non-monophyletic. The idea that genera, as well as other taxa, named under the International Code of Botanical Nomenclature should be monophyletic is well-established in the taxonomic community (Angiosperm Phylogeny Group 1998; Backlund and Bremer 1998). All our analyses and those from previous studies (Eriksson et al. 1998, 2003; Töpel et al. 2011; Feng et al. 2017; Xiang et al. 2017) show that *P. indica* is a close relative to the type species *P. reptans*, and should therefore be included in *Potentilla*.

### The Ivesioid clade

As in the Reptans clade, the division of the Ivesioid clade into subclades in the DHAR2, GAPCP1 and GBSSI-2 trees (Fig. 2A–C), and the apparent lack of diploid species (Baldwin et al. 2012; IPCN), suggests an autopolyploidization event early in the clade's history. Only a few Ivesioid species have been subject to chromosome counting, and most of them are tetraploid (4x) (Baldwin et al. 2012; IPCN). The exception is *Horkelia marinensis* (not included in this study), which is octoploid (8x) (Baldwin et al. 2012). We found between two and four gene variants in the species included in our study, but this number was not consistent across the markers, which may be indicative of extensive allele variation in addition to polyploidization.

The latest edition of *Flora of North America* classified the Ivesioids in the genera *Horkelia*, *Horkeliella* and *Ivesia* (Ertter and Reveal 2014a). All our analyses, as well as those from previous studies (Eriksson et al. 1998, 2003; Dobeš and Paule 2010; Töpel et al. 2011; Feng et al. 2017; Xiang et al. 2017; Zhang et al. 2017; Persson et al. 2020), consistently show that they are nested within the *Potentilla* clade. Thus, as with *Duchesnea*, recognition of these genera causes *Potentilla* to be non-monophyletic. Keeping the genera of the Ivesioid clade separate from *Potentilla* would mean that hundreds of species outside of the Reptans clade, instead of about 10 Ivesioid species, would have to be formally transferred to new genera. In addition, the recent study by Persson et al. (2020) suggested a history of allopolyploid speciation between the *Argentea* and Ivesioid clades. Such a close evolutionary relationship adds weight to the argument of the inclusion of the Ivesioid species in *Potentilla*.

### Explanations for incongruent gene trees

Given our sample and that the major clades are supported in our species trees, hybridization does not seem to have played a prominent role before they formed, but rather during their

diversification. Töpel *et al.* (2011) suggested allopolyploidy as a plausible explanation for why the Reptans clade and the *Fragarioides* species showed different relationships in their chloroplast and ribosomal phylogenies. However, in our gene trees the Reptans species show relationships that rather indicate an autopolyploid origin of the clade (Fig. 2A–C), and *P. dickinsii* and *P. fragarioides* are diploids in all published chromosome counts (PCN). Homoploid hybridization between diploid ancestors could explain the chromosome numbers of *P. dickinsii* and *P. fragarioides*, but both species showed several different supported relationships in the gene trees, which means that more than two parental lineages may have been involved. In that case, the incongruences cannot be explained by a single hybridization event or hybridization alone.

In addition to hybridization, ILS is an evolutionary process that can lead to conflicting gene phylogenies (Doyle 1992; Maddison 1997). Gene trees usually coalesce deeper than the speciation events and are therefore expected to differ from the actual species phylogeny (Oxelman *et al.* 2017). Figure 4 shows how the LCN phylogenies in Fig. 2 may be contained within the species phylogeny in Fig. 3A. Assuming there were no polyploidizations or hybridizations between lineages before radiation of the clades, we interpret the gene variants conserved to have evolved before the time of diversification of the different clades. In DHAR2 (Fig. 4A), the Reptans species are divided into the Reptans I and II clades, where I is sister to *P. dickinsii* and II is sister (with low support) to *P. fragarioides*. Therefore, under this interpretation, a second gene variant evolved at least before the divergence of *P. dickinsii*, where one variant is conserved in the Reptans I and *P. dickinsii* lineage. The other variant evolved into two new variants before the divergence of Reptans II, and one of those variants is conserved in the Reptans II and *P. fragarioides* lineage. In GAPCP1 (Fig. 4B), *P. dickinsii* is sister to Reptans, and *P. fragarioides* is in turn their sister. Therefore, a second gene variant evolved at least before *P. dickinsii* diverged. One of those

variants evolved into two new variants, where one is conserved in *P. fragarioides* and the other one in *P. dickinsii* and Reptans. In GBSSI-2 (Fig. 4C), *P. dickinsii* is sister to Alba, and therefore a second gene variant evolved at least before divergence of Alba, where one variant is conserved in these two lineages. There was very low support for *P. fragarioides* being sister to Reptans in the GBSSI-2 tree, but there might have evolved two new variants from the one variant not conserved in Alba and *P. dickinsii* before the divergence of the Reptans lineage. One of those variants was then conserved in Reptans and *P. fragarioides*. In SbeI (Fig. 4D), *P. fragarioides* is sister to the rest of the ingroup (due to rooting on the Anserina clade). Therefore, a second gene variant evolved before the Anserina lineage diverged. One of those variants is conserved in Anserina and *P. fragarioides*, and the other one in Alba, Reptans, *P. dickinsii*, the Ivesioids and Argentea. No marker is immune to ILS, but a larger number of unlinked nuclear low-copy markers applied in a MSC model could potentially resolve the relationships of *P. dickinsii* and *P. fragarioides* to the major clades of *Potentilla*.

### Species trees and the backbone phylogeny

Since there were no indications of reticulate relationships between the clades in our gene trees, we performed MSC analyses to infer species trees. This was done to see if the shared patterns in the gene trees when *P. dickinsii* and *P. fragarioides* were excluded would be confirmed. This kind of analysis is advantageous over concatenation, since the model is able to take ILS and different histories of loci into account (Degnan and Rosenberg 2009). In addition, concatenation would not be possible for the cloned markers, since we do not know which gene variants belong to the same chromosome sets. The MSC analysis excluding *P. dickinsii* and *P. fragarioides* showed a fully resolved tree down to the level of the previously defined clades (Fig. 3B); where Alba was sister to the rest of the ingroup (pp 0.94), in which Reptans was sister to Argentea plus the Ivesioids

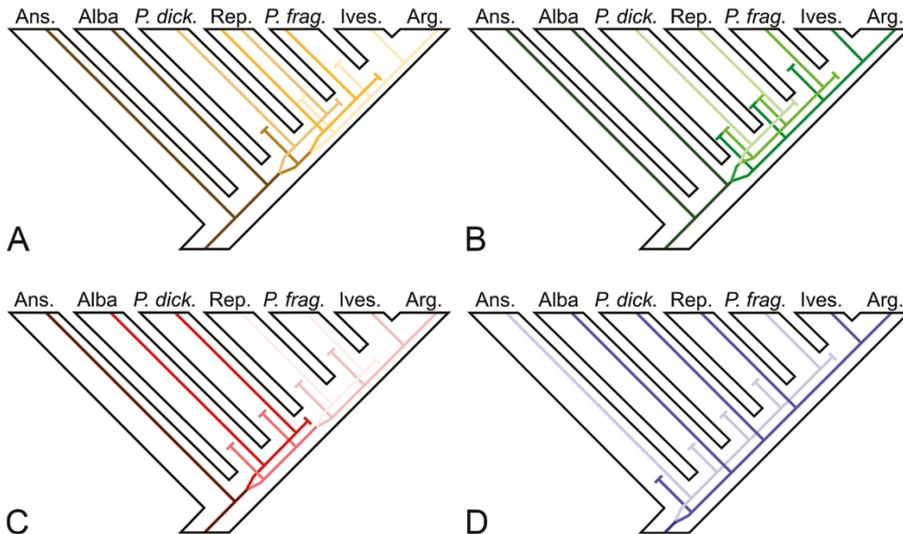


Figure 4. Plausible scenarios for how the gene trees of the nuclear low-copy markers in Fig. 2 may have evolved within the species phylogeny in Fig. 3 under ILS. Colours indicate different gene variants. (A) DHAR2, (B) GAPCP1, (C) GBSSI-2, (D) SbeI. Abbreviations: Ans. = Anserina clade; *P. dick.* = *P. dickinsii*; Rep. = Reptans clade; *P. frag.* = *P. fragarioides*; Ives. = Ivesioid clade; Arg. = Argentea clade.

(pp 1.0). As expected, the tree was not fully resolved when *P. dickinsii* and *P. fragarioides* were included (Fig. 3A) since the nodes directly related to the position of *P. dickinsii* and *P. fragarioides* were not supported. The low resolution within the Ivesioid and Reptans clades may be due to the presumably autopolyploid origins of these clades, as indicated by our interpretation of the gene tree topologies.

Recombination and hybridization are evolutionary processes that violate the MSC model (Bravo et al. 2019). Those processes result in reticulate relationships, and allopolyploid species are known to occur in *Potentilla* (Paule et al. 2011; Persson et al. 2020). Due to both auto- and allopolyploid taxa being present in the genus, it is evident that the complete evolutionary history of *Potentilla*, as opposed to the backbone relationships, may only be possible to describe correctly with a reticulate tree.

### Sampling effects

It is clear from our results that inferred relationships may be strongly affected by the inclusion or exclusion of single species. In our study, we focused on the relationships between the major clades, exploring under which sampling regimes we would get a supported phylogenetic backbone for *Potentilla*. This meant that we included representatives of the most well-supported clades, but also that some groups were excluded. In particular, we did not sample species of the Himalayan clade that were previously classified in *Sibbaldia* (Eriksson et al. 2015). In previous analyses using chloroplast and nuclear ribosomal data (Dobeš and Paule 2010; Eriksson et al. 2015; Feng et al. 2017), this clade is either resolved as sister to Alba or in an unresolved position in relation to Alba and the rest of *Potentilla*. Thus, inclusion of this clade would have been unlikely to affect the results presented here. There are possibly other species in addition to *P. dickinsii* and *P. fragarioides* that might affect the phylogeny in similar ways, but if so, they are still to be sampled for phylogenetic analysis. Inclusion of any close relatives to *P. dickinsii* and *P. fragarioides* in future studies could potentially stabilize their positions in the tree, and reveal more information about putative hybridizations in their evolutionary history.

### Conclusions

In this study, we have found a supported phylogenetic backbone of *Potentilla*, based on the relationships between the four major clades of *Potentilla*: the Alba clade as sister to the rest, then the Reptans clade, and then the Argentea clade as sister to the Ivesioid clade.

The different nuclear low-copy genes show incongruent phylogenetic relationships in our sample of *Potentilla* species, and we conclude that these incongruences are mainly caused by *P. dickinsii* and *P. fragarioides*.

*Potentilla dickinsii* and *P. fragarioides* have sometimes been joined in the informal *Fragarioides* group. We have no results that support this grouping as monophyletic, and suggest that these species should not be classified in the same infrageneric taxon.

We found no evidence in our sample for any hybridization or allopolyploidization events between the major clades, and suggest that early *Potentilla* evolution was affected by other processes such as ILS.

Possible autopolyploidization events were inferred in the Reptans and Ivesioid clades.

This study adds to the abundant molecular evidence that a monophyletic status of *Potentilla* would be achieved by an

inclusion of all the Ivesioid genera (*Horkelia*, *Horkeliella* and *Ivesia*), as well as *Duchesnea*.

### Supporting Information

The following additional information is available in the online version of this article—

**Table S1.** Primer pairs used for amplification of the markers analysed.

**Table S2.** Evolutionary models for the different markers in the Bayesian inference analyses.

**Table S3.** Evolutionary models for the different markers in the Multispecies Coalescent analyses.

**Table S4.** Log marginal likelihood values for analyses in \*BEAST.

**Figures S1–S24.** Consensus trees from the Bayesian inference analyses.

### Sources of Funding

Funding is gratefully acknowledged: University Museum of Bergen internal funding (to H.L.A.), and the Olaf Grolle Olsen's fund at the University of Bergen (to J.E.E.S., T.E. and N.L.P.).

### Contributions by the Authors

T.E. conceived the original idea and study set-up. I.T. and N.L.P. did the lab work and analyses. An early version of the manuscript was written by I.T., and later versions by N.L.P. All authors contributed by writing sections and/or commenting on drafts.

### Conflict of Interest

The authors declare that there are no conflicts of interest.

### Acknowledgements

We would like to thank the botanical gardens (Bergius Botanic Garden Stockholm, Bonn University Botanic Gardens, The Linnéan Gardens of Uppsala and Royal Botanic Garden Edinburgh) and herbaria (BG, E, GB, MARY, O, S and UPS) providing us with plant material. We also thank the staff at the Biodiversity Lab at the University of Bergen for help with the molecular methods and T. Marcussen for providing instructions on how to identify PCR recombinants. Finally, we would like to thank our colleagues in the Rosaceae Research Group and two anonymous reviewers for valuable comments on the manuscript.

### Literature Cited

- Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20:407–415.
- Angiosperm Phylogeny Group, Bremer K, Chase MW, Stevens PF, Anderberg A, Backlund A, Bremer B, Briggs BG, Endress PK, Fay MF, Goldblatt P, Gustafsson MHG, Hoot SB, Judd WS, Källersjö M, Kellogg KA, Kron KA, Les DH, Morton CM, Nickrent DL, Olmstead RG, Price RA, Quinn CJ, Rodman JE, Rudall PJ, Savolainen V, Soltis DE, Soltis PS, Sytsma KJ, Thulin M. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85:531–553.
- Backlund A, Bremer K. 1998. To be or not to be – principles of classification and monotypic plant families. *Taxon* 47:391–400.

- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* 29:2157–2167.
- Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution* 30:239–243.
- Baldwin BG, Goldman DH, Keil DJ, Patterson R, Rosatti TJ, Wilken DH, eds. 2012. *The Jepson manual: vascular plants of California*, 2nd edn. Berkeley, CA: University of California Press.
- Bento M, Gustafson JP, Viegas W, Silva M. 2011. Size matters in Triticeae polyploids: larger genomes have higher remodeling. *Genome* 54:175–183.
- Bravo GA, Antonelli A, Bacon CD, Bartoszek K, Blom MPK, Huynh S, Jones G, Knowles LL, Lamichhaney S, Marcussen T, Morlon H, Nakhleh LK, Oxelman B, Pfeil B, Schliep A, Wahlberg N, Werneck FP, Wiedenhoeft J, Willows-Munro S, Edwards SV. 2019. Embracing heterogeneity: coalescing the tree of life and the future of phylogenomics. *PeerJ* 7:e6399.
- Buti M, Moretto M, Barghini E, Mascagni F, Natali L, Brilli M, Lomsadze A, Sonogo P, Giongo L, Alonge M, Velasco R, Varotto C, Šurbanovski N, Borodovsky M, Ward JA, Engelen K, Cavallini A, Cestaro A, Sargent DJ. 2018. The genome sequence and transcriptome of *Potentilla micrantha* and their comparison to *Fragaria vesca* (the woodland strawberry). *Gigascience* 7:1–14.
- Chaoluan L, Ikeda H, Ohba H. 2003. *Duchesnea*. In: Wu Z, Raven PH, eds. *Flora of China*, Vol. 9. St. Louis, MO: Missouri Botanical Garden Press, 338–339.
- Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16:738–749.
- Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24:332–340.
- Dobes C, Paule J. 2010. A comprehensive chloroplast DNA-based phylogeny of the genus *Potentilla* (Rosaceae): implications for its geographic origin, phylogeography and generic circumscription. *Molecular Phylogenetics and Evolution* 56:156–175.
- Doyle JJ. 1992. Molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144–163.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4:e88.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29:1969–1973.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797.
- Eriksson T, Donoghue MJ, Hibbs MS. 1998. Phylogenetic analysis of *Potentilla* using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of Rosoideae (Rosaceae). *Plant Systematics and Evolution* 211:155–179.
- 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 Sciences* 164:197–211.
- Eriksson T, Lundberg M, Töpel M, Östensson P, Smedmark JEE. 2015. *Sibbaldia*: a molecular study of a remarkably polyphyletic genus in Rosaceae. *Plant Systematics and Evolution* 301:171–184.
- Erterter B. 1989. Revisionary studies in *Ivesia* (Rosaceae: Potentilleae). *Systematic Botany* 14:231–244.
- Erterter B, Reveal JL. 2014a. *Ivesia*, *Horkelia* & *Horkeliella*. In: *Flora of North America Editorial Committee*, eds. *Flora of North America North of Mexico*, Vol. 9. New York: Oxford University Press, 219–272.
- Erterter B, Reveal JL. 2014b. *Duchesnea*. In: *Flora of North America Editorial Committee*, eds. *Flora of North America North of Mexico*, Vol. 9. New York: Oxford University Press, 272–274.
- Evans RC, Campbell CS. 2002. The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *American Journal of Botany* 89:1478–1484.
- Feng T, Moore MJ, Yan MH, Sun YX, Zhang HJ, Meng AP, Li XD, Jian SG, Li JQ, Wang HC. 2017. Phylogenetic study of the tribe Potentilleae (Rosaceae), with further insight into the disintegration of *Sibbaldia*. *Journal of Systematics and Evolution* 55:177–191.
- GBIF Secretariat. 2019. GBIF Backbone Taxonomy. [www.gbif.org](http://www.gbif.org) (May 2019).
- Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. 1992. *Statistical Science* 7:457–472.
- Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27:570–580.
- Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23:254–267.
- IPCN, Index to Plant Chromosome Numbers. 1979. *IPCN database*. In: Goldblatt P, Johnson DE, eds. St. Louis, MO: Missouri Botanical Garden. [www.tropicos.org/project/ipcn](http://www.tropicos.org/project/ipcn) (May 2018).
- Kalkman C. 2004. *Potentilla*. In: Kubitzki K, ed. *Flowering plants - Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales*. Berlin: Springer, 366.
- Kass RE, Raftery AE. 1995. Bayes factors. *Journal of the American Statistical Association* 90:773–795.
- Kechaykin AA, Shmakov AI. 2016. A system of subtribe Potentillinae J. Presl (Rosaceae Juss.). *Turczaninowia* 19:114–128.
- Kendall DG. 1948. On the generalized “birth-and-death” process. *Annals of Mathematical Statistics* 19:1–15.
- Kurto A, Lampinen R, Junikka L, eds. 2004. *Atlas Florae Europaeae. Distribution of vascular plants in Europe. Rosaceae (Spiraea to Fragaria, excl. Rubus)*, Vol. 13. Helsinki: Societas Biologica Fennica Vanamo.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34:772–773.
- Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30:3276–3278.
- Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploids. *Biological Journal of the Linnean Society* 82:651–663.
- Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913–925.
- Lundberg M, Töpel M, Eriksen B, Nylander JA, Eriksson T. 2009. Allopolyploidy in Fragariinae (Rosaceae): comparing four DNA sequence regions, with comments on classification. *Molecular Phylogenetics and Evolution* 51:269–280.
- Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* 46:523–536.
- Mandáková T, Pouch M, Harmanová K, Zhan SH, Mayrose I, Lysak MA. 2017. Multisped genome diploidization and diversification after an ancient allopolyploidization. *Molecular Ecology* 26:6445–6462.
- Marcussen T, Heier L, Brysting AK, Oxelman B, Jakobsen KS. 2015. From gene trees to a dated allopolyploid network: insights from the angiosperm genus *Viola* (Violaceae). *Systematic Biology* 64:84–101.
- Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
- 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* 56:837–856.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 14 November, New Orleans, LA, USA, 1–8.
- Oxelman B, Brysting AK, Jones GR, Marcussen T, Oberprieler C, Pfeil BE. 2017. Phylogenetics of allopolyploids. *Annual Review of Ecology, Evolution, and Systematics* 48:543–557.
- Paule J, Scherbantín A, Dobeš C. 2012. Implications of hybridisation and cytotypic differentiation in speciation assessed by AFLP and plastid haplotypes—a case study of *Potentilla alpicola* La Soie. *BMC Evolutionary Biology* 12:132.
- Paule J, Sharbel TF, Dobeš C. 2011. Apomictic and sexual lineages of the *Potentilla argentea* L. group (Rosaceae): cytotype and molecular genetic differentiation. *Taxon* 60:721–732.

- Persson NL, Eriksson T, Smedmark JEE. 2020. Complex patterns of reticulate evolution in opportunistic weeds (*Potentilla* L., Rosaceae), as revealed by low-copy nuclear markers. *BMC Evolutionary Biology* 20:38.
- Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenaault M, Dickinson TA, Campbell CS. 2007. Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution* 266:5–43.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Tracer v1.6. [www.beast.community/tracer](http://www.beast.community/tracer) (March 2018).
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61:539–542.
- Shi S, Li J, Sun J, Yu J, Zhou S. 2013. Phylogeny and classification of *Prunus sensu lato* (Rosaceae). *Journal of Integrative Plant Biology* 55:1069–1079.
- 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, Troglio 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* 43:109–116.
- Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49:369–381.
- Small RL, Cronn RC, Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17:145–170.
- Smedmark JEE, Eriksson T, Bremer B. 2005. Allopolyploid evolution in Geinae (Columelidae: Rosaceae) – building reticulate species trees from bifurcating gene trees. *Organisms Diversity & Evolution* 5:275–283.
- Sokolovskaya AP, Probatova NS, Rudyka, EG. 1985. Chromosome numbers in the species of the families Asteraceae, Poaceae, Rosaceae from the Primorye Region, Kamchatka and Sakhalin. *Botanicheskii Zhurnal SSSR* 70:126–128.
- Staden R. 1996. The Staden sequence analysis package. *Molecular Biotechnology* 5:233–241.
- Takeda H. 1911. On some *Potentillas* from the Far East. *Bulletin of Miscellaneous Information (Royal Botanic Gardens, Kew)* 6:250–256.
- Töpel M, Lundberg M, Eriksson T, Eriksen B. 2011. Molecular data and ploidal levels indicate several putative allopolyploidization events in the genus *Potentilla* (Rosaceae). *PLoS Currents* 3:RRN1237.
- Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42:225–249.
- Wolf T. 1908. Monographie der Gattung *Potentilla*. Stuttgart: Bibliotheca Botanica 16:1–714.
- 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* 34:262–281.
- Yang Z, Rannala B. 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution* 14:717–724.
- Yule G. 1924. A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis F.R.S. *Philosophical Transactions of the Royal Society London* 213:21–87.
- 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 Cretaceous based on plastid phylogenomics. *The New Phytologist* 214:1355–1367.



---

# Paper II



RESEARCH ARTICLE

Open Access

# Complex patterns of reticulate evolution in opportunistic weeds (*Potentilla* L., Rosaceae), as revealed by low-copy nuclear markers



Nannie L. Persson, Torsten Eriksson and Jenny E. E. Smedmark\*

## Abstract

**Background:** Most cinquefoils (*Potentilla* L., Rosaceae) are polyploids, ranging from tetraploid (4x) to dodecaploid (12x), diploids being a rare exception. Previous studies based on ribosomal and chloroplast data indicated that Norwegian cinquefoil (*P. norvegica* L.) has genetic material from two separate clades within *Potentilla*; the Argentea and the Ivesioid clades – and thus a possible history of hybridization and polyploidization (allopolyploidy). In order to trace the putative allopolyploid origin of the species, sequence data from low-copy, biparentally inherited, nuclear markers were used. Specimens covering the circumpolar distribution of *P. norvegica* and its two subspecies were included, along with the morphologically similar *P. intermedia*. *Potentilla* species of low ploidy level known to belong to other relevant clades were also included.

**Results:** Gene trees based on three low-copy nuclear markers, obtained by Bayesian Inference and Maximum Likelihood analyses, showed slightly different topologies. This is likely due to genomic reorganizations following genome duplication, but the gene trees were not in conflict with a species tree of presumably diploid taxa obtained by Multispecies Coalescent analysis. The results show that both *P. norvegica* and *P. intermedia* are allopolyploids with a shared evolutionary history involving at least four parental lineages, three from the Argentea clade and one from the Ivesioid clade.

**Conclusions:** This is the first time that reticulate evolution has been proven in the genus *Potentilla*, and shows the importance of continuing working with low-copy markers in order to properly resolve its evolutionary history. Several hybridization events between the Argentea and Ivesioid clades may have given rise to the species of Wolf's grex Rivaies. To better estimate when and where these hybridizations occurred, other Argentea, Ivesioid and Rivaies species should be included in future studies.

**Keywords:** Low-copy nuclear markers, Hybridization, Ivesioids, Molecular cloning, Molecular phylogeny, Polyploidy, *Potentilla*, Reticulate evolution

\* Correspondence: [jenny.smedmark@uib.no](mailto:jenny.smedmark@uib.no)

Department of Natural History, University Museum, University of Bergen,  
Postboks 7800, N-5020 Bergen, Norway



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The evolution of species is usually considered to be a slow process, working over thousands or even millions of years. Sometimes, however, new species evolve within a relatively short period of time through polyploidization. This phenomenon is common throughout the vascular plants, where genome duplications can be found from the ferns [1] and lycopods [2], to the asterids [3]. Two main types of polyploidization are recognized; autopolyploidization, where the duplication occurs within a single species, and allopolyploidization, where the duplication occurs in combination with hybridization between two different species [4]. A doubling of the chromosomes can make a sterile hybrid fertile [5, 6] and cause a reproductive barrier between individuals of the new genomic state and the old state [6, 7]. This may create a new, independently evolving, lineage that could thus be regarded as a new species [8].

The rose family (Rosaceae Juss.) is well known for its many polyploid taxa, and there seem to have been a large number of independent auto- and allopolyploidization events during its evolutionary history [9–11]. Chromosome counting data, summarized by Vamossi and Dickinson [12], suggest that around half of the family's genera include at least one polyploid species. Some, as for instance *Acacina* L., *Alchemilla* L. and *Sorbaria* (Ser.) A. Braun, consist only of polyploids.

The cinquefoils, *Potentilla* L., is an example of a genus in Rosaceae with mixed ploidy levels. According to the Chromosome Counts Database [13] only a few species seem to be exclusively diploid, e.g. *P. biflora* Willd. ex Schltld., *P. freyniana* Bornm. and *P. valderia* L. At the other end, *P. gracilis* Douglas ex Hook., *P. tabernaemontani* Asch. and *P. indica* (Jacks.) Th. Wolf have been reported to have dodecaploid (12x) populations. Furthermore, it is not uncommon for single species to have multiple ploidy levels. The genus has undergone a major recircumscription since the first molecular studies of the group were performed [14, 15]; both plastid and nuclear ribosomal markers showed that it had been polyphyletic. They strongly indicated that some previous *Potentilla* species are more closely related to the strawberries, *Fragaria* L., in the Fragariinae clade, such as those species now assigned to the genera *Dasiphora* Raf. and *Drymocallis* Fourr. In contrast, the genus *Duchesnea* Sm. and some species of *Sibbaldia* L., were instead shown to belong to *Potentilla* [14, 16]. However, the debate on where to draw the generic delimitation is still ongoing; as whether to include the genus *Argentina* Hill. and its sisters [15, 17] or not [18–20]. Regardless whether *Argentina* is included or not, the genus is still polyphyletic in certain classifications where *Duchesnea* (*P. indica*) and the genera of the North American Ivesioid clade (*Horkelia* Cham. & Schltld., *Horkeliella* (Rydb.) Rydb. and *Ivesia* Torr. & A.Gray) are separated from *Potentilla* [17, 21, 22]. Within *Potentilla* in the strict sense, there are a

number of well supported subclades, such as the Alba, Reptans and Ivesioid clades [23]. The most species-rich subclade, called either “Argentea” [23] or “core group” [18] in previous studies, is, however, in itself poorly resolved [18, 20, 23].

Previous studies have found a possible connection between the Argentea and Ivesioid clades in the polyploid species *P. norvegica* L. This species has been shown to have different phylogenetic relationships depending on whether the analyses were based on chloroplast [15, 18, 23] or nuclear ribosomal data [14, 15, 23]; with chloroplast data the species groups with the Argentea clade, but with ribosomal data it groups with the Ivesioids. Töpel et al. [23] speculated that this may be due to an evolutionary history of polyploidization in combination with hybridization between these two clades. It is, however, not previously known to what extent these two processes have played a part in the formation of *P. norvegica*, or if the discordance between chloroplast and ribosomal data is the result of other processes, such as a single hybridization event followed by introgression [24].

In his monograph of *Potentilla*, Wolf [25] placed *P. norvegica* together with 20 other species in his “grex” Rivaes. Of these, *P. intermedia* L. and *P. supina* L. have a similar circumpolar distribution as *P. norvegica*, while the North American species *P. biennis* Greene and *P. rivalis* Nutt. are morphologically similar to *P. norvegica*. Another common feature is that they are annuals or short-lived perennials [17, 25]. *Potentilla norvegica* was originally described by Linnaeus [26] as two separate species based on stem and leaflet morphology of European specimens; *P. norvegica* L. and *P. monspeliensis* L. In 1803, Michaux [27] described *P. hirsuta* Michx. based on North American specimens, but Ledebour [28] later synonymized *P. monspeliensis* and *P. hirsuta* under *P. norvegica*. Nevertheless, there is striking morphological variation within the species, and today two subspecies are generally accepted. However, it has been unclear which subspecies name has priority. In 1904, Ascherson and Graebner [29] described “*P. norvegica* II. *monspeliensis*”, by some nomenclatural databases interpreted as a subspecies [30, 31]. However, Hylander [32] must have interpreted this as a variety. Since names only have priority at the same nomenclatural rank [33], he was able to list “II. *monspeliensis*” under *P. norvegica* ssp. *hirsuta* (Michx.) Hyl. The name that will be used in this study is therefore *Potentilla norvegica* ssp. *hirsuta*, which refers to specimens displaying the morphology first used to describe *P. monspeliensis*. Since *P. norvegica* ssp. *hirsuta* is the most common subspecies in North America, it is sometimes referred to as the American form, and the autonym ssp. *norvegica* as the European form, but there are numerous findings of ssp. *hirsuta* in Europe. Most floras argue for an East European origin of the species,

and that ssp. *hirsuta* later has dispersed to Europe from North America [34–38]. However, no molecular phylogenetic work has been performed in order to test these hypotheses.

The two types of molecular data most commonly used in phylogenetic studies of plants both have the inconvenience that they are not able to detect reticulate patterns in phylogenetic trees. The chloroplast is inherited uniparentally and nuclear ribosomal markers are most often subject to concerted evolution, while low-copy nuclear markers are inherited biparentally and present in each subgenome after a polyploidization event [39]. This means that they have the potential to retrieve polyploid signals in a single gene tree. For instance, Smedmark et al. [40] resolved the Colurieae clade in Rosaceae with its many polyploid species using this type of marker. However, different gene trees do not necessarily depict the same evolutionary history, due to processes such as horizontal gene transfer, deep coalescence and lineage sorting [41]. Furthermore, since it is not possible to know beforehand which sequences are homologous, low-copy markers cannot be concatenated to form larger datasets. Therefore, when polyploidy is present, it is important to investigate several low-copy markers in order to find the species tree. In a phylogenetic gene tree covering a simple polyploidization event, the gene copies of an autopolyploid (paralogues) would be each other's sisters, while the gene copies of an allopolyploid (homoeologues) would be sisters to their respective parental lineage. This has a number of effects on species trees, since the evolutionary history of an allopolyploid would be better represented by a reticulate pattern where lineages merge, rather than by a traditional bifurcating tree [24].

By using low-copy nuclear markers, this study aims to determine (1) if *Potentilla norvegica* and *P. intermedia* have an allopolyploid evolutionary history resulting from hybridization between the Argentea and Ivesioid clades; (2) if this is the case, do they share polyploidy events; and (3) if morphology and geography are concordant with intraspecific phylogeny in *P. norvegica*.

## Results

### Sequence alignment

All markers shared some identical *Potentilla norvegica* sequences across individuals, which are marked in brackets in the gene trees (Figs. 1, 2 and 3).

In addition, two GAPCP1 sequences from *P. intermedia* were identical to two *P. norvegica* sequences (P to 97E and D to 113D), while the GBSSI-1 *P. intermedia* sequence Kb and *P. norvegica* sequence 96N differed in only one base pair.

## Phylogenetic analyses

### Partitioning and model suggestions

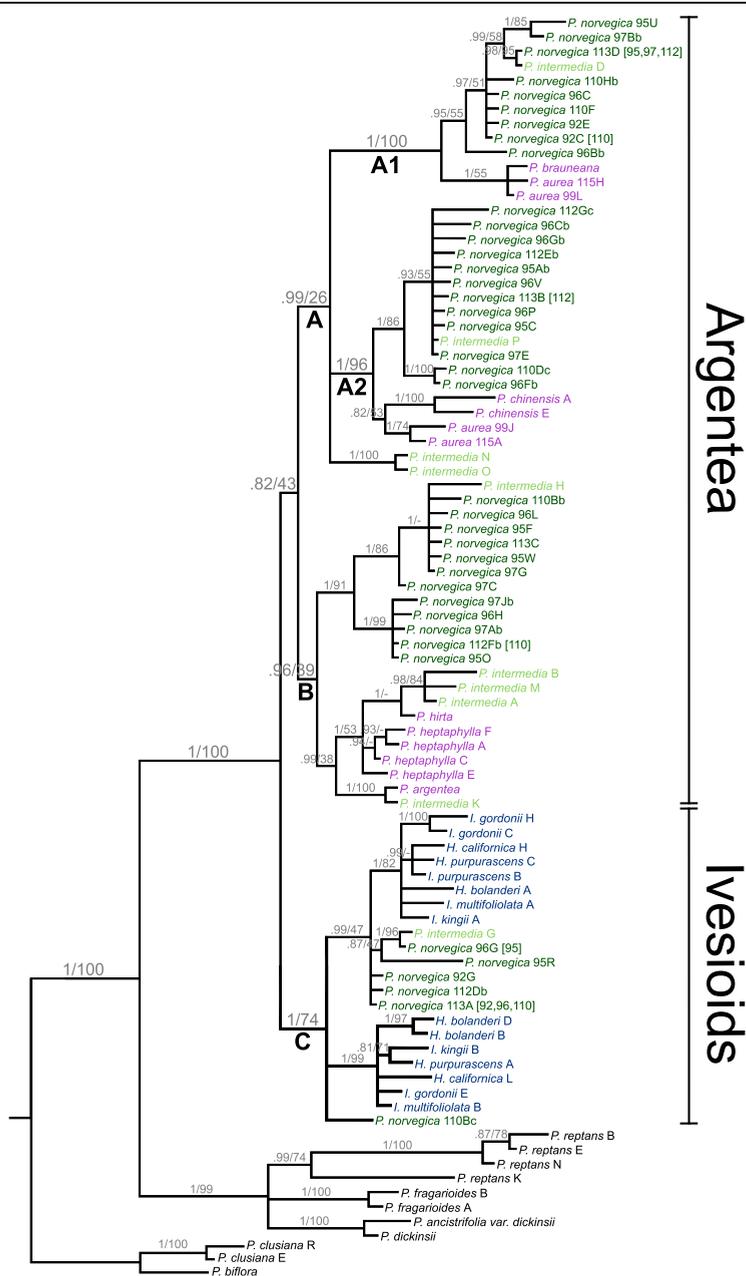
The lowest log likelihood value for the partitioning and model analyses were obtained under the AICc criterium for all markers. Partitioning schemes and their assigned models are found in Table 1.

### Bayesian and ML analyses

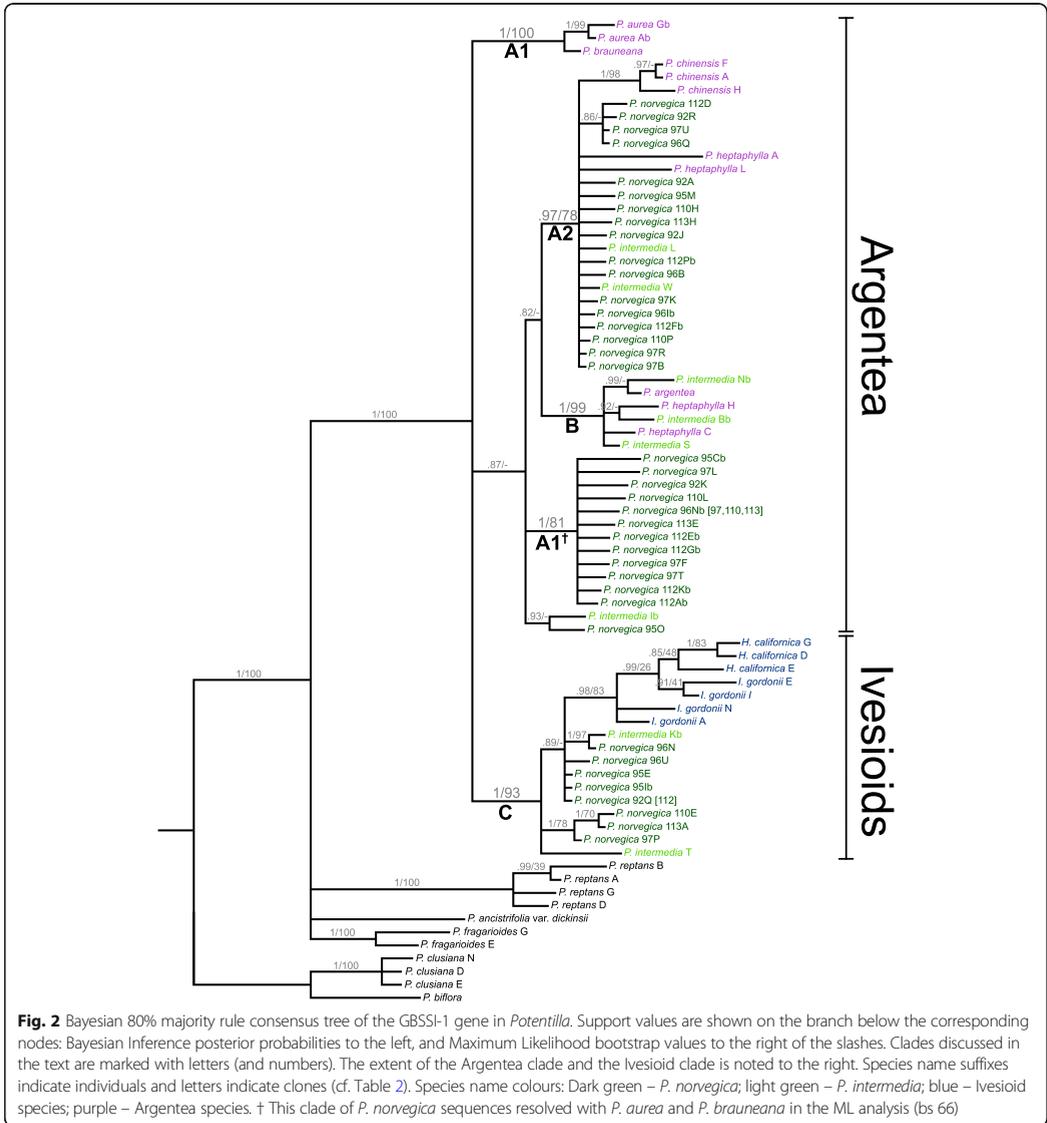
The Bayesian phylogenetic analysis of GAPCP1 resolved *Potentilla norvegica* sequences in four clades (Fig. 1). Three of these clades were sisters to Argentea species (clade A1, posterior probability 1.0; A2, pp. 1.0; B, pp. 0.96) and one was sister to the Ivesioids (C, pp. 1.0). *Potentilla intermedia* was found in the same four clades. The A1 and A2 clades formed a polytomy together with two *P. intermedia* sequences (A, pp. 0.99). The node connecting the A and B clades, i.e. corresponding to the Argentea clade, was not strongly supported (pp 0.82). The Ivesioid genera (*Horkelia*, *Horkeliella* and *Ivesia*) in clade C were divided into two subclades (both pp. 1.0), with at least one sequence from each species in each subclade. The Maximum Likelihood analysis showed the same topology, but only clades A1 and A2 were supported (bootstrap support 100 and 96, respectively).

The Bayesian analysis of GBSSI-1 showed *P. norvegica* sequences in three clades (Fig. 2), of which two correspond to A2 (pp 0.97) and C (pp 1.0) in the GAPCP1 tree. There was, however, no *P. norvegica* homoeologue associated with the Argentea species in clade B (pp 1.0). *Potentilla intermedia* homoeologues were found in clades A2, B and C. Clades A2 and B were sisters with low support (pp 0.82). They formed a polytomy (pp 0.87) with the third *P. norvegica* clade (pp 1.0) and a small clade consisting of one *P. norvegica* and one *P. intermedia* sequence (pp. 0.93). This polytomy was in turn in a polytomy (pp 1.0) with clade C and the Argentea species from clade A1 (pp 1.0). Thus, there was no Argentea clade in this tree. Within clade C, the Ivesioid species formed one subclade (pp 0.98), in which two of the four *Ivesia* sequences were sisters to *Horkelia* (pp 0.99), while the other two were unresolved. The ML analysis showed clades A1 (bs 66), A2 (bs 78), B (bs 99) and C (bs 93), but their relative positions were not supported. The clade with only *P. norvegica* sequences, present in the Bayesian tree, was placed as sister to *P. aurea* and *P. brauneana* (A1) in the ML tree. Even though bootstrap support was low, we will refer to this *P. norvegica* clade as A1<sup>†</sup>.

The Bayesian analysis of DHAR2 (Fig. 3) also showed *P. norvegica* in three clades, two of them corresponding to A1 (pp 1.0) and C (pp 0.93) in the other trees, while the third had not been seen previously. This clade consisted of *P. norvegica*, *P. intermedia* and one *P. heptaphylla* sequence, and was supported as sister to clade C (pp 1.0), while the clade itself had low support (pp 0.86). There was no supported Argentea clade in this tree. The



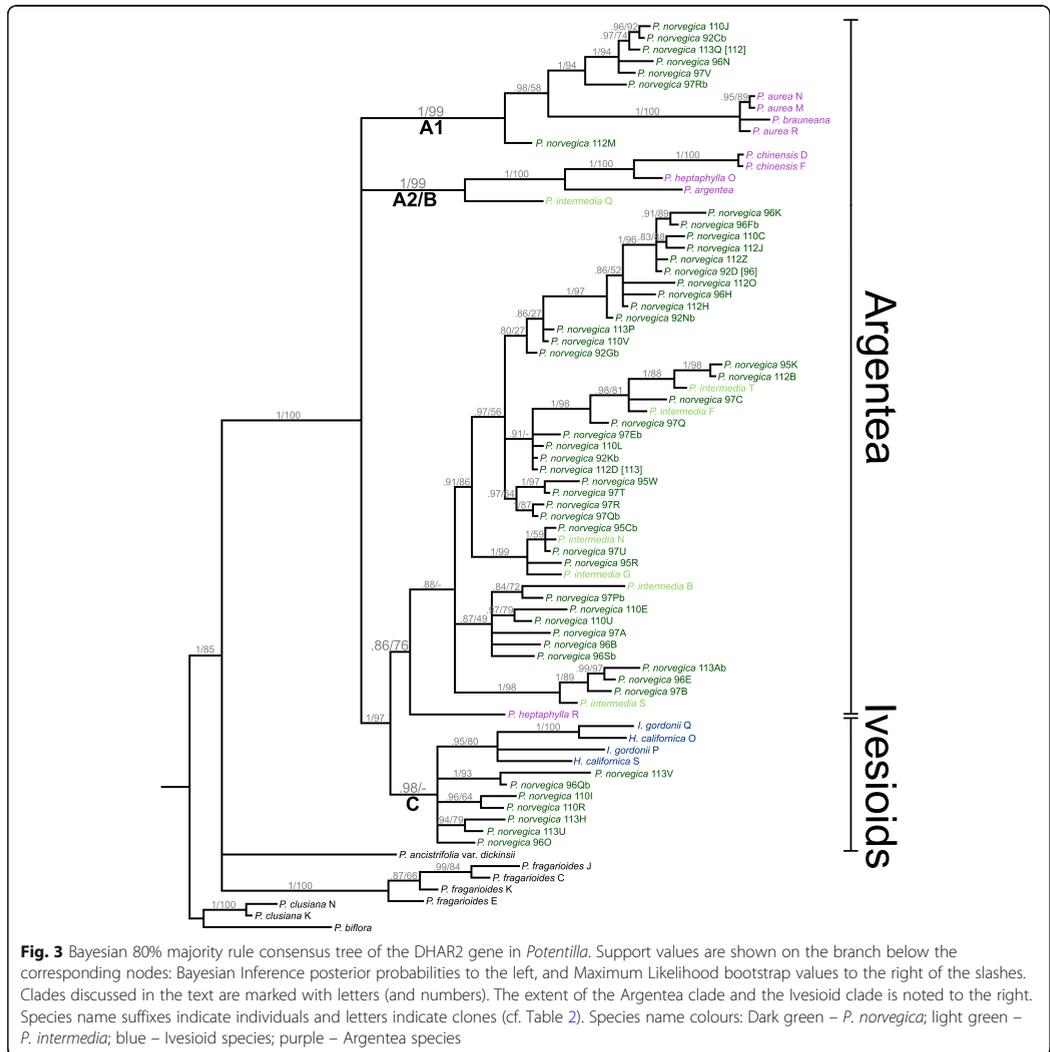
**Fig. 1** Bayesian 80% majority rule consensus tree of the GAPCP1 gene in *Potentilla*. Support values are shown on the branch below the corresponding nodes: Bayesian Inference posterior probabilities to the left, and Maximum Likelihood bootstrap values to the right of the slashes. Clades discussed in the text are marked with letters (and numbers). The extent of the Argentea clade and the Ivesioid clade is noted to the right. Species name suffixes indicate individuals and letters indicate clones (cf. Table 2). Species name colours: Dark green – *P. norvegica*; light green – *P. intermedia*; blue – Ivesioid species; purple – Argentea species



Ivesioids formed one subclade in clade C, where one of two *Horkelia* sequences and one of two *Ivesia* sequences were sisters (pp 1.0), while the other two were unresolved. The ML analysis showed no conflicting topology of the major clades, but there were two Ivesioid subclades (bs 83 and 100), with one *Ivesia* and one *Horkelia* sequence in each, and those were supported as sisters (bs 80). The sister clade to clade C was also supported (bs 76).

No clade was specific to, or missing, any of the two *P. norvegica* subspecies or seven individuals throughout all three gene trees. For instance, clade C was missing individual 97 in the GAPCP1 tree and individuals 92, 95, 97 and 112 in the DHAR2 tree, while all individuals were represented in this clade in the GBSSI-1 tree.

Five species with previously published diploid chromosome counts [13], *P. aurea*, *P. chinensis*, *P. clusiana*, *P. fragarioides* and *P. heptaphylla*, failed direct sequencing



and were therefore molecularly cloned. In the GBSI-1 and DHAR2 trees, *P. aurea* was sister to *P. brauneana* in clade A1 (pp 1.0). However, in the GAPCP1 tree two *P. aurea* sequences were placed in clade A1, but the other two were placed in clade A2 as sisters to *P. chinensis* (pp 0.82). In the GAPCP1 tree, all *P. heptaphylla* sequences were placed in clade B, but in the GBSI-1 tree two sequences were found in A1 and two found in A2. In the DHAR2 tree they were even further apart, with one sequence as sister to *P. chinensis* in A2/B and one as sister to *P. norvegica* and *P. intermedia* in the sister

clade to clade C. The sequences of *P. chinensis*, *P. clusiana* and *P. frugarioides* formed clades of their own.

**Control analyses**

The control ML analyses for putatively missed *P. norvegica* gene copies did not reveal any new clades or overlooked patterns in terms of subspecies or geographical origin. However, two excluded *P. intermedia* GBSI-1 sequences were indicated to belong in clade A1. One of these was added to the dataset, but the Bayesian analysis resulted in the collapse of clades B and C, which received

**Table 1** Partitioning and evolutionary models used for analysis in MrBayes, as suggested by PartitionFinder2

GAPCP1				
Subset	1st codon	2nd codon	3rd codon	introns
Model	F81 + G	JC	GTR + G	HKY + G
GBSSI-1				
Subset	1st codon	2nd codon	3rd codon + introns	
Model	GTR + I + G	JC + I	GTR + I + G	
DHAR2				
Subset	intron			
Model	HKY + I + G			

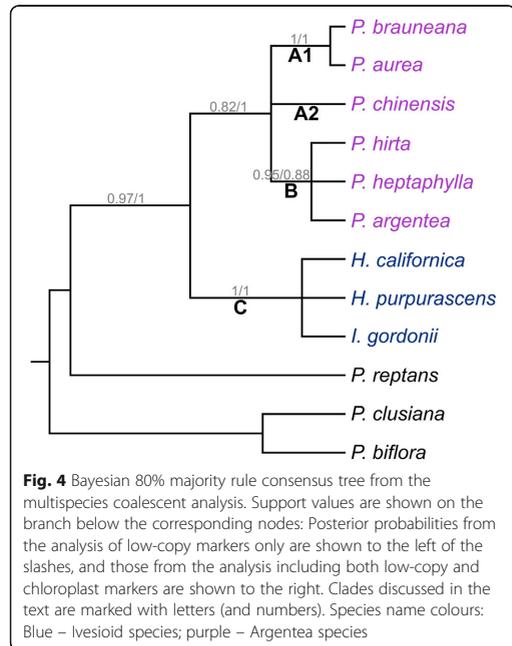
high support in the other trees. Similarly, one *P. intermedia* DHAR2 sequence was indicated to belong in clade C, but when added to the dataset it also resulted in the collapse of several clades. Both sequences were therefore excluded again from their respective datasets.

#### Multispecies coalescent analysis

The substitution model suggested for all markers was HKY [42], with gamma as site heterogeneity model for GAPCP1 and GBSSI-1, and invariant sites for DHAR2. The clock model and tree prior that was best fitted to the low-copy marker only dataset was a relaxed uncorrelated lognormal clock with a birth-death process, and for the combined low-copy and chloroplast marker dataset a relaxed uncorrelated lognormal clock with a birth process. The two trees had the same topology, but some of the support values differed (Fig. 4). In both trees, the Ivesioid clade was supported (pp 1.0) and *P. aurea* and *P. brauneana* were sisters (pp 1.0), corresponding to clade A1 in the gene trees. *Potentilla hirta*, *P. heptaphylla* and *P. argentea* formed a polytomy (pp 0.95 in the low-copy marker dataset and pp. 0.88 in the combined dataset) corresponding to clade B, while *P. chinensis* of clade A2 was unresolved. The Argentea clade received low support (pp 0.82) in the low-copy marker tree and full support (pp 1.0) in the combined tree.

#### Morphological study

Most specimens studied from the collections of BG, GB, O, S and UPS were of intermediate morphology. They had, for instance, whole stipules (ssp. *norvegica*), but obovate leaflets and obtuse leaflet teeth (ssp. *hirsuta*). For European specimens, there was approximately equal occurrence of typical individuals of the two subspecies. For the North American and East Russian specimens, typical individuals showing the ssp. *hirsuta* morphology were more common than those showing the ssp. *norvegica* morphology. The few North American specimens showing the ssp. *norvegica* morphology were all but one (Alaska,



**Fig. 4** Bayesian 80% majority rule consensus tree from the multispecies coalescent analysis. Support values are shown on the branch below the corresponding nodes: Posterior probabilities from the analysis of low-copy markers only are shown to the left of the slashes, and those from the analysis including both low-copy and chloroplast markers are shown to the right. Clades discussed in the text are marked with letters (and numbers). Species name colours: Blue – Ivesioid species; purple – Argentea species

USA) collected in the East (Ontario, Canada, to New York, USA), a pattern also seen by Rydberg [43].

#### Discussion

Despite the slightly different topologies of the three single-copy nuclear markers presented in this study, it is clear that both *Potentilla norvegica* and *P. intermedia* are allopolyploids with a shared evolutionary history involving one parental lineage in the Ivesioid clade and multiple parental lineages in the Argentea clade. These results rule out a simple case of introgression, and reveal a complex reticulate evolutionary history of several hybridization events in combination with polyploidization. For *P. norvegica*, there was no concordance between geography and intraspecies phylogeny. Thus, on the basis of our data we see no support for species differentiation, as first suggested by Linnaeus [26], since the majority of the individuals studied in the herbaria were of intermediate morphological form. Neither did our molecular data support a division into subspecies, but a more extensive study involving more individuals of especially ssp. *norvegica* would be better able to investigate the relationship between them.

As previously shown in studies based on chloroplast and ribosomal data [14, 15, 18, 20, 23, 44], the Ivesioid clade is deeply nested in *Potentilla* (Figs. 1, 2 and 3). Thus, following the established practice of only recognizing monophyletic taxa, the Ivesioid genera *Horkelia*, *Horkeliella* and *Ivesia* should be incorporated in

*Potentilla*. The type species of *Potentilla*, *P. reptans*, is part of the small Reptans clade, which is the sister clade to the Argentea and Ivesioid clades. If the Ivesioid genera were to be retained, the many species of the large Argentea clade would have to be reclassified, and it is probable that almost all would have to change names. However, the new evidence presented here of a hybridization event between the Argentea and Ivesioid clades indicates a close relationship between the groups, and adds a compelling argument for including the Ivesioid genera in *Potentilla*.

The three gene trees conform well to the backbone reference (Fig. 4), apart from some *P. aurea* and *P. heptaphylla* sequences. It is, however, clear that one *P. norvegica* GBSSI-1 homoeologue (subgenome-specific gene copy) is missing in clade B and one *P. intermedia* GBSSI-1 homoeologue is missing in clade A1<sup>†</sup> (Fig. 2). In the DHAR2 tree (Fig. 3), there is a major rearrangement in which the Ivesioid clade C is sister to what could be assumed to be parts of clade A2 or B. In addition, contrary to previous analyses based on chloroplast and nuclear ribosomal data [16, 18, 20, 23], the support for the Argentea clade was low both in the individual gene trees (Figs. 1, 2 and 3) and in the species tree based on low-copy markers only (Fig. 4). Thus, it is evident that phylogenetic relationships of low-copy nuclear genes are complicated by a number of evolutionary processes. A polyploid genome with high genetic redundancy may be subjected to large genomic alterations, such as deletions, insertions, or recombinations, to a high extent without causing fatal effects [45]. For instance, entire homoeologues may be lost as a response to genomic reorganization after polyploidization [46, 47] or via incomplete lineage sorting during speciation after hybridization [41]. Furthermore, if an interallelic recombination [24] splits a gene in two unequal parts during meiosis, the new recombinant will position itself as sister to its major donor in the gene tree, and such a process might explain the clade rearrangement seen in the DHAR2 tree.

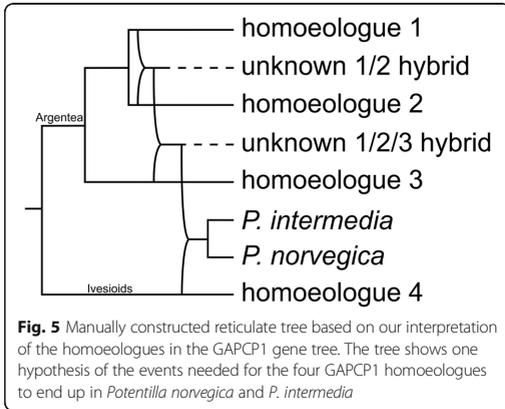
Previous dating analyses have assigned somewhat different ages to the *Potentilla* crown group (excluding *Argentina*), either between ca 36 to 15 Mya [18, 20, 48] or between ca 56 to 32 Mya [44]. Estimations of the Argentea-Ivesioid split also varies, with ages between 15.2–9.8 Mya [18, 48] and 36.6–18.7 Mya [44]. There is also disagreement as to whether the Argentea crown clade is younger [44] or older [18, 48] than the Ivesioid crown clade, but this may be a sampling issue since undersampling of a species rich sister clade would tend to result in underestimating the age of its crown. Today, the Argentea clade consists of the majority of the *Potentilla* species. They have a circumpolar distribution in the Northern Hemisphere, are adapted to a variety of climates, and are of multiple ploidy levels. In contrast, the Ivesioids are limited to dry areas in western United States [21] and are, as far as known, tetraploid [13].

According to Töpel et al. [44] they also evolved in the same area, while Dobeš and Paule [18] estimated an origin in East Asia both for the *Potentilla* crown group and the Ivesioids. However, considering the Ivesioids being geographically restricted and ecologically specialized, the Western American origin of the crown clade found by Töpel et al. [44] may be the most plausible. It is notable, however, that if they are indeed sister groups, their stem lineages are of the same age, and any species that would fall below the crown clades of Argentea or the Ivesioids are either unsampled or extinct.

During the Eocene (56–33.9 Mya [49]), before or in the early stages of the diversification of the *Potentilla* crown group, the North Atlantic land bridge was broken up [50, 51] and the Turgai strait still separated Asia from Europe [51]. A land bridge over the Bering strait existed during most of the later Tertiary to mid Pliocene [51–53], and the original dispersal of the Ivesioid and Argentea ancestors from Asia to North America is most likely to have occurred before its breakup. Today the Bering Strait area is subject to very cold and long winters, but the clade ages suggested by Töpel et al. [44] indicate that the dispersal may have coincided with the Mid Miocene Climatic Optimum, when the Earth was on average 3°C warmer than present [54]. However, considering the current cold climate tolerance of both *P. norvegica* and *P. intermedia* [17, 38], dispersal did not necessarily have to have coincided with warmer periods. Therefore, the younger clade ages estimated by Dobeš and Paule [18] and Feng et al. [20] need not be dismissed.

Regardless of their relative ages, and judging from extant species, the Argentea clade has gone through many more speciations, polyploidizations and hybridizations than the Ivesioid clade. Nonetheless, there is an indication of an early autopolyploid event in the Ivesioids, and this is especially evident in the GAPCP1 tree (Fig. 1); the two subclades in clade C each contain one or two sequences of all Ivesioid species included.

The single *P. norvegica* homoeologue in clade C (Figs. 1, 2 and 3) indicates that the Argentea-Ivesioid hybridization may have happened before polyploidization and diversification of the Ivesioid crown group. This makes the hybridization event difficult to pinpoint geographically; Töpel et al. [44] predicted a wide climate preference for the Ivesioid ancestor, and both *P. norvegica* and *P. intermedia* have weedy growth habits and can be found all around the Northern Hemisphere. Neither is it possible to say, based on our species sample and the resolution of our gene trees, if the Argentea-Ivesioid hybridization is the oldest or the most recent. To illustrate the mode of speciation that *P. norvegica* and *P. intermedia* have gone through, one possible chain of events is shown in Fig. 5 based on our interpretation of the GAPCP1 tree.



The phylogenetic pattern seen in chloroplast markers of Rivalles species (in the sense of Wolf [25]) occurring in North America suggests that other species than *P. norvegica* and *P. intermedia* may have connections to the Argentea-Ivesioid hybridization event [18, 44]. Based on chloroplast data, *P. norvegica*, *P. newberryi*, *P. rivalis* and *P. supina* resolve with the Argentea clade, while *P. biennis* is sister to the Ivesioid clade. For *P. norvegica*, it is evident that the pollen donor came from the Ivesioid clade [23], and therefore it is notable that *P. biennis* is the only Rivalles species that resolves with the Ivesioid clade. *Potentilla biennis*, *P. newberryi* and *P. rivalis* have a limited central to western North American distribution similar to that of the Ivesioids [17]. In addition, *P. biennis* and *P. rivalis* are morphologically similar to *P. norvegica*. Thus, it seems likely that the Argentea-Ivesioid hybridization event occurred in North America rather than in Asia. That would make the Eastern European origin of *P. norvegica*, as proposed by many floras [35, 37, 38], doubtful. It is therefore possible that the Rivalles group originated following multiple hybridization events between the two clades. To better pinpoint where they occurred and which evolutionary routes that were then taken by the lineages that emerged, additional Argentea and Rivalles species of various ploidy levels should be included in future analyses, such that all continents are better covered.

The four homoeologues that were found in *P. norvegica* had a high degree of variation. In the case of *P. intermedia*, this variation seemed even greater, since it is found in more subclades than *P. norvegica*. Both *P. norvegica* and *P. intermedia* have more than one ploidy level reported [13], and there are many other examples of plant populations with mixed ploidy levels [42, 55–57]. Sterile hybrids may still be able to produce offspring through apomixis, and this apomixis is in turn heritable [58]. According to Asker [59], both *P. norvegica* and *P. intermedia* can reproduce in this manner, which could explain the

existence of multiple ploidy levels and high sequence variation within the two species. In addition, several of the putatively diploid species (*P. aurea*, *P. chinensis*, *P. clusiana*, *P. fragarioides* and *P. heptaphylla*) [13] included in this study failed direct sequencing of all markers, and showed a remarkable sequence variation. *Potentilla heptaphylla* was resolved together with *P. argentea* and *P. hirta* in the backbone reference (pp 0.95/0.88) (Fig. 4), but was seen in three different clades (A2, B and C) in the separate gene trees. This suggests allopolyploidy rather than single gene duplications, since the gene copies were resolved as sisters to different species in the same gene tree. The ploidy level of *P. aurea* is difficult to determine solely from the results presented here, since it is found in clades A1 and A2 in the GAPCP1 tree, but only in clade A1 in the GBSSI-1 and DHAR2 trees. However, as seen for *P. norvegica* in the GBSSI-1 and DHAR2 trees, it is possible that *P. aurea* and *P. heptaphylla* have lost homoeologues too. Future studies of polyploid species in *Potentilla* should consider chromosome counting and flow cytometry of the specimens included in order to more securely connect the gene trees with ploidy level, in addition to recreate a more accurate, reticulate species tree.

### Conclusions

This is the first study of species level relationships and reticulate patterns in *Potentilla* based on low copy nuclear markers. With this type of data it was possible to reveal a complex evolutionary history of polyploidizations and hybridizations, not only within previously identified subclades, but also between subclades. The nature of the results, and implications for the interpretation of evolutionary events and distribution patterns, demonstrate the importance of continued work with this kind of data.

The gene trees showed that *P. norvegica* and *P. intermedia* are allopolyploids with multiple parental lineages in the Argentea clade, and one in the Ivesioid clade. This close relationship between the two clades is one of several arguments for an inclusion of the genera of the Ivesioid clade (*Horkelia*, *Horkeliella* and *Ivesia*) in *Potentilla*. This inclusion would help to make *Potentilla* monophyletic.

Gene sequences from both *Potentilla norvegica* and *P. intermedia* are present in the same major clades. This indicates that the allopolyploidy events occurred in their common ancestral lineage.

This study shows no support for species differentiation of *P. norvegica*, as previously suggested, since there was no concordance between geography and intraspecies phylogeny. In addition, the majority of the preserved specimens studied were of intermediate morphological form between the two subspecies. A more extensive study including more specimens is needed in order to determine the support for recognition of the subspecies.

Hybridization between the Argentea and Ivesioid clades may have occurred several times and given rise to the species of Wolf's grex Rivales [25]. To better estimate when and where these hybridizations occurred, other Argentea and Rivales species of various ploidy levels should be included in future studies, such as *P. rivalis* and *P. biennis*.

## Methods

### Taxon selection

To cover the circumpolar distribution of *Potentilla norvegica* L., [60] herbarium material of one morphologically typical individual of each subspecies, ssp. *norvegica* and ssp. *hirsuta* (Michx.) Hyl., were included from Scandinavia and central Europe, in addition to two North American and one eastern Russian specimen of ssp. *hirsuta*. From the Argentea clade, species were selected if they had reported diploid populations [13], and from the Ivesioid clade the type species of *Horkelia* and *Ivesia* were selected. Low-ploidy outgroup species were selected from the Reptans, Fragarioides and Alba clades. *Potentilla intermedia* L. was also included since it shares several features with *P. norvegica*: similar morphology, weedy growth habit and assigned to grex Rivales by Wolf [25], and could therefore be suspected to have a similar evolutionary history as *P. norvegica*. All specimens included are listed in Table 2.

### Primer design

New primer pairs were designed for three low-copy nuclear markers (Table 3); GAPCP1 (glyceraldehyde-3-phosphate dehydrogenase) with primer sites in exons 11 and 14, GBSSI-1 (granule-bound starch synthase 1) in exons 1 and 4 and DHAR2 (dehydroascorbate reductase 2) in exons 4 and 5. In order to find suitable primer placements, the 150 base pair long Illumina raw reads of a *Potentilla argentea* genome (putatively diploid [63]), were assembled using SOAPdenovo2 [64] on the Abel cluster (hosted by the University of Oslo, Norway). Alignments of the resulting contigs to available Rosaceae sequences at GenBank were used to screen for conserved regions in the markers. Candidate sequences were blasted in Geneious version 10.2 [65] to the *Fragaria vesca* genome published at Genbank [66] and to the *P. argentea* contigs to ensure that they would not amplify multiple regions. Annotation was based on the *F. vesca* genome (GAPCP1: XM\_004306515; GBSSI-1: XM\_004306569; DHAR2: XM\_004307358).

The *P. argentea* sequences used in this study were taken from these contigs, and were therefore not produced as the rest of the sequences (see below).

### Molecular methods

#### DNA extraction and PCR

Twenty milligrams of silica gel-dried or herbarium leaf material were extracted using the Qiagen DNeasy Plant

Mini Kit (Qiagen, Valencia, CA, USA) following the manufacturer's instructions, with the exception that the samples were left overnight at 56 °C and then allowed to lyse at 65 °C for 10 min. PCR mixtures included 2.5 µl 10x buffer (Mg<sup>2+</sup> plus, 20 mM), 2 µl dNTP (2.5 mM each), 1 µl forward and reverse primers (10 µM), 0.15 µl TaKaRa Ex Taq HotStart DNA polymerase (5 U/µl) (Takara Bio, Shiga, Japan), 1–2 µl template, and ddH<sub>2</sub>O to add up to 25 µl. The reactions were run on a PCR C1000™ Touch Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA). For GAPCP1, the reactions were amplified through 3 min initial denaturation at 95 °C, followed by 35 cycles of denaturation for 30 s at 95 °C, annealing for 30 s at 51 °C and extension for 1 min at 72 °C. A final extension was performed for 5 min at 72 °C. For GBSSI-1 and DHAR2, the reactions were amplified through a touch-down program with 3 min initial denaturation at 94 °C, followed by 10 cycles of denaturation for 45 s at 94 °C, annealing for 30 s starting at 55 °C and then 0.5 °C lower for each cycle, and 60 s extension at 72 °C. Thirty-five cycles with a constant annealing temperature at 49 °C followed, and a final extension for 7 min at 72 °C. The reactions were checked on a 1% agarose GelRed-stained (Biotium Inc., Fremont, CA, USA) gel under UV light.

### Cloning

Cloning of PCR products was performed on polyploids and specimens failing direct sequencing, using the StrataClone PCR Cloning kit (Agilent Technologies, Santa Clara, CA, USA) following the manufacturer's instructions, with the exceptions that 40 and 80 µl of the transformation mixture were plated and that the reaction mixture was halved for species of lower ploidy level (4x). PCR reactions were performed on positive transformants with primers M13–20 and M13 reverse (as found in the manual) together with Ex-Taq HS polymerase as described above. Amplification started with an initial denaturation for 10 min at 94 °C, followed by 35 cycles of denaturation for 45 s at 94 °C, annealing for 45 s at 55 °C and extension for 3 min at 72 °C. A final extension was performed for 10 min at 72 °C. PCR products were then checked on a 1% agarose gel.

### Purification and sequencing

All PCR products were purified using the Exo-Sap method [67]. The number of clones sequenced corresponded to 95% probability of finding all gene copies, that is at least 6 clones for tetraploids, 11 clones for hexaploids, 16 clones for octoploids and 21 clones for decaploids (Lundberg et al., unpublished). Two species of *Ivesia* have been reported to be tetraploids [61], and therefore the Ivesioid species included in this study were also treated as such. The samples were prepared using a BigDye Terminator Cycle sequencing kit (Applied Biosystems, Waltham, MA,

**Table 2** Voucher list; low-copy markers

Taxon	Voucher	Collection site	Ploidy level	Clade	GAPCP1	GBSSH1	DHAR2	Suffix
<i>Horkelia bolanderi</i> A.Gray	5353 (no voucher)	Cult. in Bergius Botanic Garden, Sweden	4 <sup>a</sup>	Ivesioids	MIN346800, MIN346802	–	–	–
<i>Horkelia californica</i> Cham. & Schtdl.	Balls 9326 (S)	California, USA	4 <sup>a</sup>	Ivesioids	MIN346803, MIN346804	MIN346888, MIN346889, MIN346890	MIN346707, MIN346708	–
<i>Horkiella purpurascens</i> (S.Watson) Rydb.	Eriksson 818 (S)	Cult. in Gothenburg Botanical Garden, Sweden (from wild seeds, California, USA)	4 <sup>a</sup>	Ivesioids	MIN346805, MIN346806, MIN346807	–	–	–
<i>Ivesia gordonii</i> (Hook.) Torr. & A.Gray	Porter 6666 (UPS)	Wyoming, USA	4 <sup>a</sup>	Ivesioids	MIN346808, MIN346809, MIN346810	MIN346891, MIN346892, MIN346893, MIN346894	MIN346709, MIN346710	–
<i>Ivesia kingii</i> S.Watson	Reveal et al. 4782 (MARY)	Nevada, USA	4 <sup>a</sup>	Ivesioids	MIN346811, MIN346812	–	–	–
<i>Ivesia multifoliolata</i> (Torr.) D.D.Keck	Eriksson 820 (S)	Cult. in Bergius Botanic Garden, Sweden	4 <sup>a</sup>	Ivesioids	MIN346813, MIN346814	–	–	–
<i>Potentilla ancistrifolia</i> Franch. & Sav.	2002–0674 (no voucher)	Cult. in Royal Botanic Garden Edinburgh, United Kingdom	4 <sup>a</sup>	Fragarioideis	MIN346826	MIN346895	MIN346711	–
<i>Potentilla argentea</i> L.	Fiatberg, Bendiksbj. Østreng Nygård (TRH)	Sør-Trøndelag, Norway	2,4,6,8	Argentea	MIN346815	MIN346896	MIN346712	–
<i>Potentilla aurea</i> L.	Persson & Eriksson NP36 (BG)	Salzburg, Austria	2	Argentea	MIN346818, MIN346819	–	–	115
<i>Potentilla aurea</i> L.	Persson & Eriksson NP38 (BG)	Carinthia, Austria	2	Argentea	MIN346816, MIN346817	MIN346897, MIN346898	MIN346713, MIN346714, MIN346715	99
<i>Potentilla biflora</i> Willd. ex Schtdl.	Gabrielsen & Jørgensen (O)	Alaska, USA	2	Alba	MIN346820	MIN346899	MIN346716	–
<i>Potentilla brauneana</i> Hoppe	Persson & Eriksson NP31 (BG)	Styria, Austria	2	Argentea	MIN346821	MIN346900	MIN346717	–
<i>Potentilla chinensis</i> Ser.	Guocheng-yong 20, 061–446-5 (S)	Shilazhichen, Xintai Co., Shangdong Province, China	2	Argentea	MIN346822, MIN346823	MIN346901, MIN346902, MIN346903	MIN346718, MIN346719	–
<i>Potentilla clusiana</i> Jacq.	Persson & Eriksson NP29 (BG)	Styria, Austria	2,6	Alba	MIN346824, MIN346825	MIN346904, MIN346905, MIN346906	MIN346720, MIN346721	–
<i>Potentilla dickinsii</i> Franch. & Sav.	Crompton, D'Arcy & Coke 139 (E)	Cult. in Royal Botanic Garden Edinburgh, UK	2	Fragarioideis	MIN346827	–	–	–
<i>Potentilla fragarioideis</i> L.	32,074	Cult. in Botanical Garden of Bonn University, Germany	2	Fragarioideis	MIN346828, MIN346829	MIN346907, MIN346908	MIN346722, MIN346723, MIN346724, MIN346725	–
<i>Potentilla heptaphylla</i> L.	Persson & Eriksson NP28 (BG)	Lower Austria, Austria	2,4,6	Argentea	MIN346830, MIN346831, MIN346832, MIN346833	MIN346909, MIN346910, MIN346911, MIN346912	MIN346726, MIN346727	–

**Table 2** Voucher list; low-copy markers (Continued)

Taxon	Voucher	Collection site	Ploidy level	Clade	GAPC1	GBSSH1	DHAR2	Suffix
<i>Potentilla hirta</i> L.	1962-1846 (no voucher)	Cult. in Royal Botanic Garden Edinburgh, UK	2	Argentea	MN346834	-	-	-
<i>Potentilla intermedia</i> L.	Smedmark 226 (BG)	Hedmark, Norway	4,6,8	Argentea, lvsoids	MN346835, MN346836, MN346837, MN346838, MN346839, MN346840, MN346841, MN346842, MN346843, MN346844	MN346913, MN346914, MN346915, MN346916, MN346917, MN346918, MN346919, MN346920	MN346728, MN346729, MN346730, MN346731, MN346732, MN346733, MN346734, MN346735	
<i>Potentilla norvegica</i> L. ssp. <i>hirsuta</i> (Michx.) Hy.	Eriksen & Topel (GB)	Kamchatka, Russia	8,10	Argentea, lvsoids	MN346876, MN346877, MN346878, MN346879	MN346949, MN346950, MN346951, MN346952, MN346953, MN346954, MN346955	MN346782, MN346786, MN346787, MN346788, MN346791, MN346792	112
<i>Potentilla norvegica</i> L. ssp. <i>hirsuta</i> (Michx.) Hy.	Gillespie, Saarela, Consaul & Bull 7439 (O)	Northwest Territories, Canada	8,10	Argentea, lvsoids	MN346880, MN346881, MN346882, MN346883	MN346956, MN346957, MN346958	MN346793, MN346794, MN346795, MN346796, MN346798, MN346799	113
<i>Potentilla norvegica</i> L. ssp. <i>hirsuta</i> (Michx.) Hy.	Persson NP45 (BG)	Cult. in Bergen Museum Garden, Norway (from wild seeds, Salzburg, Austria)	8,10	Argentea, lvsoids	MN346871, MN346872, MN346873, MN346874, MN346875	MN346945, MN346946, MN346947, MN346948	MN346773, MN346774, MN346776, MN346778, MN346779, MN346781	110
<i>Potentilla norvegica</i> L. ssp. <i>hirsuta</i> (Michx.) Hy.	Rouleau 6057 (UPS)	Newfoundland, Canada	8,10	Argentea, lvsoids	MN346845, MN346846, MN346847	MN346921, MN346922, MN346923, MN346924, MN346925	MN346736, MN346737, MN346738, MN346740, MN346741	92
<i>Potentilla norvegica</i> L. ssp. <i>hirsuta</i> (Michx.) Hy.	Svenson AS07135 (S)	Uppland, Sweden	8,10	Argentea, lvsoids	MN346855, MN346856, MN346857, MN346858, MN346859, MN346860, MN346861, MN346862, MN346863, MN346864	MN346931, MN346932, MN346933, MN346934, MN346935, MN346936	MN346747, MN346749, MN346750, MN346751, MN346753, MN346754, MN346755, MN346756, MN346757	96
<i>Potentilla norvegica</i> L. ssp. <i>norvegica</i>	Eriksson 1058 (BG)	Oppland, Norway	8,10	Argentea, lvsoids	MN346865, MN346866, MN346867, MN346868, MN346869, MN346870	MN346937, MN346938, MN346939, MN346940, MN346941, MN346942, MN346943, MN346944	MN346758, MN346759, MN346760, MN346761, MN346764, MN346765, MN346766, MN346767, MN346768, MN346769, MN346770, MN346771	97
<i>Potentilla norvegica</i> L. ssp. <i>norvegica</i>	Gugnacka (UPS)	Kujawsko-Pomorske, Poland	8,10	Argentea, lvsoids	MN346848, MN346849, MN346850, MN346851, MN346852, MN346853, MN346854	MN346926, MN346927, MN346928, MN346929, MN346930	MN346743, MN346744, MN346745, MN346746	95
<i>Potentilla reptans</i> L.	Salvesen 16.45 (BG)	Vest-Agder, Norway	4	Reptans	MN346884, MN346885, MN346886, MN346887	MN346959, MN346960, MN346961, MN346962	-	-

Ploidy level (CCDB [13], IPCN [61]), clade [23] and Genbank accessions

\*Based on *liesia baileyi* var. *beneolens* (A.Nelson & J.F.Macbr.) Erter and *I. rhypana* var. *shellyi* Erter

**Table 3** Primer sequences used for PCR and sequencing

Marker	Primer name	Sequence 5'-3'	Reference
GAPCP1	11F	TGT CGA CTT GAG AAG GGT GGT TC	This paper
	14R	CTT ATG CTG CCA CCA ATG CCA TG	This paper
	CGPPB5575 FWD	CAT GTG CTC TAT GAG GTC CA	[62]
	CGPPB5575 REV	ATC AGG TAT GCT GCT GAT GG	[62]
GBSSI-1	1F	TGG AG CAA GAC TGG TGG ACT TG	This paper
	4R	GCA CAA CAA GCT GAA TCT AAG TTG G	This paper
DHAR2	4F	AAG TAC ACT GAG GTA TGC TGT TC	This paper
	5R	GTT GAC TTT CGG CTC CCA TC	This paper
Cloning vector	M13-20	GTA AAA CGA CGG CCA G	StrataClone manual
	M13 Reverse	CAG GAA ACA GCT ATG AC	StrataClone manual

USA) and run on an ABI 3730XL DNA Analyser (Applied Biosystems). For DHAR2, some samples were sent to MacroGen Sequencing Service (Amsterdam, The Netherlands) after purification. All other molecular lab-work was carried out at the Biodiversity Laboratories (DNA Section) at the University of Bergen.

#### Sequence treatment and alignment

For each marker, forward and reverse reads for each specimen or clone were assembled using PreGap4 and Gap4 of the Staden Package [68]. Automatic alignment of each cloned species separately (and specimen, in the case of *P. norvegica*) was performed in AliView v. 1.18 [69] using MUSCLE [70]. Putative PCR errors were corrected and identical sequences were removed. An alignment with all *P. norvegica* specimens was then performed, in order to remove identical sequences shared between individuals.

To detect PCR recombinants, the alignments of cloned specimens were loaded into SplitsTree v. 4.14.4 [71]. Sequences identified as putative PCR recombinants had no, or very short, individual edges and long, parallel, connecting edges to their parental sequences [72]. All remaining sequences were automatically aligned together in AliView followed by manual adjustments.

#### Phylogenetic analyses

##### Model testing

Substitution model testing was performed on each marker with PartitionFinder2 [73], with GAPCP1 and GBSSI-1 divided into subsets of introns and the three codon positions, under the BIC and AICc criteria for the models available in MrBayes. DHAR2 was not divided into subsets, since the amplified region almost exclusively consists of the intron between exons 4 and 5.

##### Indel coding

Indels found in two or more sequences were manually coded according to the Simple Indel Coding method as present (1), absent (0) or inapplicable (N) [74].

##### Bayesian inference

Bayesian Inference analyses were run for each marker separately in MrBayes v. 3.2.6 [75, 76], using the Metropolis Coupled Markov Chain Monte Carlo algorithm [77], including one cold chain and three heated chains for each of two runs. Division of the alignments into subsets and assignment of models were coded according to the results from PartitionFinder2 (Table 1). The Mk model [78] was applied for the indels, where the likelihood prior Coding and rate prior were set to variable. The analyses were run for 5 million generations for GAPCP1 and GBSSI-1, and 7.5 million generations for DHAR2, with sampling from the chain every 1000th generation and with a burnin of 20%. An analysis was accepted if the standard deviation of split frequencies was below 0.01, the chain swap was between 20 and 80% [79] (McGuire et al. 2007), no trend was seen in the overlay plot and the Potential Scale Reduction Factor [80] values had reached 1.0 for all parameters. A clade was fully accepted if its Bayesian posterior probability was 0.95 or higher. In order for the DHAR2 analysis to converge, 13 *P. norvegica* sequences and one *P. intermedia* sequence that were suspected to cause problems had to be removed. These were identified by inspecting the whole dataset in SplitsTree. PartitionFinder2 and MrBayes were run at the CIPRES Science Gateway [81].

##### Maximum likelihood

Maximum Likelihood analyses were performed in RAXML version 7.2.8 [82, 83], under the GTR + G (nucleotides, DNA) [84] and Mk (indels, MULTI) [71] models with

1000 rapid bootstrap replicates [85]. A clade was fully accepted if its Bootstrap support was 75 or higher.

#### Rooting and tree graphics

The resulting consensus trees from the BI and ML analyses were inspected using FigTree version 1.4.1 [86] and rooted on *P. biflora* and *P. clusiana* Jacq. of the Alba clade. The Alba clade is the sister clade to the rest of the species included in this study [18, 23]. All branches with posterior probabilities below 0.8 were collapsed in Mesquite version 3.10 [87]. The layouts were further edited using GIMP version 2.8.10 ([www.gimp.org](http://www.gimp.org)) and Inkscape version 0.48 ([www.inkscape.org](http://www.inkscape.org)).

#### Control analyses

To ensure that no gene copies were incorrectly discarded as PCR recombinants, all unique sequences of the Ivesioids (*Horkelia*, *Horkeliella* and *Ivesia*), *P. intermedia* and *P. norvegica* were subjected to an ML analysis each (without coded indels), together with a reduced dataset of the species representing the larger clades seen in the gene trees.

#### Multispecies coalescent analysis

Due to initial results from the BI and ML analyses showing somewhat different topologies for the different

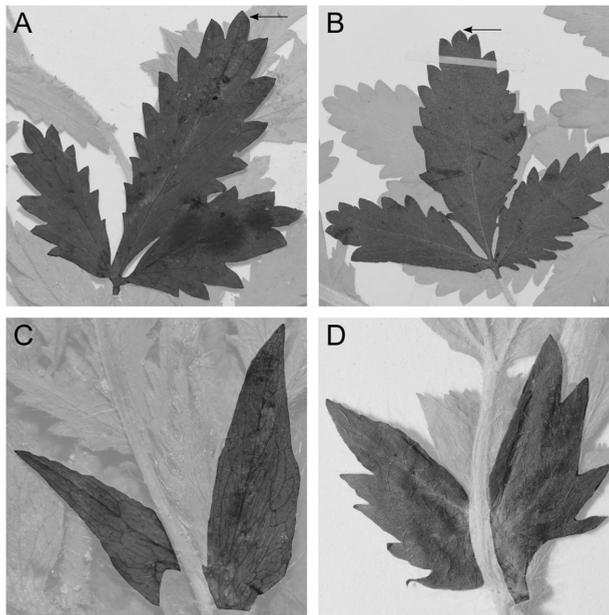
markers, some species were subjected to a Multispecies Coalescent analysis [88] in BEAST v. 1.8.0 [89] at CIPRES [81], in order to create a species tree as a backbone reference. Two datasets were created, one with the three low-copy markers only, and one with the low-copy markers in combination with three chloroplast regions from previous studies (trnL-F, trnC-ycf6 and trnS-ycf9) (Table 4) [18, 44, 90]. Substitution model testing was performed in Partition-Finder2 on each region, not accounting for codon positions. Two clock models were tested; strict and relaxed uncorrelated log normal [91]. For each of these, two tree priors were tested; a birth-death process [92] and a birth process [93]. The analysis of the dataset with low-copy markers only was run for 50 million generations with sampling every 1000th generation, and the combined dataset for 150 million generations with sampling every 1000th generation. To test the fit of the models to the data, path sampling and stepping-stone sampling [94, 95] were performed with 50 steps, each with a length of 1 million iterations for the low-copy marker dataset, and 150 steps with a length of 1 million iterations for the combined dataset. Log marginal likelihood differences larger than three were considered significant [96]. The analysis with the models best fit to the data was run two independent times, and the results were inspected using Tracer v. 1.7.1 [97]. In order to test if the prior, rather than the data,

**Table 4** Voucher list; chloroplast markers

Taxon	Voucher	Collection site	Ploidy level	Clade	trnL-trnF	trnS-ycf9	trnC-ycf6
<i>Horkelia californica</i> Cham. & Schtdl.			4 <sup>a</sup>	Ivesioids	FR872958	–	–
<i>Horkeliella purpurascens</i> (S.Watson) Rydb.	Erterter 4980 (UC)	California, USA	4 <sup>a</sup>	Ivesioids	GQ384737	GQ384569	GQ384891
<i>Ivesia gordonii</i> (Hook.) Torr. & A.Gray	Huber 1182 (MO)	Utah, USA	4 <sup>a</sup>	Ivesioids	GQ384725	GQ384557	–
<i>Potentilla argentea</i> L.	Gregor (HEID)	Rhineland-Palatinate, Germany	2,4,6,8	Argentea	GQ384652	GQ384485	GQ384820
<i>Potentilla argentea</i> L.	Dobeš (HEID)	Lower Austria, Austria	2,4,6,8	Argentea	GQ384665	GQ384497	GQ384833
<i>Potentilla argentea</i> L.	Krämer et al., Botanical Garden Bonn	North Rhine-Westphalia, Germany	2,4,6,8	Argentea	GQ384675	GQ384507	GQ384843
<i>Potentilla aurea</i> L.	Dobeš (HEID)	Tyrol, Austria	2	Argentea	GQ384667	GQ384499	GQ384835
<i>Potentilla aurea</i> L.	Paule (HEID)	Julian Alps, Slovenia	2	Argentea	GQ384673	GQ384505	GQ384841
<i>Potentilla biflora</i> Willd. ex Schtdl.	S. Kharkevich, T. Buch	Magadan Oblast, Russia	2	Alba	GQ384682	GQ384514	GQ384850
<i>Potentilla brauneana</i> Hoppe	Dobeš (HEID)	Tyrol, Austria	2	Argentea	GQ384668	GQ384500	GQ384836
<i>Potentilla chinensis</i> Ser.	Zhechai (CM)	China	2	Argentea	KT991783	–	–
<i>Potentilla clusiana</i> Jacq.	Leopoldinger, University of Salzburg Botanical Garden	Upper Austria, Austria	2,6	Alba	GQ384640	GQ384473	GQ384808
<i>Potentilla heptaphylla</i> L.	Dobeš (HEID)	Lower Austria, Austria	2,4,6	Argentea	GQ384666	GQ384498	GQ384834
<i>Potentilla hirta</i> L.	Dobeš (HEID)	Alpes-Maritimes, France	2	Argentea	GQ384634	GQ384467	GQ384802
<i>Potentilla reptans</i> L.	Botanical Garden Nantes Mairie	Pays de la Loire, France	4	Reptans	GQ384638	GQ384471	GQ384806

Ploidy level (CCDB [13], IPCN [61]), clade [23] and Genbank accessions

<sup>a</sup>Based on *Ivesia baileyi* var. *beneolens* (A.Nelson & J.F.Macbr.) Erterter and *I. rhypara* var. *shellyi* Erterter



**Fig. 6** Images of *Potentilla norvegica* leaves and stipules, illustrating the typical characters for *ssp. norvegica* and *ssp. hirsuta*: Leaflet and tooth shape (**a** and **b**) and stipule dentation (**c** and **d**). The specimen in **a** and **c** was collected by Gugnacka s.n. (UPS) in Poland, here denoted individual 95. The specimen in **b** and **d** was collected by Rouleau nr. 6057 (UPS) in Canada, here denoted individual 92

was driving the results, an additional run with sampling from prior only was performed. The tree files were then combined using TreeAnnotator of the BEAST package with a burnin of 20% of each run.

### Morphological study

*Potentilla norvegica* specimens were inspected at, or on loan from, the herbaria of Stockholm (S), Uppsala (UPS) and Gothenburg (GB) in Sweden, and the herbaria of Bergen (BG) and Oslo (O) in Norway. They were used to study the defining characters of the two *P. norvegica* subspecies (*ssp. norvegica* and *ssp. hirsuta*); leaflet form, leaflet dentation and stipule dentation [29, 38, 98] (Fig. 6 and Table 5).

**Table 5** Morphological characters used to differentiate between the two subspecies of *Potentilla norvegica*

	<i>ssp. norvegica</i>	<i>ssp. hirsuta</i>
<b>Basal leaflets</b>	oblongeolate	obovate
	acute terminal tooth	obtuse terminal tooth
	long terminal tooth	short terminal tooth
<b>Stipule teeth</b>	0–3	2 – several

### Acknowledgements

We are grateful to the curators and other staff at the herbaria at the University Museum of Bergen (BG), University of Gothenburg (GB), University of Oslo (O), Swedish Museum of Natural History (S), Uppsala University (UPS), Natural History Museum, Vienna (W) and University of Vienna (WU) for assistance during visits, field trips and with loans; and the technicians at the Biodiversity Laboratory at the University of Bergen for help with the labwork. We also thank the other members of the Rosaceae research group at UiB for fruitful discussions, especially Ingrid Toresen who produced several of the GAPCP1 sequences. The unpublished raw reads of the *Potentilla argentea* genome were kindly provided by Malene Østreng Nygård and Mika Bendiksy at the Norwegian University of Science and Technology, Trondheim.

### Authors' contributions

NLP, TE and JEEES contributed intellectually to the work and approved the manuscript for submission. JEEES designed the study. NLP, TE and JEEES authors collected the study material. NLP carried out the molecular labwork, statistical analyses, morphological study and designed the graphics. NLP led the writing of the manuscript but TE and JEEES contributed.

### Funding

This work was supported by the University Museum of Bergen and the Olaf Grolle Olsen fund at the University of Bergen to NLP, TE and JEEES. The funding sources had no participation in the design or execution of this study.

### Availability of data and materials

Most specimens included are deposited at herbaria (see Table 1). Vouchers are missing for a few of the specimens obtained from botanical gardens. The

DNA sequences are deposited at GenBank under the accession numbers [MN346707–MN346962].

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 7 December 2018 Accepted: 24 February 2020

Published online: 18 March 2020

#### References

- Werth CR, Guttman SJ, Eshbaugh WH. Electrophoretic evidence of reticulate evolution in the Appalachian *Asplenium* complex. *Syst Bot.* 1985;10(2):184–92.
- Takamiya M, Tanaka R. Polyploid cytotypes and their habitat preferences in *Lycopodium clavatum*. *Bot Mag Tokyo.* 1982;95:419–34.
- Ownbey M. Natural hybridization and amphiploidy in the genus *Tragopogon*. *Am J Bot.* 1950;37(7):487–99.
- Kihara H, Ono T. Chromosomenzahlen und systematische Gruppierung der *Rumex*-Arten. *Z Zellforsch Mik Ana.* 1926;4:475–81.
- Stebbins GL. Variation and evolution in plants. New York: Columbia University Press; 1950.
- Grant V. Plant speciation. 1st ed. New York: Columbia University Press; 1971.
- Rieseberg LH, Willis JH. Plant speciation. *Science.* 2007;317(5840):910–4.
- de Queiroz K. A unified concept of species and its consequences for the future of taxonomy. *Proc Calif Acad Sci.* 2005;56(18):196–215.
- Evans RC, Campbell CS. The origin of the apple subfamily (Malioideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *Am J Bot.* 2002;89(9):1478–84.
- Lundberg M, Töpel M, Eriksen B, Nylander JAA, Eriksson T. Allopolyploidy in *Fragariaeae* (Rosaceae): comparing four DNA sequence regions, with comments on classification. *Mol Phylogenet Evol.* 2009;51:269–80.
- Kessler M, Kühn A, Solís Neffa VG, Hensen I. Complex geographical distribution of ploidy levels in *Polyplepis australis* (Rosaceae), an endemic tree line species in Argentina. *Int J Plant Sci.* 2014;175(8):955–61.
- Vamosi JC, Dickinson TA. Polyploidy and diversification: a phylogenetic investigation in Rosaceae. *Int J Plant Sci.* 2006;167(2):349–58.
- Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. *New Phytol.* 2014;206(1):19–26.
- Eriksson T, Donoghue MJ, Hibbs MS. Phylogenetic analysis of *Potentilla* using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of Rosoideae (Rosaceae). *Pl Syst Evol.* 1998;211:155–79.
- Eriksson T, Hibbs MS, Yoder AD, Delwiche CF, Donoghue MJ. 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. *Int J Plant Sci.* 2003;164(2):197–211.
- Eriksson T, Lundberg M, Töpel M, Östenson P, Smedmark JEE. *Sibbaldia*: a molecular phylogenetic study of a remarkably polyphyletic genus in Rosaceae. *Pl Syst Evol.* 2015;301:171–84.
- Erter B, Elven R, Reveal JL, Murray DF. *Potentilla*. In: Flora of North America editorial committee, editors. 1993+. Flora of North America north of Mexico. 20+ vols, vol. 9. New York: Oxford University Press; 2014. p. 121–218.
- Dobeš C, Paule J. A comprehensive chloroplast DNA-based phylogeny of the genus *Potentilla* (Rosaceae): implications for its geographic origin, phylogeography and generic circumscription. *Mol Phylogenet Evol.* 2010;56:156–75.
- Soják J. *Argentina Hill*, a genus distinct from *Potentilla* (Rosaceae). *Thaiszia - J Bot.* 2010;20:91–7.
- Feng T, Moore MJ, Li J, Wang H. Phylogenetic study of the tribe *Potentilleae* (Rosaceae), with further insight into the disintegration of *Sibbaldia*. *J Syst Evol.* 2017;55(3):177–91.
- Erter B, Reveal JL, vesia, Horkelia & Horkeliella. In: Flora of North America editorial committee, editor. editors. 1993+ Flora of North America north of Mexico. 20+ vols, vol. 9. New York: Oxford University Press; 2014a. p. 219–72.
- Erter B, Reveal JL, Duchesnea. In: Flora of North America editorial committee, editor. 1993+ Flora of North America north of Mexico. 20+ vols, vol. 9. New York: Oxford University Press; 2014b. p. 275–6.
- Töpel M, Lundberg M, Eriksson T, Eriksen B. Molecular data and ploidal levels indicate several putative allopolyploidization events in the genus *Potentilla* (Rosaceae). *PLoS Curr.* 2011;3:RRN1237.
- Wendel JF, Doyle JJ. Phylogenetic incongruence: Window into genome history and molecular evolution. In: Soltis P, Soltis D, Doyle J, editors. Molecular systematics of plants. II. Dordrecht: Kluwer Academic Press; 1998. p. 265–96.
- Wolf T. Monographie der Gattung *Potentilla*. *Bibliotheca Botanica.* 1908;16(71):1–714.
- Linnaeus C. *Species plantarum*. Stockholm: Salvius; 1753. Vol. 1, p. 1109.
- Michaux A. *Flora boreali americana*. Paris. 1803;1:303.
- Ledebour CF. *Flora Rossica*, vol. 2. Stuttgart: Sumptibus Librariae E. Schweizerbart; 1844. p. 36.
- Ascherson P, Graebner P. Synopsis der Mitteleuropäischen Flora. Leipzig: Engelmann; 1904. Vol. 6 pt.1 p. 748.
- IPNI - International Plant Names Index. The Royal Botanic Gardens, Kew, Harvard University Herbaria & Libraries and Australian National Botanic Gardens. <http://www.ipni.org>. Accessed 14 Sept 2018.
- Tropicos. Missouri Botanical Garden. <http://www.tropicos.org/Name/27802541>. Accessed 15 Sept 2018.
- Hylander N. Studien über nordische Gefäßpflanzen. Uppsala Univ Årsskr. 1945;1945(7):203.
- Turland NJ, Wiersma JH, Barrie FR, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Kusber W-H, Li D, Marhold K, May TW, McNeill J, Monro AM, Prado J, Price MJ, Smith GF. International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. *Regnum Vegetabile* 159. Glashütten: Koeltz Botanical Books; 2018.
- Hitchcock CL, Conquist A. Vascular plants of the Pacific northwest. Part 2: Salicaceae to Saxifragaceae. Seattle: University of Washington Press; 1968.
- Tutin TG, Heywood VH, Burges NA, Valentine DH, Walters SM, Webb DA. *Flora Europaea. Rosaceae to Umbelliferae*, vol. 2. London: Cambridge University Press; 1968. p. 42.
- Hultén E. The circumpolar plants. II – Dicotyledons. *Kungliga Vetenskapsakademiens Handlingar. Fjärde serien, band 13, nr. 1*. Stockholm: Almqvist & Wiksell; 1971.
- Kurto A, Lampinen R, Junikka L. *Atlas Florae Europaeae. Distribution of vascular plants in Europe*, vol. 13. Helsinki: Societas Biologica Fennica Vanamo; 2004.
- Lid J, Lid DT. *Potentilla L.* In: Elven R, editor. *Norsk flora*. 7th ed. Oslo: Det Norske Samlaget; 2013. p. 413–23.
- Small RL, Cronn RC, Wendel JF. Use of nuclear genes for phylogeny reconstruction in plants. *Aust Syst Bot.* 2004;17:145–70.
- Smedmark J, Eriksson T, Bremer B. Allopolyploid evolution in *Geinae* (Colurieae: Rosaceae) – building reticulate species trees from bifurcating gene trees. *Org Divers Evol.* 2005;5:275–83.
- Maddison WP. Gene trees in species trees. *Syst Biol.* 1997;46(3):523–36.
- Hasegawa M, Kishino H, Yano T. Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 1985;22(2):160–74.
- Rydberg KA. A monograph of the North American *Potentilleae*, vol. 2. Lancaster: The New Era Printing Company; 1898. p. 47.
- Töpel M, Antonelli A, Yesson C, Eriksen B. Past climate change and plant evolution in Western North America: a case study in Rosaceae. *PLoS One.* 2012;7(12):e50358.
- Comai L. The advantages and disadvantages of being polyploid. *Nat Rev Genet.* 2005;6:836–46.
- Leitch IJ, Bennett MD. Genome downsizing in polyploids. *Biol J Linn Soc.* 2004;82:651–63.
- Bento M, Gustafsson JP, Viegas W, Silva M. Size matters in *Triticaceae* polyploids: larger genomes have higher remodeling. *Genome.* 2011;54:175–83.
- Xiang Y, Huang C, Hu Y, Wen J, Li S, Tingshuang Y, Chen H, Xiang J, Ma H. Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol Biol Evol.* 2017;34(2):262–81.
- Cohen KM, Finney SC, Gilbard PL, Fan J. The ICS international Chronostratigraphic chart. *Episodes.* 2013;36(3):199–204.
- Dietz RS, Holden JC. Reconstruction of Pangaea: breakup and dispersion of continents, Permian to present. *J Geophys Res.* 1970;75(26):4939–56.
- Tiffney BH. The Eocene North Atlantic land bridge: its importance in tertiary and modern phytogeography of the northern hemisphere. *J Arnold Arboretum.* 1985;66:243–73.

52. Hopkins DM. Cenozoic history of the Bering land bridge. *Science*. 1959; 129(3362):1519–28.
53. Tiffney BH, Manchester SR. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the northern hemisphere tertiary. *Int J Plant Sci*. 2001;162(6):3–17.
54. You Y, Huber M, Müller RD, Poulsen CJ, Ribbe J. Simulation of the middle Miocene climate optimum. *Geophys Res Lett*. 2009;36:L04702.
55. Eriksen B. Mating systems in two species of *Potentilla* from Alaska. *Folia Geobot Phytotax*. 1996;31:333–44.
56. Renny-Byfield S, Ainouche M, Leitch IJ, Lim Y, Le Comber SC, Leitch AR. Flow cytometry and GISH reveal mixed ploidy populations and *Spartina* nonaploids with genomes of *S. alterniflora* and *S. maritima* origin. *Ann Bot-London*. 2010;105(4):527–33.
57. Čertner M, Fenclová E, Kúr P, Kolář F, Koutecký P, Krahulcova A, Suda J. Evolutionary dynamics of mixed-ploidy populations in an annual herb: dispersal, local persistence and recurrent origins of polyploids. *Ann Bot-London*. 2017;120:303–15.
58. Müntzing A. Heteroploidy and polymorphism in some apomictic species of *Potentilla*. *Hereditas*. 1958;44(2–3):280–329.
59. Asker S. Apomictic biotypes in *Potentilla intermedia* and *P. norvegica*. *Hereditas*. 1970;66(1):101–7.
60. GBIF. Global Biodiversity Information Facility Home Page. 2018. <https://www.gbif.org>. Accessed 2 July 2018.
61. IPCN - Index to Plant Chromosome Numbers. Goldblatt, P., Johnson, D. E. editors. Missouri Botanical Garden, St. Louis. 1979–. <http://www.tropicos.org/Project/IPCN>. Accessed 7 May 2017.
62. Le Dantec L, Cardinet G, Bonet J, Fouché M, Boudehri K, Monfort A, Poëssel J, Moing A, Dirlewanger E. Development and mapping of peach candidate genes involved in fruit quality and their transferability and potential use in other Rosaceae species. *Tree Genet Genomes*. 2010;6:995–1012.
63. Paule J, Sharbel TF, Dobeš C. Apomictic and sexual lineages of the *Potentilla argentea* L. group (Rosaceae): Cytotype and molecular genetic differentiation. *Taxon*. 2011;60(3):721–32.
64. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T, Wang J. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*. 2012;1:18.
65. Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647–9.
66. 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 J, DJG J, Williams KP, Holt SH, Ruiz Rojas JJ, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin SA, Troglio M, Viola R, Ashman T, Wang H, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant DW Jr, Fox SE, Givan SA, Wilhelm LJ, Naithani S, Christoffels AY, 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, Fliinn B, Aharoni A, Bennetzen JL, Salzberg SL, Dickerman AW, Velasco R, Borodovsky M, Veilleux RE, Folta KM. The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet*. 2011;43(2):109–16.
67. Dugan K, Lawrence H, Hares D, Fisher C, Budowle B. An improved method for post-PCR purification for mtDNA sequence analysis. *J Forensic Sci*. 2002;47(4):1–8.
68. Staden R. The Staden sequence analysis package. *Mol Biotechnol*. 1996;5:233.
69. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014;30(22):3276–8.
70. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–7.
71. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006;23(2):254–67.
72. Marcussen T, Heier L, Bryusting AK, Oxelman B, Jakobsen KS. From gene trees to a dated allopolyploid network: insights from the angiosperm genus *Viola* (Violaceae). *Syst Biol*. 2015;64(1):84–101.
73. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol*. 2016;34(3):772–3.
74. Simmons MP, Ochoterena H. Gaps as characters in sequence-based phylogenetic analyses. *Syst Biol*. 2000;49(2):369–81.
75. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*. 2001;17:754–5.
76. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–42.
77. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*. 2004;20:407–15.
78. Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol*. 2001;50(6):913–25.
79. McGuire JA, Witt CC, Altshuler DL, Remsen JV. Phylogenetic systematics and biogeography of humming-birds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. *Syst Biol*. 2007;56:837–56.
80. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457–72.
81. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans: Heidelberg Institute for Theoretical Studies; 2010. p. 1–8.
82. Stamatakis A. RaxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22(21):2688–90.
83. Stamatakis A. RaxML version 7.2.8. 2010. <https://sco.h-its.org/exelixis/web/software/raxml/>.
84. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on mathematics in the life sciences. Am Mathematical Soc. 1986;17:57–86.
85. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RaxML web servers. *Syst Biol*. 2008;57(5):758–71.
86. Rambaut A. FigTree: Tree figure drawing tool version 1.4.2. <http://tree.bio.ed.ac.uk/software/figtree>: Institute of Evolutionary Biology, University of Edinburgh; 2014.
87. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.10. 2016. <http://mesquiteproject.org>.
88. Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. *Mol Biol Evol*. 2009;27(3):570–80.
89. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUTI and the BEAST 1.7. *Mol Biol Evol*. 2012;29:1969–73.
90. Koski MH, Ashman T. Macroevolutionary patterns of ultraviolet floral pigmentation explained by geography and associated bioclimatic factors. *New Phytol*. 2016;211(2):708–18.
91. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4(5):e88.
92. Kendall DG. On the generalized "birth-and-death" process. *Ann Math Stat*. 1948;19:1–15.
93. Yule GU. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos T Roy Soc B*. 1925;213:21–87.
94. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol*. 2012;29:2157–67.
95. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol*. 2013;30:239–43.
96. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995;90:773–95.
97. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarisation in Bayesian phylogenetics using tracer 1.7. *Syst Biol*. 2018;67(5):901–4.
98. Mossberg B, Stenberg L. *Gyldendals store nordiske flora*. 2nd ed: Hung Hing: Gyldendal; 2014.

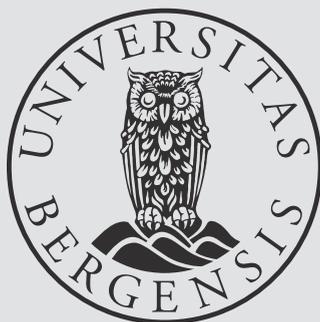
## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





Graphic design: Communication Division, UIB / Print: Skjipes Kommunikasjon AS



[uib.no](http://uib.no)

ISBN: 9788230840573 (print)  
9788230858813 (PDF)