Clarification of basal relationships in Rubus (Rosaceae) and the origin of Rubus chamaemorus

Karen Michael
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses

Part of the Fruit Science Commons, and the Medical Sciences Commons

Recommended Citation

http://digitalcommons.wku.edu/theses/250

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
CLARIFICATION OF BASAL RELATIONSHIPS IN *RUBUS* (ROSACEAE) AND THE ORIGIN OF *RUBUS CHAMAEMORUS*

A Thesis
Presented to
The Faculty of the Department of Biology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Biology

By
Karen Ambrose Michael
December 2006
CLARIFICATION OF BASAL RELATIONSHIPS IN RUBUS (ROSACEAE) AND THE ORIGIN OF RUBUS CHAMAEMORUS

Date Recommended 8/17/06
Director of Thesis

Date 9/18/2006
Dean, Graduate Studies and Research
ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge the fact that this research project would not have been possible without the blessings, guidance and love of God. The past two years have been a pleasant and eventful time in my life, both professionally and personally. I would like to express my utmost gratitude to my Masters degree advisor Dr. Lawrence Alice, without whom I would not have been able to embark on this wonderful journey discovering the pleasures of evolutionary biology. His unflinching patience, unparalleled knowledge, constant hard work, generous financial assistance and unwavering support all these times has made this effort what it is today. I would like to sincerely thank Dr. Claire Rinehart and Dr. Keith Philips, my wonderfully supportive graduate committee advisors. Their kindness, understanding, and co-operation have helped me gain a better understanding of the bigger picture of being in graduate school.

My sincere appreciation to Dr. Sigrid Jacobshagen, who graciously accepted my request to be a proxy for Dr. Philips during my defense. Thank you to Dr. John Andersland, who always lightened up even the dullest of moments with his smile and friendliness; my poster presentation at the Botany 2006 conference would not have been a success without him. I would like to express my gratefulness to Dr. Michael Stokes, who was always there to give moral support.

To all my dear friends in the Department of Biology, this journey would not have been humanly possible without all of you. Your love, encouragement, wisdom, kindness and emotional support made this place my home away from home, thank you from the bottom of my heart. I would like to especially show my appreciation to Brittany Sutherland, Michael Dodson, Ashley Wint, Kate Hertweck, Natalina Elliot, Dr. Karen
Bell, Clement Akotsen-Mensah, Eva Ngulo, Cassandra Cantrell, Tia Hughes, Ali Wright, Christa Gaskill, and Debbie Wolfe. My sincere appreciation to all faculty members for your kindness, and the lovely office staffs, Jessica Harazin and Erin Cummings for your patience. Lastly but certainly not the least, I would not be the person I am today without my loving family. Appa, Amma, Kaminiy, Anneh, Susan, and Shankar, hope this makes some sense to you guys!
# TABLE OF CONTENTS

Acknowledgements ........................................................................................................ iii  
Table of Contents ........................................................................................................... v  
List of Tables ................................................................................................................ vii  
List of Figures ............................................................................................................... viii  
Abstract ......................................................................................................................... x  

Introduction 1  
  Traditional Classification 1  
  Definition of Basal Species 2  
  Characteristics of the Basal Species in *Rubus* 3  
  Previous Studies in *Rubus* Systematics 6  
  Objectives 7  
  DNA Sequences 8  
  Scope of the Current Study 10  

Materials and Methods 12  
  Plant Samples 12  
  Polymerase Chain Reaction (PCR) 13  
  Cloning 14  
  Sequencing PCR Products and Clones 15  
  Phylogenetic Analyses 15  

Results 16  
  Sequence Characteristics 16  
  Phylogenetic Relationships of *Rubus* 17
Discussion

Phylogeny of Rubus

Putative Allopolyploid Origin of Octoploid R. chamaemorus

Molecular Evolution of Duplicated Genes

Literature Cited
LIST OF TABLES

Table 1. Rubus accessions and outgroup .................................................. 33
Table 2. trnS-trnG and GBSSI-1 sequence characteristics ......................... 36
Table 3. Characterization of GBSSI-1 clone sequences for basal species of Rubus and Dalibarda repens based on position in the GBSSI-1 phylogeny (Fig. 15). ................................................................. 37
LIST OF FIGURES

Figure 1. Image of (a) *R. deliciosus* and
(b) geographic distribution in the US ............................................. 38

Figure 2. Image of (a) *R. neomexicanus* and
(b) geographic distribution in the US ............................................. 39

Figure 3. Image of (a) *R. odoratus* and
(b) geographic distribution in the US ............................................. 40

Figure 4. Image of (a) *R. parviflorus* and
(b) geographic distribution in the US ............................................. 41

Figure 5. Image of *R. trilobus* from Mexico .................................. 42

Figure 6. Image of (a) *R. chamaemorus* and
(b) geographic distribution in the US ............................................. 43

Figure 7. Image of (a) *R. lasiococcus* and
(b) geographic distribution in the US ............................................. 44

Figure 8. Image of (a) *R. pedatus* and
(b) geographic distribution in the US ............................................. 45

Figure 9. Image of (a) *Dalibarda repens* and
(b) geographic distribution in the US ............................................. 46

Figure 10. Organization of the chloroplast trnS-trnG region ................. 47

Figure 11. Organization of the nuclear GBSSI-1 gene in Rubus .............. 48

Figure 12. Strict consensus phylogeny based on
nuclear ribosomal ITS sequences ............................................... 49

Figure 13. Strict consensus phylogeny based on
$trnS-trnG$ sequences .......................................................... 50

Figure 14. Strict consensus phylogeny based on
combined analysis of cpDNA ................................................ 51

Figure 15. Strict consensus phylogeny based on
nuclear GBSSI-1 gene sequences ........................................ 52

Figure 16. Strict consensus phylogram based on
nuclear GBSSI-1 sequence .................................................. 53

Figure 17. Strict consensus phylogeny based on
nuclear GBSSI-1γ sequences ............................................... 54
Determination of phylogenetic relationships among ancestral species of *Rubus* has been elusive. Most *Rubus* species (including blackberries and raspberries), representing nine of the 12 subgenera, occur in a large, well-supported clade (named ‘A’ for reference). The remaining nine species are excluded from this group and represent three subgenera: subg. *Anoplobatus* (*R. bartonianus*, *R. deliciosus*, *R. neomexicanus*, *R. odoratus*, *R. parviflorus*, *R. trilobus*), subg. *Chamaemorus* (*R. chamaemorus*), and subg. *Dalibarda* (*R. lasiococcus*, *R. pedatus*). In addition, *Rubus dalibarda* L. is often treated in its own monotypic genus as *Dalibarda repens* L. Phylogenetic analyses of DNA sequence data from chloroplast regions and the nuclear ribosomal DNA internal transcribed spacer (ITS 1 – 5.8S – ITS 2; ITS) region have not resolved basal relationships in *Rubus* and the position of *Dalibarda repens* has varied from being the sister group to *Rubus* to nested within it. However, monophyly of American subg. *Anoplobatus* species is supported by both genomic regions. Our goal was to clarify ancestral relationships, investigate the position of *Dalibarda repens* relative to *Rubus*, and examine the origin of the circumboreal, octoploid species *R. chamaemorus* using sequence data from one additional chloroplast DNA region, *trnS-trnG*, and the single-copy nuclear gene Granule-Bound Starch Synthase (GBSSI-1). Parsimony analyses of
trnS-trnG sequences indicate a basal trichotomy, while *R.chamaemorus* is strongly supported as sister to *R. pedatus*. A combined cpDNA (*trnS-trnG* and three other regions) parsimony analysis indicates that subg. *Anoplobatus* is sister to clade A, and strongly supports *Dalibarda repens* as sister to *R. lasiococcus*. This suggests that *Dalibarda repens* be classified as *R. dalibarda* consistent with Linnaeus (1762) and Focke (1910). Parsimony analyses of GBSSI-1 sequences result in a large polytomy and do not recover clade A. The presence of three (GBSSI-1α, GBSSI-1β and GBSSI-1γ) putative forms of the gene is observed. However, separate parsimony analysis of GBSSI-1γ sequences demonstrates strong support for clade A and the monophyly of subg. *Anoplobatus*. In this analysis, two different alleles are present in *R. chamaemorus*; one occurs outside clade A (sister to *R. lasiococcus*) and the other nests within clade A (sister to *R. arcticus*). Thus these data suggest that *R. chamaemorus* may be an ancient allopolyploid. The phylogenetic position of *Dalibarda repens* relative to *Rubus* cannot be resolved by existing GBSSI-1 data.
INTRODUCTION

The genus *Rubus* (Rosaceae) is systematically challenging due to hybridization, polyploidy, agamospermy (asexual seed production), and vegetative reproduction. This group, which includes raspberries and blackberries, has approximately 750 species and is found worldwide except in Antarctica. While some species in this genus are used as ornamentals, most are grown for their edible fruits.

Traditional Classification. Focke (1910, 1911, 1914) recognized ~ 429 *Rubus* species in 12 subgenera, the three largest being *Idaeobatus* (raspberries, 117 species), *Malachobatus* (115 primarily Asian species) and *Rubus* (= *Eubatus* Focke; blackberries, ~ 117 species) in the most recent global monograph of *Rubus*. Among the other nine subgenera (*Anoplobatus, Chamaebatus, Chamaemorus, Comaropsis, Cylactis, Dalibarda, Dalibardastrum, Lampobatus and Orobatus*), only three have more than six species.

One taxonomic issue concerns the placement of *Dalibarda repens* whose classification has been fraught with controversy. Linnaeus first described *Dalibarda repens* in 1753; however, he later treated it in *Rubus* as *R. dalibarda* L. (Linnaeus, 1762). Focke (1910) retained this taxon in *Rubus* and created the subgenus *Dalibarda* that comprised *R. dalibarda* and four other species. However, recent botanists (Rydberg, 1913; Bailey, 1941 - 1945; Fernald, 1950; Gleason and Cronquist, 1991) prefer to follow Linnaeus’ (1753) original classification and place *R. dalibarda* in the monotypic genus
Dalibarda. *Rubus* and *Dalibarda* are thought to be closely related and are placed in the tribe Rubeae of subfamily Rosoideae (Robertson, 1974).

**Definition of Basal Species.** In their study of *Rubus*, Alice and Campbell (1999) used nuclear ribosomal DNA internal transcribed spacer (ITS 1 – 5.8S – ITS 2; ITS) sequences to generate the first *Rubus* phylogeny based on a large taxonomic sample. They examined 57 taxa including 20 species of subg. *Rubus* (blackberries), one to seven species from each of the remaining 11 subgenera, and the monotypic and closely related *Dalibarda*. Their results showed that Focke’s (1910,1911, 1914) classification of *Rubus* contains mostly non-monophyletic subgenera although several lineages were strongly supported.

Alice and Campbell (1999) defined the basal species in *Rubus* as those excluded from the strongly supported clade A. The basal species comprise nine *Rubus* species in three subgenera plus *Dalibarda repens*: *R. bartonianus, R. deliciosus, R. neomexicanus, R. odoratus, R. parviflorus* and *R. trilobus* in subg. *Anoplobatus*, *Rubus chamaemorus* in the monotypic subg. *Chamaemorus*, and *R. lasiococcus* and *R. pedatus* in subg. *Dalibarda*.

Their study conflicted with Lu (1983) who suggested that evolution in Rosaceae proceeded from woody to herbaceous and from compound to simple leaves. Alice and Campbell (1999) demonstrated that primarily semi-herbaceous, simple-leaved species occupied basal positions in their trees, which contrasted with Lu’s (1983) hypotheses of
subg. *Idaeobatus* being the most primitive group and the most advanced to be subg. *Chamaemorus*. Alice and Campbell (1999) suggested western North America or far eastern Asia (e.g., Japan or eastern Russia) as the center of origin of *Rubus*. Their observation was in contrast with Lu’s (1983) and Kalkman’s (1988) hypotheses of an origin of *Rubus* in southwestern China, or Gondwanaland, respectively.

**Characteristics of the basal species in Rubus.** A high degree of morphological diversity is evident in *Rubus*. This includes large, woody, upright armed species; delicate, semi-herbaceous, prostrate unarmed species; and climbing species with highly reduced leaf blades (Alice and Campbell, 1999). Slight morphological variations are found among the basal species in *Rubus* though all lack stem armature. The basal species in *Rubus* are found primarily in northern temperate regions except for *R. trilobus* (southern Mexico to Guatemala). Base chromosome number in *Rubus* is $x = 7$, and ploidy ranges from diploid ($2n = 2x = 14$) to dodecaploid ($2n = 12x = 84$). The most frequent ploidy is tetraploid ($2n = 4x = 28$) (Thompson, 1995, 1997).

**Subg. Anoplobatus.** There are six American species in subg. *Anoplobatus*. These are *R. bartonianus* (not sampled in this study), *R. deliciosus* (delicious raspberry; Fig. 1), *R. neomexicanus* (Fig. 2), *R. odoratus* (Fig. 3), *R. parviflorus* (thimbleberry; Fig. 4) and *R. trilobus* (southern Mexico and Guatemala; Fig. 5). All (except *R. bartonianus* and *R. trilobus* for which it is unknown) species in this subg. are diploids.
Flowers in these species are large, showy and usually solitary. They have white or purple petals. All species in this subgenus have simple, digitately lobed leaves. Rydberg (1913) separated subg. *Anoplobatus* into two separate genera in his treatment of North American flora. He classified *R. odoratus* and *R. parviflorus* in the genus *Rubacer*, and *R. deliciosus*, *R. neomexicanus* and *R. trilobus* in the genus *Oreobatus*. The segregation was done on the basis of style morphology.

There are several morphological variations within this group. *Rubus odoratus* and *R. parviflorus* have numerous drupelets that are coherent with each other and separate from the flat receptacle, and are capped with a dry, rather hard, finely and densely pubescent cushion. These two species have glabrous styles. On the other hand, *R. deliciosus*, *R. neomexicanus* and *R. trilobus* are branched prostrate or reclining or erect shrubs. Drupelets are very large, few, fleshy, and are without hairy cushions. These species have pubescent styles (Rydberg, 1913).

*Rubus deliciosus* and *R. neomexicanus* occur in southwestern North America (Figs. 1, 2) and *R. trilobus* is found in Mexico and Guatemala. *Rubus odoratus* is found in eastern North America, ranging from Quebec to Ontario, and south to Tennessee (Fig. 3) and *R. parviflorus* occurs throughout western North America from Alaska to Ontario, and south to California (Fig. 4).

**Subg. Chamaemorus.** *Rubus chamaemorus*, known as cloudberry or baked-apple berry, is the only species in subg. *Chamaemorus*. It is an herbaceous perennial, with
creeping rootstock. The plants are dioecious. It has solitary flowers with ovate sepals and broad, white petals. Its leaves are lobed and drupelets are large. Fruits having pulps are red at first, and become yellow or golden when ripe (Rydberg, 1913). *Rubus chamaemorus* is circumboreal in distribution (Fig. 6). This species is octoploid ($2n = 8x = 56$) and unique among the basal species in being polyploid. Its origin as either an allopolyploid or autopolyploid has not been previously documented.

**Subg. Dalibarda.** Focke (1910) recognized five species in this subgenus. The basal species *R. lasiococcus* (Fig. 7) and *R. pedatus* (Fig. 8) are herbaceous, hermaphrodite perennial plants, with creeping stems. Both species have broad, spreading, and white petals. Leaves in *R. lasiococcus* are ternate or simple, while *R. pedatus* has digitately 5-foliate leaves. The fruits in *R. lasiococcus* are greenish, yellowish, or rarely rose-colored and its drupelets are few, pubescent, and large. *Rubus pedatus* has red, glabrous, pulpy fruits; and its drupelets are few, between 1–6, and distinct (Rydberg, 1913). Both species are diploids and occur in northwestern North America (Figs. 7, 8).

**Dalibarda repens.** Morphologically, the diploid *Dalibarda repens* is distinct from *Rubus* (Fig. 9). It is a low perennial herbaceous plant with creeping, slender stems. Two kinds of flowers, fertile and infertile, are found in this species. Fertile flowers lack petals, and its sepals close around the fruit. Infertile flowers have white, oblong petals, and several but usually abortive carpels. Its leaves are simple and petioled with narrow stipules. Drupelets, usually from 5–10, are large and nearly dry (Rydberg, 1913).
*Dalibarda repens* is found in eastern North America, ranging from Quebec to Ontario, and south to Tennessee (Fig. 9).

**Previous Studies in Rubus Systematics.** Alice and Campbell (1999) in their study of *Rubus* based on ITS data sampled four out the five species described by Focke (1910, 1914) in subg. *Dalibarda*, and their results showed the group to be non-monophyletic. Two of these four species, *R. pedatus* and *R. lasiococcus*, were outside clade A, but did not form a monophyletic group. *Rubus pedatus* was sister to all *Rubus* species sampled plus *Dalibarda*; however the bootstrap (BS) support was less than 50%, and the decay (D) value was only 2. *Rubus lasiococcus* was the second most basal species. ITS data were unable to fully resolve relationships of *R. chamaemorus* apart from placing it outside clade A (BS less than 50%, D1). Moreover, there was absence of nucleotide polymorphism in *R. chamaemorus* ITS sequences. The American subg. *Anoplobatus* species were monophyletic and divided into two subclades. The first subclade strongly placed *R. odoratus* and *R. parviflorus* together (BS 100%, D 9), and the second subclade contained *R. deliciosus* and *R. trilobus* (BS 100%, D 6). Their findings demonstrated that *Dalibarda repens* nested within *Rubus*, although with low support (BS less than 50% and D 2); therefore it could be sister to *Rubus*.

They concluded that in order to resolve phylogenetic relationships within *Rubus*, better sampling of species and either faster evolving or longer nuclear DNA regions were needed. They proposed this need because of the weak support of several nodes in ITS-based trees was due likely to a limited number of characters. Apart from that, rapid
concerted evolution that acts upon the ribosomal gene family may result in distinct nuclear DNA sequences derived biparentally being homogenized (Sang, 2002). Another limitation of ITS is that its sequence variation may be inadequate for the study of closely related species or intraspecific relationships (Baldwin et al., 1995).

Based on the chloroplast DNA (cpDNA) sequence analysis by Dodson and Alice (2004), the positions of *Dalibarda repens* and *R. pedatus* are in conflict with the results obtained from ITS data (Alice and Campbell, 1999). *Dalibarda repens* occurred outside *Rubus* with strong support (BS 98%, D 8). *Rubus pedatus* (BS 87%) was found to be sister to all of the remaining *Rubus* species. The remaining basal species were in agreement with the Alice and Campbell (1999) study.

Although cpDNA sequences are widely used in phylogeny reconstruction, several limitations exist. Notably, the uniparental inheritance, generally maternal, enables cpDNA to reveal only half of the parental lineage. This especially is a disadvantage when putative hybrids are being analyzed (Soltis et al., 1998). Moreover, the slow mutational rate of cpDNA, even in non-coding regions, often presents a significant drawback of these regions for use in phylogeny reconstruction (Small et al. 2004, Shaw et al. 2005). These utility constraints result in the need for more data to produce a robust phylogeny.

**Objectives.** The main goal of this research was to enhance the understanding of evolutionary relationships among the basal *Rubus* species and provide insight into
speciation via polyploidization by using DNA sequence data to: 1) resolve basal relationships within *Rubus*, 2) investigate whether the taxon *Dalibarda repens* is sister to or nested within *Rubus*, and 3) examine the origin of 8x *R. chamaemorus.*

This knowledge is needed because a robust phylogeny of the basal species in *Rubus* is essential to assess changes in morphological traits and elucidate the historical biogeography.

**DNA Sequences.** Use of molecular data has had a tremendous effect on the field of plant systematics, and DNA sequence data are now commonly used to gain insight into phylogenetic questions (Soltis et al., 1998). Molecular data are considered preferable to morphology primarily because of the greater number of characters available, and because the molecular characters studied are assumed to be selectively neutral. However, in order to overcome the incongruence between gene trees and the underlying organismal trees, phylogenetic hypotheses from multiple unlinked genes are needed (Hillis, 1991). DNA sequences are highly useful because they provide the ultimate source of genetic variation. The rapid development in the DNA sequencing field has also enabled data to be obtained with relative ease and reduced costs. Most importantly, this method allows for usage at all levels of the evolutionary hierarchy. In comparison with DNA sequencing, other markers such as microsatellites are better suited for population level studies and are too variable for comparisons of divergent species. Gene duplications that result in complex gene products are difficult to interpret in allozymes. SCARs and RAPDs may pose complications in phylogenetic reconstruction as these may not provide independent loci.
Markers routinely used in plant molecular systematics can be assigned into three categories: nuclear genome, chloroplast genome and mitochondrial genome (mtDNA). The general slow rate of sequence evolution, fast rate of structural evolution, and the common presence of small, unstable extra chromosomal plasmids of unknown origin in plant mitochondrial genomes hinder its usage as a potential source of data (Soltis et. al., 1998). Although plant molecular systematists have relied primarily on cpDNA throughout history, investigators have come to realize the potential of nuclear gene sequences in phylogenetic comparisons (Soltis et al., 1998).

There is high variability among plant nuclear genomes. Even between closely related species, chromosome number, the degree of gene clustering, and chromosome size can all differ tremendously. These allow plant nuclear genomes to contain sufficient variability to be powerful differentiating factors (Soltis et al., 1998). In addition, nuclear genes have a significant advantage in phylogenetic analysis because of their faster evolutionary rates relative to genes found in plastids. Previous studies have documented that synonymous substitution rates of nuclear genes were found to be five times greater than that of chloroplast genes and up to 20 times higher compared to mitochondrial genes (Soltis et al., 1998 and references therein).

Aside from having the main advantage of an elevated rate of sequence variation, nuclear genes are important because they represent multiple unlinked loci that are useful for inferring independent phylogenies. Confidence level in a phylogeny increases when it is corroborated by independent datasets. Another desirable trait possessed by nuclear
genes is their biparental Mendelian inheritance. In cases of hybridization or allopolyplloidization, nuclear genes, especially low-copy ones, are the favored markers to help identify parental lineages.

Work on the chloroplast genome has dominated plant molecular evolutionary studies. Most importantly there is an abundance of cpDNA present in total cellular DNA. Chloroplast DNA also has a conservative rate of nucleotide substitution; this is seen as ideal for studying plant phylogenetic relationships at or above the family level. Finally, cpDNA data is able to provide an important tool for the reconstruction of plant relationships in the controversial deeper levels of plants evolution because of the abundance of extensively researched cpDNA encoded-genes. Taken in combination, these factors have contributed to cpDNA playing an essential role in plant molecular systematics (Soltis et al., 1998 and references therein).

Scope of the Current Study. Chloroplast DNA sequences and nuclear DNA sequences were generated in this study from the trnS-trnG intergenic spacer and trnG intron from the chloroplast genome and a low-copy number nuclear gene GBSSI.

The cpDNA trnS-trnG regions consist of the trnG intron and the trnS-trnG intergenic spacer. This region was initially used by Hamilton (1999) to study population dynamics in a tropical tree species of Corythophora (Lecythidaceae) (Shaw et al., 2005 and references therein). The trnS-trnG intergenic spacer has been shown to be among the most informative of nine non-coding cpDNA regions within two closely related
subgenera of *Glycine* (Shaw et al., 2005 and references therein). The *trnS-trnG* spacer is reported to contain many more potentially informative characters (PICs) than 3 - 5 other cpDNA regions. This region is also five times more variable than the *trnL-trnF* spacer and contained more variable characters than ITS in *Tamarix* (Shaw et al., 2005 and references therein).

The *trnG* intron, by itself, provided nearly double the number of variable characters when compared to *trnL-trnF*, although it did not contain as many as the *rpl16* intron (Shaw et al., 2005 and references therein). Shaw et al. (2005) studied the phylogenetic utility of 21 non-coding cpDNA regions across angiosperms and included *Prunus*, a member of Rosaceae. They categorized regions into three tiers based on their usefulness in phylogeny reconstruction. Based on their study, the *trnS-trnG* spacer was placed in Tier 1 and the *trnG* intron in Tier 2. They reported that the *trnS-trnG* spacer gives both the highest number of potentially informative characters (PICs) and variability averaging 4.74%. They also noted that when the *trnS-trnG* intergenic spacer and *trnG* intron are combined as one co-amplifiable unit, averaging ~ 1500 bp, the highest number of PICs per two (very rarely three) sequence reactions was obtained. These results were in comparison with all other non-coding cpDNA regions, single or combined, surveyed by them.

The GBSSI codes for granule-bound starch synthase and is single copy in most diploid angiosperms (Mason-Gamer, 1996). In *Rubus* and most Rosaceae, GBSSI is represented by two paralogous loci, GBSSI-1 and GBSSI-2 (Evans et al., 2000).
Previous phylogenetic studies have shown that GBSSI exons and introns were useful in resolving relationships among closely related genera and species (Alice, 1997, Evans et al., 2000 and references therein). This gene has been very useful in detecting ancient hybridizations as demonstrated by Mason-Gamer (1996) and Evans et al. (2000). In a study done by Peralta et al. (1997) to examine phylogenetic relationships of nine wild tomato species, GBSSI data provided useful information about intraspecific variation and evolution of characters regarding breeding system and fruit color.

There are several advantages of using GBSSI as a molecular marker that were documented by Mason-Gamer et al. (1996) in their study of the phylogenetic utility of the gene in grasses. First, GBSSI has a high rate of nuclear gene intron variability, which makes it a desirable marker to examine relationships between closely related species. Secondly, it was found that the GBSSI exons are also able to resolve relationships among more distantly related taxa and reconstruct phylogenetic trees that are consistent with those produced from other genes. Finally, the number of characters using the introns of this gene is potentially greater because of the presence of 12 introns, each varying from 100 - 150 bp in length. These properties make GBSSI a molecular marker with the potential to be used in a wide range of taxonomic studies.

MATERIALS AND METHODS

**Plant samples.** Single accessions from 29 *Rubus* species representing all 12 subgenera were sampled along with *Dalibarda repens* and *Rosa multiflora* as outgroup
(Table 1). Total genomic DNA was extracted and sequenced for ITS by Alice and Campbell (1999), cpDNA sequences (Dodson and Alice, 2004; Alice and Dodson, in prep.) and GBSSI-1 for clade A species (Alice, unpubl. data).

**Polymerase Chain Reaction (PCR).** Two new DNA regions were examined: (1) the chloroplast DNA $trnS^{GCU}$-$trnG^{UUC}$ that comprises the $trnS$-$trnG$ intergenic spacer region and the $trnG^{UUC}$ intron, and (2) the granule-bound starch synthase gene (GBSSI-1) comprising four complete exons, two partial exons and five introns.

$trnS$-$trnG$. PCR amplification generally followed the suggestion of Shaw et al. (2005). Primers $trnS^{GCU}$ (AGA TAG GGA TTC GAA CCC TCG GT) and 3′$trnG^{UUC}$ (GTA GCG GGA ATC GAA CCC GCA TC) were used to directly amplify double-stranded DNA by symmetric PCR. Reaction volumes were 25 μL and contained 12.5 μL FailSafe™ 2X PCR Premix C (Epicentre Biotechnologies, Madison, WI), 0.25 μL Taq DNA Polymerase (5 μ/μL), 1.0 μL of each primer and 10.25 μL genomic DNA serial diluted to 1:100. PCR was performed in a PTC-100 thermal cycler and consisted of 40 cycles of 1 min at 94°C for template denaturation, 1 min at 48°C for primer annealing, and 2 min at 72°C for primer extension, followed by a final extension of 10 min at 72°C. PCR products were purified by gel electrophoresis in 0.8% agarose followed by band isolation. For direct sequencing of cpDNA $trnS$-$trnG$, gel slices containing the PCR product were purified using QIAquick gel extraction columns (QIAGen Inc., Valencia, CA) following the manufacturer’s instructions.
**GBSSI-1.** PCR amplification of GBSSI using primers GBSSI-3F (TAC AAA CGA GGG GTT GAT CG) and GBSSI-8R (GAT TCC AGC TTT CAT CCA GT) was done by symmetric PCR following a modification of Alice (1997). Reaction volumes were 25 μL and contained 1.75 μL nanopure water, 2.5 μL 100X purified bovine somatrophin albumin (BSA) (New England Biolabs, Beverly, MA.), 2.5 μL Taq DNA Polymerase 10X Buffer with MgCl₂ (Promega, Madison, WI), 1.0 μL dNTPs (5 μM each) (Invitrogen Inc., Carlsbad, CA), 1.0 μL (10μM) of each primer (QIAgen Inc., Valencia, CA), 0.25 μL Taq DNA Polymerase (5 μ/μL) (Promega, Madison, WI), and 15.0 μL genomic DNA serially diluted to 1:100. PCR was performed in a PTC-100 thermal cycler (MJ-Research, Inc., Watertown, MA) and consisted of 10 cycles of 45 sec at 94°C for template denaturation, 2 min at 65°C for primer annealing, and 1 min 30 sec at 72°C for primer extension, followed by 35 cycles of 30 sec at 94°C for template denaturation, 1 min 20 sec at 65°C for primer annealing, and 2 min at 72°C for primer extension, and a final extension of 20 min at 72°C. PCR products were visualized by gel electrophoresis in 0.8% OmniPur agarose (Merck KGaA, Darstadt, Germany) followed by band isolation.

**Cloning.** The GBSSI-1 region was cloned for six species. Fresh PCR products were ligated into the pCR4.0-TOPO cloning vector in the TOPO TA-cloning Kit (Invitrogen Inc., Carlsbad, CA) following the manufacturer’s instructions. The resulting ligation reaction was used to transform competent *E. coli* cells (One Shot® Chemically Competent) provided in the kit. The transformation mix was incubated in 250 μL LB medium for 1 hour at 37°C on a rotary shaker then plated on LB agar with 50 μg/mL
kanamycin. Plates were incubated overnight at 37°C. Positive selection of clones was
done after 18 hours. Eight colonies per taxon were selected for growth in LB-kanamycin
broth overnight at 37°C. Plasmid DNA was isolated using a QIAprep Spin MiniPrep Kit
(QIAgen Inc., Valencia, CA) following the manufacturer’s instructions. Potentially
recombinant plasmids were digested with EcoRI in order to determine if they contained
the GBSSI insert. Reaction volumes were 10 µL and contained 1X EcoRI buffer, 10
units enzyme (New England Biolabs, Beverly, MA) and 2.0 µL plasmid DNA. Digested
DNA was electrophoresed in 0.8% agarose gels, and visualized and photographed under
UV light.

**Sequencing PCR Products and Clones.** Double-stranded DNA was sequenced
using the dideoxy chain termination method using an ABI PRISM BigDye Terminator
v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Removal of
unincorporated dye terminators from the sequencing reactions was done using a DyeEx
2.0 Spin Kit (QIAgen Inc., Valencia, CA) following the manufacturers’ instructions.
Samples were electrophoresed in an ABI3130 Genetic Analyzer (Applied Biosystems,
Foster City, CA). Sequences were edited and aligned visually in Sequencher version 4.6
(Gene Codes Corporation, Ann Arbor, MI).

**Phylogenetic analysis.** *Rosa multiflora* was selected as outgroup based on the
recent analysis of GBSSI sequences in Rosoideae (Eriksson et al., 2005). Four data sets
were analyzed: (1) ITS, (2) trnS-trnG, and (3) combined cpDNA with 31 total taxa, and
(4) GBSSI-1 for 21 taxa. Parsimony analysis was done using HEURISTIC searches
comprising 1000 reps of RANDOM stepwise-addition of taxa. Sets of equally parsimonious trees were summarized using strict consensus. Decay indices (Bremer, 1988; Donoghue et al., 1992) and bootstrap analyses (Felsenstein, 1985) with 500 replicates were calculated as measures of support for individual clades. Decay analyses were performed with AutoDecay (Eriksson and Wikstrom, 1996).

RESULTS

Sequence Characteristics.

*trnS-trnG* intergenic spacer. The total number of aligned characters is 740 (Table 2). This region contains 116 (15.68%) variable characters, and 46 (6.22%) are parsimony-informative characters. Mean pairwise divergence ranges from 0.0% (*R. tricolor - R. assamensis*, and *R. tricolor - R. neomexicanus*) to 5.1% (*R. crataegifolius - Dalibarda repens*).

*trnG* intron. The total number of aligned characters is 751 (Table 2). This region contains 76 (10.12 %) variable characters, and 35 (4.66%) parsimony-informative characters. Mean pairwise divergence ranges from 0.0% (*R. parviflorus - R. odoratus*) to 3.2% (*R. rosifolius - R. glabrus*).

GBSSI-1. Total number of aligned characters for the region is 1621. This region contains 477 (29.43%) variable characters, and 199 (12.28%) are parsimony-informative characters.
Phylogenetic Relationships of Rubus.

ITS region (Alice and Campbell, 1999). Parsimony analysis of the ITS regions yields 102 equally parsimonious trees of length 298 (strict consensus in Fig. 12). Excluding uninformative sites, the consistency index (CI) is 0.571, and the retention index (RI) is 0.759. The phylogeny shows a basal trichotomy composed of Dalibarda repens, the remaining basal species, and clade A. Moreover, the relationships involving the basal species are low. Subgenus Anoplobatus is monophyletic with bootstrap (BS) support of 74% and decay value (D) of 3, and divided into two subclades. Rubus deliciosus, R. neomexicanus and R. trilobus are nested together in a well-supported clade (BS 99%, D 4) and the other clade groups R. odoratus and R. parviflorus with strong support (BS 100%, D 9). Rubus chamaemorus together with R. lasiococcus and R. pedatus form a weakly supported (BS < 50%, D 1). Clade A is recovered with strong support (BS 96%, D 8).

trnS - trnG region. Analysis of the trnS-trnG regions yields 2559 equally parsimonious trees of length 291 (strict consensus in Fig. 13). Excluding uninformative sites, the CI is 0.607, and the RI is 0.695. The base of the phylogeny shows Dalibarda repens, R. lasiococcus, and the remaining species as three unresolved lineages. The next node comprises five lineages: (1) R. chamaemorus and R. pedatus (BS 95%), (2) American subg. Anoplobatus (BS 75%), (3) most members of clade A (BS 65%), (4) R. corchorifolius, R. crataegifolius, and R. rosifolius (BS 80%), and (5) R. idaeus. Clade A is not recovered. Rubus deliciosus, R. neomexicanus, and R. trilobus are nested together in a clade (BS 75%) and R. deliciosus sister to the two. There is strong support (BS 85%)
for *R. neomexicanus* and *R. trilobus* as sister species. The other clade comprises *R. odoratus* and *R. parviflorus* (BS 82%). Clade A is not recovered.

Combined cpDNA (Dodson and Alice, 2004; Alice and Dodson, in prep.). Analysis of all four cpDNA regions recovered eight equally parsimonious trees of length 1141 (strict consensus in Fig. 14). Excluding uninformative sites, the CI is 0.610 and the RI is 0.741. The strict consensus phylogeny shows the basal species in *Rubus*, including *Dalibarda repens*, in three strongly supported clades occurring as a paraphyletic grade.

*Dalibarda repens* and *R. lasiococcus* form the basal group (BS 98%, D 6) followed by *R. chamaemorus* and *R. pedatus* (BS 99%, D 6), then American subg. *Anoplobatus* sister to clade A (BS 98%, D 3). *Rubus deliciosus, R. neomexicanus*, and *R. trilobus* are together in a well-supported clade (BS 100%, D 5) and *R. deliciosus* sister to the two. The other clade comprises *R. odoratus* and *R. parviflorus* (BS 100%, D 9). Clade A is also recovered with strong support (BS 100%, D 16). There is strong support (BS 98%, D 3) for subg. *Anoplobatus* to be sister to clade A.

GBSSI-1. Parsimony analysis of the GBSSI-1 gene yields 777 equally parsimonious trees of length 860 (strict consensus in Fig. 15). Excluding informative sites, the CI is 0.653, and the RI is 0.797. Clade A is not recovered in this analysis. Cloned sequences from individual accessions do not group together and appear to show multiple forms (labeled GBSSI-1α, β, and γ). There is strong support (BS 100%) for the GBSSI-1α clade though there is no resolution. Clade GBSSI-1β has reasonable support (BS 79%), but also lacks resolution. The GBSSI-1γ clade, containing clones of American
subg. *Anoplobatus* species, is well supported (BS 71%) and has the highest resolution of the three groups even the species included are more closely related. An unlabelled fourth group exists composed of one clone each of *Dalibarda repens* and *R. deliciosus*. Notably, these two cloned sequences have a diagnostic 2-bp deletion.

Two clones of *Dalibarda repens* occur in the GBSSI-1α clade and four in GBSSI-1β. None of the clones occur in the GBSSI-1γ clade. Two clones of *R. deliciosus* occur in GBSSI-1α clade, and three in clade GBSSI-1γ. One *R. neomexicanus* clone occurs in clade GBSSI-1α, while six in the GBSSI-1γ clade. None occur in the GBSSI-1β clade. There are three clones of *R. pedatus* in clades GBSSI-1α, and GBSSI-1β, respectively. None occur in clade GBSSI-1γ. None of the *R. parviflorus* clones are nested in clade GBSSI-1α. There are two clones nested in clade GBSSI-1β, and three in the GBSSI-1γ clade. One clone of *R. chamaemorus* nests in clade GBSSI-1β. Two other clones occur in a single clade, excluded from the GBSSI-1α, GBSSI-1β and GBSSI-1γ clades, and are sister species to *R. arcticus*. This clade is weakly supported (BS 54%). The only *R. odoratus* clone occurs in clade GBSSI-1γ, and is sister to three clones of *R. parviflorus*. This relationship is strongly supported (BS 87%). The single *R. lasiococcus* clone is found in a different clade by itself. For each basal species, a minimum of two different alleles was detected except for *R. lasiococcus* and *R. odoratus*. Three different alleles were detected for *Dalibarda repens* and *R. deliciosus*. Two different alleles were detected for *R. neomexicanus, R. pedatus, R. parviflorus* and *R. chamaemorus* (Table 3).
Lack of sequence divergence among clones of different species individuals and low resolution within GBSSI-1α and β was detected (Fig. 16). No GBSSI-1γ sequences were found for species not included in the American subg. *Anoplobatus*. When GBSSI-1γ DNA sequences are analyzed separately, clade A is resolved with strong support (BS 83%) (Fig. 17). Consistent with cpDNA and ITS DNA results, subg. *Anoplobatus* is further divided into two subclades. The first subclade comprises *R. deliciosus* and *R. neomexicanus* clones (BS 99%) and the second subclade groups clones of *R. odoratus* and *R. parviflorus* (BS 100%). *Rubus odoratus* is shown to be sister species to three *R. parviflorus* clones (BS 60%).

Two alleles of *R. chamaemorus* are detected. One *R. chamaemorus* clone is sister to *R. lasiococcus* (BS 62%). Two other *R. chamaemorus* clones occur inside clade A. These clones are sister to *R. arcticus* in a well-supported clade (75%).

**DISCUSSION**

*Phylogeny of Rubus.* Lack of resolution in the current ITS tree resulting in a basal trichotomy provides no additional insight as regards the placement of *Dalibarda repens* (Fig. 12). However, analysis of the Alice and Campbell (1999) data with a reduced taxonomic sample does change the position of *Dalibarda repens*. In our phylogeny, *Dalibarda repens* occurs as an unresolved lineage at the base; whereas Alice and Campbell (Figs. 2,3; 1999) showed *Dalibarda repens* as sister to clade A albeit with low
support. This variation in position using the same ITS data highlights the instability of
the basal nodes and indicates the need for alternatives nuclear sequences.

GBSSI-1 data do not resolve the relationship of *Dalibarda repens* relative to
*Rubus* (Fig.15). Six out of seven clones of this species that were obtained occur only in
GBSSI-1α and GBSSI-1β clades. Even then, these clones show low sequence
divergence with the clones of other species that were found in each clade (Fig.16). There
are no *Dalibarda repens* clones recovered in the GBSSI-1γ tree, and this hampers our
effort in gaining more evidence to resolve the position of *Dalibarda repens* (Fig.15 and
16).

The *trnS-trnG* phylogeny is similar to ITS in that the base is unresolved (Fig. 13).
When *trnS-trnG* data are combined with other cpDNA data, a very robust phylogeny is
obtained (Fig. 14). Combined analysis of 6461 cpDNA characters strongly supports
*Dalibarda repens* and *R. lasiococcus* as sister to the remaining species. Our combined
chloroplast phylogeny does not support the classification of *Dalibarda repens* as a
monotypic genus.

The occurrence of *R. chamaemorus* as sister to *R. lasiococcus* and *R. pedatus*
based on current ITS tree (Fig. 12) is consistent with Alice and Campbell (1999). In ITS
(Fig. 12) and combined cpDNA trees (Figs.13 and 14), there is a close relationship
between *R. pedatus* and *R. chamaemorus* which is supported by shared morphology (Fig.
6 and 8). Both are herbaceous perennial plants with red pulpy fruits, and share a northern
geographic distribution. The GBSSI-1 tree did not indicate the presence of the GBSSI-1γ form of the gene in *R. pedatus* and three clones occur in each of the α and β clades, respectively (Fig. 15). However the α and β clades in the GBSSI-1 tree show low divergence among clones, resulting in lack of information pertaining to the position of *R. pedatus* (Fig. 16). The inability to recover clones having the γ copy of the GBSSI-1 gene hampers the effort to fully characterize the relationship of *R. pedatus* relative to others in this tree. However, *R. chamaemorus* and *R. pedatus* are sister species to American subg. *Anoplobatus* and clade A (Fig. 15).

American species of subgenus *Anoplobatus* consistently form a monophyletic group regardless of data. Current ITS and cpDNA phylogenies strongly reinforce the position of these five taxa as being several of the basal species in *Rubus* in addition to cpDNA tree providing insight that subg. *Anoplobatus* is sister clade to clade A. The absence of *R. trilobus* clones sequences did not cause any loss of resolution in the GBSSI-1γ as the remaining four species positions are congruent with other data. There is no conflict regarding the monophyly of this group is all phylogenies except in GBSSI-1 (Fig.15). Monophyly is further supported by the overall similar morphology shared by these species. The division of this clade into two subclades is also seen in all trees except in GBSSI-1 (Fig.15). Moreover, two subclades corresponding to Rydberg’s (1903, 1913) genera *Rubacer* and *Oreobatus* can be distinguished. However all data confirm the placement of these species in *Rubus*. We further suggest that only American species be classified in subg. *Anoplobatus* excluding the two Asian species, *R. ribisoideus* and *R. trifidus*. Geographically, these American species occur in similar places in the United
States, with the exception of *R. trilobus*, a native of Mexico. However, the close proximity of Mexico with southwestern US may have facilitated the gene flow between these two regions.

Based on these findings, two alternatives may be considered in re-classifying basal species in *Rubus* and *Dalibarda repens*, ranging from conservative to radical suggestions. Because the basal species, including *Dalibarda repens*, occur in three strongly supported clades (Fig. 14), one could recognize three genera that are separate from *Rubus*. The alternative would be to treat *Dalibarda repens* as a species of *Rubus* rendering *Rubus* monophyletic. The primary reasons for treating *Dalibarda repens* as separate from *Rubus* is based on its dry fruits as opposed to pulpy fruits and the presence of fertile and infertile flowers. This taxonomy to treat *Dalibarda repens* as *R. dalibarda* L. has been proposed by Linnaeus (1762) and followed by Focke (1910). We support this classification and propose that *Dalibarda repens* be put back into *Rubus*.

**Putative Allopolyploid Origin of Octoploid *R. chamaemorus***. The process of becoming and the attributes of being polyploid play a major role in the development and maintenance of genetic variation in allopolyploid species. Allopolyploidy may arise following hybridization and chromosome doubling between two genetically distinct diploid species, and it is an important mechanism of speciation in flowering plants (Soltis and Soltis, 1999 and references therein). The duplication of chromosomes giving rise to an octoploid results in “instant” speciation and may confer complete reproductive isolation from the parental diploid taxa.
The position of *R. chamaemorus* occurs outside clade A in current ITS and cpDNA trees. Based on the ITS tree, *R. chamaemorus* is sister to *R. lasiococcus* and *R. pedatus* (Fig. 12). A robust cpDNA phylogeny shows that *R. pedatus* is sister to *R. chamaemorus* (Fig. 14). The GBSSI-1γ phylogeny indicates the presence of two divergent alleles in 8x *R. chamaemorus* (Fig. 16). One of these alleles is detected in clone 6 that occurs among basal *Rubus* species outside clade A and is sister to *R. lasiococcus*. The second allele is present in clones 2 and 4, which are sister to *R. arcticus* inside clade A. We infer *R. chamaemorus* to be an ancient allopolyploid because of the lack of persistent nucleotide site polymorphism in ITS (Alice and Campbell, 1999). Based on combined cpDNA and GBSSI-1 phylogenies, *R. chamaemorus* may be an ancient allopolyploid involving *R. pedatus* and *R. lasiococcus* as its putative parental species.

A minimum number of hybridization events may be estimated from the maximum total number of alleles derived from the diploid parental species at any given locus (Wendel, 2000). For example, if each of the diploid species contributed two different alleles to an allotetraploid, then a single hybridization event could result in an allotetraploid with four different alleles at any given locus. Since only two alleles were recovered in this study, it could not be ascertained whether the hybridization event resulting in *R. chamamemorus* was single or multiple.
However, previous molecular data indicate that multiple origins of polyploids are the rule rather than the exception (Soltis and Soltis, 1999). In addition, recurrent hybridization events have been documented occurring over relatively short time spans and geographic distances (Soltis and Soltis, 1999 and references therein). Thus, it is likely that the origin of allooctoploidy in *R. chamaemorus* may have resulted from multiple hybridization events rather than a single one. It is interesting to note that a wider range in geographic distribution of polyploids relative to their diploid progenitors has been reported (Soltis and Soltis, 1999 and references therein). Further studies incorporating *R. chamaemorus* sequences from Alaskan and northern European samples would provide further insight into multiple hybridization events and dispersal patterns in this species.

Investigation into the origin of allopolyploids in *Rubus* becomes extremely interesting when considering previously documented evidence that the genomes of some plants are the products of polyploidy, but had later become diploidized, known sometimes as “degenerate” polyploids (Soltis and Soltis, 1999, and references therein). These occurrences, although uncommon, raise a multitude of questions about the evolution of all species in *Rubus*, with respect to whether similar “diploidization” events have occurred in this genus and the evolutionary stage of *R. chamaemorus*. It is noted that extensive and rapid genome restructuring can occur after polyploidization. Such changes can be mediated by transposable elements. Polyploidization is believed to represent a period of transilience, a period during which the genome is more amenable to,
or tolerant of, change, such as recombination, potentially producing new gene complexes and driving rapid evolution (Soltis and Soltis, 1999).

**Molecular Evolution of Duplicated Genes.** Duplication of individual genes, chromosomal segments, or entire genome have long been considered a primary force in genomic evolution, contributing material for the origin of evolutionary novelties, including new gene functions and expression patterns (Lynch and Conery, 2000 and references therein). Gene duplication creates a requirement for distinguishing orthologous from paralogous genes with respect to phylogeny reconstruction (Wendel et al., 2000 and references therein). It is acknowledged that inadvertent inclusion of paralogous sequences in a phylogenetic analysis could result in a well supported but misleading topology (Adams and Wendel, 2005).

The situation becomes more complicated if the duplicated genes are subject to interlocus concerted evolution. If concerted evolution is strong, then duplicated genes will be homogenized to a single type; if concerted evolution is absent, then duplicated genes will maintain their identity and construction of an orthology–paralogy tree is possible (Adams and Wendel, 2005). Two copies of the GBSSI-1 gene were found to show low levels of sequence divergence amongst species in this study (Fig. 16). This finding could be the result of weak concerted evolution acting upon the gene since if concerted evolution is present but weak, then some copies (or parts) of duplicated genes may be homogenized, whereas others are maintained as distinct. This occurrence may
possibly provide an explanation for the nature of GBSSI-1γ, which maintains the highest level of sequence divergence amongst species.

It is widely accepted that accurate reconstruction of a species tree from duplicated genes is fraught with difficulties (Wendel, 2000 and references therein). Moreover, it is unclear how duplicate genes evolve from an initial state of complete redundancy, where gene copies are likely to be expendable, to a stable situation in which both or more copies are maintained by natural selection. The frequency of these events is also unclear (Zhang, 2003).

Theoretically, there are three possible outcomes in the evolution of duplicate genes: (i) one copy may simply become silenced by degenerative mutations or non-functionalization; (ii) one copy may acquire a novel, advantageous function and become preserved via natural selection, with the other copy retaining the original function (neofunctionalization); or (iii) all copies may become partially compromised by mutation accumulation to the point at which their total capacity is reduced to the level of the single-copy ancestral gene (subfunctionalization). Since the vast majority of mutations affecting fitness are deleterious, and because gene duplicates are generally assumed to be functionally redundant at the time of origin, almost all models predict that the usual fate of a duplicate-gene pair is the non-functionalization of one copy (Lynch and Conery, 2000 and references therein).
However, recent studies have found these alternatives are only partially consistent. Lynch and Conery (2000) found that a conservative estimate of the average rate of origin of new eukaryotic gene duplicates is on the order of 0.01 per gene per million years, with rates in different species ranging from about 0.02 down to 0.002. Given this range, 50% of all of the genes in a genome are expected to duplicate and increase to high frequency at least once on time scales of 35 to 350 million years. Thus, even in the absence of direct amplification of entire genomes (polyploidization), gene duplication has the potential to generate substantial molecular material for the origin of evolutionary novelties. The rate of duplication of a gene is of the same order of magnitude as the rate of mutation per nucleotide site, and most duplicated genes experience a brief period of relaxed selection early in their history, with a moderate fraction of them evolving in an effectively neutral manner during this period (Lynch and Conery, 2000).

Interestingly, the vast majority of gene duplicates are silenced (rather than preserved) within a few million years, with the few survivors subsequently experiencing strong purifying selection. Although duplicate genes may only rarely evolve new functions, the stochastic silencing of such genes may play a significant role in the passive origin of new species. This phenomenon is noteworthy especially since the primary evolutionary dynamic in *Rubus* is characterized by polyploidy and hybridization. If concerted evolution were acting to homogenize sequences, little to no intraspecific polymorphism would be expected to be seen. This expectation is consistent with our results, which show low intraspecific divergence in GBSSI-1α and GBSSI-1β genes.
(Fig.16). It is important to consider the issue of incomplete sampling with respect to alleles. It may be that the allelic phylogeny obtained shows a combination of processes, including presence of concerted evolution or recombination, gene flow, and non-coalescence.

The findings here contribute further to the widely accepted notion that pervasive gene duplication is an extremely common occurrence in plant low-copy genes throughout the evolutionary history of angiosperms (Duarte et al., 2006). With respect to GBSSI, it is slowly emerging that the events described here are not isolated since a recent study using this gene in the invasive allopolyploid *Spartina anglica* resulted in similar findings (Ainouche, 2006). As a measure of caution, Southern hybridization analyses may need to be included in phylogenetic studies in order to assess the copy number especially when using nuclear genes. This additional step may provide insight into the presence of heterozygosity and/or multiple loci, which may assist in providing more information about instances of gene duplication events that may seem spurious at first.

In conclusion, our data suggest that *Dalibarda repens* be re-classified as *Rubus dalibarda* L. following Linnaeus (1762) and Focke (1910). Although ITS and cpDNA data place 8x *R. chamaemorus* outside clade A, GBSSI-1 indicate an allopolyploid origin as divergent alleles occur inside clade A and outside clade A. Combined chloroplast DNA shows *R. chamaemorus* and *R. pedatus* as sister species implying that *R. pedatus* represents its maternal ancestors. Finally, the presumed single copy nuclear gene GBSSI-1 apparently has multiple copies. Of these, only one form, GBSSI-1γ, shows levels of divergence among species consistent with other phylogenetic hypotheses.
LITERATURE CITED


Dodson, T.M. and L.A. Alice. 2004. Phylogeny of Rubus based on cpDNA sequences. Abstracts of the Kentucky Academy of Science Scientific Meeting, Murray State University, Murray, KY.


National Center for Biotechnology Information


United States Department of Agriculture Plants Database
http://plants.usda.gov


Table 1: Basal species in *Rubus* and outgroup accessions used in this study. Subgeneric classifications follow Focke (1910, 1911, 1914). Geographic origin is by country, except from the USA for which two-letter state abbreviations are used. Accession number is for DNA sequences in GenBank (the prefix GBAN has been added for linking the on-line version of American Journal of Botany to GenBank and is not part of the actual GenBank accession number).

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Origin</th>
<th>P.I. Number</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subg. Anoplobatus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. <em>R. deliciosus</em> Torr.</td>
<td>OK - USA</td>
<td>N/A</td>
<td>GBANAF055733</td>
</tr>
<tr>
<td>2. <em>R. odoratus</em> L.</td>
<td>ME - USA</td>
<td>N/A</td>
<td>GBANAF055734</td>
</tr>
<tr>
<td>3. <em>R. parviflorus</em> Nutt.</td>
<td>MI - USA</td>
<td>PI 553785</td>
<td>GBANAF055735</td>
</tr>
<tr>
<td>4. <em>R. trilobus</em> Moc. et Sesse.</td>
<td>Mexico</td>
<td>N/A</td>
<td>GBANAF055738</td>
</tr>
<tr>
<td>5. <em>R. neomexicanus</em> A. Gray</td>
<td>USA</td>
<td>PI 553719</td>
<td>GBANAF83358</td>
</tr>
<tr>
<td><strong>Subg. Chamaemorus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. <em>R. chamaemorus</em> L.</td>
<td>ME-USA</td>
<td>N/A</td>
<td>GBANAF055740</td>
</tr>
<tr>
<td><strong>Subg. Dalibarda</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. <em>R. lasiococcus</em> A. Gray</td>
<td>OR- USA</td>
<td>PI 618518</td>
<td>GBANAF055750</td>
</tr>
<tr>
<td>8. <em>R. pedatus</em> Sm.</td>
<td>Canada</td>
<td>N/A</td>
<td>GBANAF055751</td>
</tr>
<tr>
<td>9. <em>Dalibarda repens</em> L. (= <em>R. dalibarda</em> L.)</td>
<td>ME - USA</td>
<td>N/A</td>
<td>GBANAF055748</td>
</tr>
<tr>
<td><strong>Subg. Chamaebatus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Species</td>
<td>Location</td>
<td>Accession</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>10.</td>
<td><em>R. nivalis</em> Douglas</td>
<td>OR- USA</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Subg. Comaropsis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td><em>R. geoides</em> Sm.</td>
<td>Argentina</td>
<td>CRUB 1981</td>
</tr>
<tr>
<td><strong>Subg. Cylactis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td><em>R. arcticus</em> L.</td>
<td>Sweden</td>
<td>PI 553831</td>
</tr>
<tr>
<td>13.</td>
<td><em>R. pubescens</em> Raf.</td>
<td>ME - USA</td>
<td>N/A</td>
</tr>
<tr>
<td>14.</td>
<td><em>R. saxatalis</em> L.</td>
<td>Sweden</td>
<td>PI 370130</td>
</tr>
<tr>
<td><strong>Subg. Dalibardastrum</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td><em>R. tricolor</em> Focke</td>
<td>China</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Subg. Idaebatus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td><em>R. crataegifolius</em> Bunge</td>
<td>Korea</td>
<td>PI 553173</td>
</tr>
<tr>
<td>17.</td>
<td><em>R. corchorifolius</em> L.</td>
<td>China</td>
<td>PI 618400</td>
</tr>
<tr>
<td>18.</td>
<td><em>R. idaeus</em> L.</td>
<td>ME - USA</td>
<td>N/A</td>
</tr>
<tr>
<td>19.</td>
<td><em>R. occidentalis</em> L.</td>
<td>ME - USA</td>
<td>N/A</td>
</tr>
<tr>
<td>20.</td>
<td><em>R. phoenicolasius</em> Maxim</td>
<td>DC - USA</td>
<td>N/A</td>
</tr>
<tr>
<td>21.</td>
<td><em>R. lasiostylus</em> Focke</td>
<td>China</td>
<td>PI 553668</td>
</tr>
<tr>
<td>22.</td>
<td><em>R. rosifolius</em> Sm.</td>
<td>Seychelles</td>
<td>N/A</td>
</tr>
<tr>
<td>23.</td>
<td><em>R. spectabilis</em> A. Gray</td>
<td>CA - USA</td>
<td>PI 553980</td>
</tr>
<tr>
<td><strong>Subg. Lampobatus</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td><em>R. australis</em> G. Forst</td>
<td>New Zealand</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td><strong>Subg. Malachobatus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>25</td>
<td><em>R. assamensis</em></td>
<td>China</td>
<td>PI 618433</td>
</tr>
<tr>
<td></td>
<td><strong>Subg. Orobatus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td><em>R. nubigenus</em> H.B.K.</td>
<td>Ecuador</td>
<td>N/A</td>
</tr>
<tr>
<td>27</td>
<td><em>R. glabratus</em> H.B.K.</td>
<td>Ecuador</td>
<td>PI 548901</td>
</tr>
<tr>
<td></td>
<td><strong>Subg. Rubus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td><em>R. cuneifolius</em> Pursh</td>
<td>AL - USA</td>
<td>N/A</td>
</tr>
<tr>
<td>29</td>
<td><em>R. trivialis</em> Michx.</td>
<td>SC - USA</td>
<td>N/A</td>
</tr>
<tr>
<td>30</td>
<td><em>R. urticifolius</em> Poir.</td>
<td>Ecuador</td>
<td>PI 548929</td>
</tr>
<tr>
<td></td>
<td><strong>Outgroup</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td><em>Rosa multiflora</em> Thunb.</td>
<td>KY - USA</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 2: *rpl16, trnL-trnF, trnK, and trnS-trnG*, and GBSSI-1 sequence characteristics in *Rubus* and *Dalibarda repens*.

<table>
<thead>
<tr>
<th></th>
<th>rpl16 intron</th>
<th>trnL-trnF</th>
<th>trnK</th>
<th>trnS-trnG</th>
<th>trnS-trnG spacer</th>
<th>trnG intron</th>
<th>GBSSI-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total aligned</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nucleotides (bp)</td>
<td>1179</td>
<td>1134</td>
<td>2635</td>
<td>1514</td>
<td>740</td>
<td>751</td>
<td>1621</td>
</tr>
<tr>
<td>% Variable</td>
<td>13.8</td>
<td>12.9</td>
<td>12.1</td>
<td>12.7</td>
<td>15.7</td>
<td>10.1</td>
<td>29.4</td>
</tr>
<tr>
<td>% Parsimony</td>
<td>4.8</td>
<td>3.6</td>
<td>4.9</td>
<td>5.4</td>
<td>6.2</td>
<td>4.7</td>
<td>12.3</td>
</tr>
<tr>
<td>informative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Table 3. Characterization of GBSSI-1 clone sequences for basal species of *Rubus* and *Dalibarda repens* based on position in the GBSSI-1 phylogeny (Fig. 15).

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of clones</th>
<th>Number of unique clones</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>Types of different alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>R. deliciosus</em></td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><em>R. neomexicanus</em></td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td><em>R. odoratus</em></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>R. parviflorus</em></td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td><em>R. chamaemorus</em></td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><em>R. pedatus</em></td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><em>R. lasiococcus</em></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>Dalibarda repens</em></td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 1. (a) Image of *Rubus deliciosus* (Jennifer Ackerfield) and (b) geographic distribution in the US (USDA website).
Figure 2. (a) Image of *Rubus neomexicanus* (David Rodriguez) and (b) geographic distribution in the US (USDA website).
Figure 3. (a) Image of *Rubus odoratus* (L.A. Alice) and (b) geographic distribution in the US (USDA website).
Figure 4. (a) Image of *Rubus parviflorus* (L.A. Alice) and (b) geographic distribution in the US (USDA website).
Figure 5. Image of *Rubus trilobus* from Mexico (L.A. Alice).
Figure 6. (a) Image of *Rubus chamaemorus* (Josef Hlasek) and (b) geographic distribution in the US (USDA website).
Figure 7. (a) Image of *Rubus lasiococcus* (L.A. Alice) and (b) geographic distribution in the US (USDA website).
Figure 8. (a) Image of *Rubus pedatus* (G.D. Carr) and (b) geographic distribution in the US (USDA website).
Figure 9. (a) Image of *Dalibarda repens* (L.A. Alice) and (b) geographic distribution in the US (USDA website).
Figure 10. Organization of the trnS-trnG region in Rubus. Intergenic spacer and introns names are italicized below and amplification and sequencing primer are above the directional arrow. Boxes represent exons. Length of the non-coding region is centered below the intergenic spacer and intron.
Figure 11. Organization of GBSSI-1 gene in *Rubus*. Arrowheads indicate position and direction of amplification (3F/8R) and internal sequencing (4F/6R) primers of the 5' portion of the gene used in the study.
Figure 12. Strict consensus phylogeny based on nuclear ribosomal DNA ITS sequence. Subgenus *Anoplobatus* (red), subg. *Dalibarda* (green). CI = 0.571 and RI = 0.759
Figure 13. Strict consensus phylogeny based on trnS-trnG region sequences. Subgenus *Anoplobatus* (red), subg. *Dalibarda* (green). CI = 0.607, RI = 0.695
Figure 14. Strict consensus phylogeny based on combined analysis of cpDNA sequences. Subgenus *Anoplobatus* (red), subg. *Dalibarda* (green). CI = 0.610, RI = 0.741
Figure 15. Strict consensus phylogeny based on nuclear GBSSI-1 sequences. The three different forms of GBSSI-1α (green), GBSSI-1β (red) and GBSSI-1γ (blue) are shown. CI = 0.6534, R1 = 0.7969
Figure 16. Strict consensus phylogram based on nuclear GBSSI-1 sequences. Three forms of GBSSI-1α (green), GBSSI-1β (red) and GBSSI-1γ (blue) are shown. CI = 0.6534, RI = 0.7969
Figure 17. Strict consensus phylogeny based on nuclear GBSSI-1γ sequences. Subg. *Anoplobatus* (red), subg. *Dalibarda* (green) and *R. chamaemorus* clones (purple).