Syst. Biol. 52(3):374–385, 2003

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150390197000

Ancient Allopolyploid Speciation in Geinae (Rosaceae): Evidence from Nuclear Granule-Bound Starch Synthase (GBSSI) Gene Sequences

JENNY E. E. SMEDMARK,[1,2] TORSTEN ERIKSSON,[1] RODGER C. EVANS,[3] AND CHRISTOPHER S. CAMPBELL[4]

[1]*Bergius Foundation, Royal Swedish Academy of Sciences, Box 50017, SE-104 05 Stockholm, Sweden; E-mail: jenny@bergianska.se (J.E.E.S.), torsten.eriksson@bergianska.se (T.E.)*
[2]*Department of Botany, Stockholm University, SE-106 91 Stockholm, Sweden*
[3]*Biology Department, Acadia University, 24 University Avenue, Wolfville, Nova Scotia B0P 1X0, Canada*
[4]*Department of Biological Sciences, University of Maine, Orono, Maine 04469-5751, USA*

Abstract.— A nuclear low-copy gene phylogeny provides strong evidence for the hybrid origin of seven polyploid species in Geinae (Rosaceae). In a gene tree, alleles at homoeologous loci in an allopolyploid species are expected to be sisters to orthologues in the ancestral taxa rather than to each other. Alleles at a duplicated locus in an autopolyploid, however, are expected to be more closely related to each other than they are to any orthologous copies in closely related species. We cloned and sequenced about 1.9 kilobases from the 5′ end of the GBSSI-1 gene from two diploid, one tetraploid, and six hexaploid species. Each of the three loci in the hexaploid species forms a separate group, two of which are more closely related to copies in other species than they are to each other. This finding indicates that the hexaploid lineage evolved through two consecutive allopolyploidization events. Based on the GBSSI-1 gene tree, we hypothesized that there was an initial hybridization between a diploid species from the ancestral lineage of *Coluria* and *Waldsteinia* and an unknown diploid species to form the tetraploid *Geum heterocarpum* lineage. Backcrossing of *G. heterocarpum* with a representative of the unknown diploid lineage then resulted in a hexaploid lineage that has radiated considerably since its origin, comprising at least 40 extant species with various morphologies. A penalized likelihood analysis indicated that Geinae may be about 17 million years old, implying that the hypothesized allopolyploid speciation events are relatively ancient. Six of the 22 cloned Geinae GBSSI-1 copies in this study, which all are duplicate copies in polyploid taxa, may have become pseudogenes. We compared the GBSSI-1 phylogeny with one from chloroplast data and explored implications for the evolution of some fruit characters. [Allopolyploidy; GBSSI; *Geum*; molecular phylogeny; reticulate evolution; *waxy*.]

The avens and their relatives belong to a taxonomically complicated group of plants where morphological evidence has resulted in ambiguous and conflicting classifications. Hybridization in combination with polyploidy may have generated many of the extant species, contributing to the problematic classification. This is the first study conducted to investigate the occurrence of allopolyploidy within this group using low-copy gene phylogenies to clarify historical relationships.

The Rosaceae clade Colurieae Rydb. includes most of what was formerly called Dryadeae Lam. & DC and consists of about 75 species in *Fallugia* Endl., *Sieversia* Willd., and the herbaceous perennial clade Geinae Schulze-Menz (Smedmark and Eriksson, 2002). Geinae includes *Waldsteinia* Willd., *Coluria* R.Br., *Taihangia* Yü & Li, and *Geum* L. *Geum* contains the majority of the species in the group and has been split into up to nine genera, e.g., *Erythrocoma*, *Acomastylis* (Greene, 1906), and *Oncostylus* (Bolle, 1933). Relationships within Geinae are currently poorly understood, but it appears that none of the previously suggested circumscriptions of *Geum* are monophyletic (Smedmark and Eriksson, 2002).

A majority of species in Geinae are polyploid, from 4*x* to 16*x*. Cytogenetic analyses (Gajewski, 1957, 1958) have suggested that all the Geinae species included in this study, except the diploids *Coluria geoides* and *Waldsteinia geoides*, are of allopolyploid origin. The hexaploid subgenus *Geum*, which is represented by *Geum urbanum* and *G. rivale* in this study, was suggested to belong to a lineage that arose by hybridization between the presumed tetraploid *G. montanum* and a diploid ancestor of *Waldsteinia* or *Coluria*. Achenes of *G. montanum* have persistent styles that become strongly elongated and feathery hairy at maturity. *Waldsteinia* and *Coluria* have deciduous styles that are shed at a joint, a layer of thin-walled cells at the base of the style. Gajewski proposed that this hybridization led to the formation of the fruit type characteristic of subgenus *Geum*, called the fish-hook fruit, type (Iltis, 1913). In this type of fruit, the stiff lower part of the style with a hook at its apex remains, but the outer stigmatic part is shed after pollination. Gajewski's hypothesis was also based on morphological evidence.

The fish-hook fruit type is also present in another hexaploid, *G. vernum*. In crosses with *G. montanum*, *G. vernum* showed a much lower level of meiotic chromosome pairing than did *G. urbanum* and *G. rivale*, which like other species of subgenus *Geum* form the maximum number of bivalents with *G. montanum*. Mainly because of very similar morphology, *G. vernum* was still thought to have the same parent species as *G. urbanum* and *G. rivale*. The hexaploids *Erythrocoma triflora* and *G. reptans*, also included in this study, were suggested to have originated in two additional allopolyploid events.

None of Gajewski's hypotheses were supported by a molecular phylogenetic study including the species mentioned (Smedmark and Eriksson, 2002). However, another species, not included in the cytogenetic analyses (Gajewski, 1957, 1958), *G. andicola*, was placed in different well-supported groups in chloroplast DNA (cpDNA) and nuclear DNA (nrDNA) gene trees. This discordance may be the result of inheritance of cpDNA from one parent and fixation of nrDNA from the other parent. Studies of synthetic allopolyploids of *Brassica* have shown that the nuclear genome in a newly formed allopolyploid may undergo directional changes, possibly because of cytoplasm–nucleus interactions (Song et al., 1995). The

nuclear genome donated by the pollen parent changed significantly during the first generations, whereas that from the ovulate parent did not. If this is a general phenomenon, the comparison of cpDNA and nrDNA data is not likely to provide much information about reticulate evolution because the nrDNA donated by the ovulate parent would tend to become fixed in the allopolyploid genome.

When reconstructing hybrid speciation with single or low-copy nuclear gene sequences, alleles from a putative allopolyploid species are analyzed phylogenetically with sequences from closely related species. This method provides a direct reconstruction of phylogenetic relationships of parental lineages of a hybrid species (Sang and Zhang, 1999). In an allotetraploid there are two homoeologous loci of a specific gene that have been contributed by different ancestral taxa. Their descendant copies are expected to be sister to orthologues in the ancestral taxa rather than to each other. Autotetraploid individuals, however, have two paralogous loci that are more closely related to each other than to orthologous sequences in closely related species.

Low-copy nuclear genes, especially their introns, can be very useful in phylogenetic reconstruction of closely related plant species. However, evolutionary dynamics such as duplications and deletions must be carefully examined to ensure that appropriate loci are compared (Tank and Sang, 2001). Genes that have been thought to be single copy, for example, are often shown to belong to larger gene families even in putatively diploid organisms (Wendel, 2000). Orthologous and homoeologous relationships among genes and inferences of diploid parentage can be assessed with more certainty with recent polyploids (Wendel, 2000). Sang and Zhang (1999), however, pointed out that single-copy genes are especially useful in studying origins of ancient hybrids, because it is not essential that the parents or close relatives of the parents still exist. Their phylogenetic positions are determined by phylogenetic relationships between the parental alleles and those of related species.

The main aim of this study is to use the 5′ portion of the low-copy nuclear gene encoding granule-bound starch synthase (GBSSI or *waxy*) to determine whether seven species of Geinae are allopolyploids. The entire gene includes 13 translated exons and 12 introns. The introns have provided useful information among very closely related species of grasses (Mason-Gamer et al., 1998). They have a high rate of nucleotide substitution and include a large number of characters. So far, the starch synthase gene has been used in a limited number of phylogenetic studies (3′ end, Mason-Gamer et al., 1998; Miller et al., 1999; 5′ end, Evans et al., 2000; Peralta and Spooner, 2001). Data from this region has been used to study reticulate evolution in Poaceae (Mason-Gamer, 2001) and Rosaceae (Evans and Campbell, 2002). Southern hybridization of genomic DNA with locus-specific probes and phylogenetic analyses have revealed two paralogous loci in Rosaceae and Rhamnaceae, GBSSI-1 and GBSSI-2 (Evans et al., 2000), and there are indications that this duplication may be present in a more inclusive clade (Small, 2002). A second duplication has also taken place in relation to the origin of the Maloideae (Evans et al., 2000). Only one locus has been found in most other studied diploid flowering plants (see Mason-Gamer et al., 1998).

We analyzed GBSSI-1 sequences from two diploids, one tetraploid, and six hexaploids in Geinae using parsimony, maximum likelihood, and Bayesian inference. The phylogeny was used to test Gajewski's (1957, 1958) cytogenetic hypotheses that allopolyploid speciation has been prevalent in the group. We also explored whether it is possible to determine the origin of homoeologues of putative allopolyploids despite the limited access to potential ancestral diploid species. To get a rough estimate of the age of Geinae and thereby an idea of when hypothesized allopolyploidizations may have taken place, a molecular dating analysis using the Penalized likelihood method (Sanderson, 2002a) was performed. We compared our GBSSI-1 phylogeny with a chloroplast phylogeny (Smedmark and Eriksson, 2002) in an attempt to shed more light on the phylogeny of the group. On the basis of these two phylogenies, implications for the evolution of some fruit characters, e.g., the abscission layer (joint) present on the style of some species, were explored.

Materials and Methods

Selection of Taxa

We selected two diploid, one tetraploid, and six hexaploid species from Geinae (Table 1) that were indicated by cytogenetic data (Gajewski, 1957, 1958) as possibly having been involved in or produced by allopolyploidy. The choice was governed by access to living material because we have not succeeded in amplifying the desired fragment (1.75–1.90 kilobases [kb]) from herbarium material. We avoided species with high ploidy levels to minimize the problem of paralogy. *Rosa* and *Rubus* were included in the analyses for rooting purposes (Morgan et al., 1994; Eriksson et al., 1998), and trees were rooted on the branch leading to *Rosa multiflora*.

DNA Isolation, Amplification, Cloning, and Sequencing

Total genomic DNA was obtained from fresh material using a slightly modified version of the CTAB extraction method described by Doyle and Doyle (1990). Primers 1F and 9R (Alice, 1997) were initially used to amplify approximately 1.7–2 kb from the 5′ end of GBSSI. Subsequently, new amplification primers were constructed to facilitate amplification of GBSSI-1 specifically (Table 2). These new primers were located 25 base pairs (bp) in the 3′ direction (primer 1F1C) and 70 bp in the 5′ direction (primer 9R1C), respectively, of the earlier primers. The amplified fragment includes 22 bp at the 3′ end of the first exon, seven complete exons, eight complete introns, and 35 bp from the 5′ end of the ninth exon.

Polymerase chain reaction (PCR) amplification was performed in a total volume of 25μl using standard reaction components (Gibco BRL) in a Techne Genius thermal

TABLE 1. List of included Geinae taxa. The list follows the classification of Bolle (1933) and includes information on voucher, ploidy level, clones included in the analyses, number of clones screened, and EMBL accession numbers of the sequences. For information about *Rosa multiflora* and *Rubus odoratus*, see Evans et al. (2000).

Species	Voucher	Ploidy	Clone	No. clones screened	EMBL no.
Erythrocoma triflora (Pursh) Greene	J. Smedmark No. 8 (SBT); cult. Bergius Botanic Garden	Hexaploid	*E. triflora* 1-4	13	AJ534189
			E. triflora 2-4		AJ534199
			E. triflora 5-1		AJ534203
Geum reptans L.	J. Smedmark No. 9 (SBT); cult. Bergius Botanic Garden	Hexaploid	*G. reptans* 1-1	7	AJ534187
			G. reptans 1-5		AJ534185
			G. reptans 1-6		AJ534186
G. montanum L.	J. Smedmark No. 10 (S); wild origin, Switzerland	Hexaploid (Tetraploids have been reported)	*G. montanum* 5-1	4	AJ534190
			G. montanum 5-2		AJ534191
			G. montanum 5-4		AJ534204
G. heterocarpum Boiss.	J. Smedmark No. 5 (SBT); cult. Bergius Botanic Garden	Tetraploid	*G. heterocarpum* 1-1	9	AJ534196
			G. heterocarpum 3-1		AJ534184
G. rivale L.	J. Smedmark No. 7 (S); wild origin, Sweden	Hexaploid	*G. rivale* 1-2	10	AJ534200
			G. rivale 1-3		AJ534192
			G. rivale 1-9		AJ534201
					AJ534202
G. urbanum L.	J. Smedmark No. 3 (S); wild origin, Sweden	Hexaploid	*G. urbanum* 1-1	18	AJ534193
			G. urbanum 1-2		AJ534194
			G. urbanum 1-3		AJ534195
G. vernum (Raf.) Torr. & Gray	J. Smedmark No. 6 (SBT); cult. Bergius Botanic Garden	Hexaploid	*G. vernum* 5-1	17	AJ534198
			G. vernum 5-5		AJ534188
			G. vernum 5-7		AJ534197
Coluria geoides (Pall.) Ldb.	J. Smedmark No. 2 (SBT); cult. Bergius Botanic Garden	Diploid	*C. geoides* 5-2	3	AJ534183
Waldsteinia geoides Willd.	J. Smedmark No. 4 (SBT); cult. Stockholm University	Diploid	*W. geoides* 2-3	4	AJ534182

cycler. The $MgCl_2$ concentration was 3.8 mM. Amplification of other Rosaceae GBSSI with a single annealing temperature has been problematical. Therefore we used the stepdown PCR procedure (Evans et al., 2000), with successively decreasing annealing temperatures (60°C, 56°C, 52°C, 48°C). PCR products were run out on 1% agarose gels, and bands corresponding to a fragment of the expected length (approximately 1.9 kb for GBSSI-1) were excised and purified using a QIAquick Gel Extraction Kit (QIAgen).

Cleaned PCR products were cloned into either pGEM-T Easy (Promega) or TOPO-TA (Invitrogen) vectors. For the pGEM-T Easy vectors, ligation, transformation, plating and selection of clones followed the procedures outlined by Evans et al. (2000), except that DHα competent cells (Gibco BRL) were used. The cloning using the TOPO TA Cloning Kit for Sequencing followed the manufacturer's instructions, except that ligation and transformation reaction volumes were halved and the ligation reaction was allowed to incubate for 1 hr. Positive transformants were used directly as templates in a PCR with standard primers T3 and T7, which attach on either side of the cloning site in the plasmid. To avoid sequencing products from transformants with the wrong insert, the product from the first PCR was tested by attempting to amplify with one of the new amplification primers (9R1C) and one primer internal to GBSSI (3FC). Clones for which amplification products were obtained were sequenced.

TABLE 2. Primers used in the study.

Primer	Sequence (5′–3′)	Use
1F1C	TGGTCTRCCGCCGGCCATGKC	Amplification
9R1C	AGGGCTTCTTTCAAAATAGGC	Amplification, check of paralogue
3FC	GCTTCAAGCGAGGAGTGGATCG	Sequencing, check of paralogue
4RC	GCTGAATCTAAGTTGGTTGTCC	Sequencing
7FC	GCTTACCAAGGGAGATTTGC	Sequencing
7RC	GCAAATCTCCCTTGGTAAGC	Sequencing

Sequencing reactions were performed using ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) according to the manufacturer's instructions, except Ready Reaction mix volumes were diluted four times. All clones were sequenced in both directions using primers T3 and T7 and four internal primers (Table 2). Purification of sequencing reactions followed the ethanol/sodium acetate precipitation protocol provided with the sequencing kit. Sequencing was performed by KISeq at the Center for Genomics Research (Karolinska Institute, Stockholm, Sweden) using 373A and 377 automated sequencers (Applied Biosystems).

Alignment

For sequence editing and assembly, we used the Staden package (Staden, 1996) under Linux. The boundaries between exons and introns were determined by aligning Geinae GBSSI sequences with other Rosaceae GBSSI sequences (Evans et al., 2000). GBSSI-1 sequences were aligned using ClustalX 1.81 with the default settings,

followed by minor corrections made by eye using the sequence alignment editor Se-Al (Rambaut, 1996). All aligned positions were included in the analyses except the fourth intron of *Rosa multiflora* and *Rubus odoratus*, which were excluded from the analyses because they could not be unambiguously aligned with the sequences of the ingroup. Twenty-two potentially informative gaps in the matrix were coded for absence or presence and then included in the parsimony analyses.

Phylogenetic Analyses

Parsimony analyses were conducted using PAUP* 4.0b10 (Swofford, 2002). All characters were weighted equally and were unordered. Indels were treated as missing data. Potentially informative gaps were scored for presence or absence and added as separate characters. A branch-and-bound search with MULTREES was conducted.

Maximum likelihood (ML) analyses were also performed using PAUP* 4.0b10 (Swofford, 2002). Modeltest 3.06 (Posada and Crandall, 1998) was used in combination with PAUP* to select the evolutionary model. The hierarchical likelihood ratio tests in Modeltest selected the Hasegawa–Kishino–Yano (HKY; Hasegawa et al., 1985) nucleotide substitution model with gamma distribution of among-site rate variation (HKY + G; Yang, 1996). In the ML phylogenetic analyses, heuristic searches with 100 random-addition start trees, tree bisection–reconnection (TBR) branch swapping and MULTREES, were performed. For each replicate, PAUP* was allowed to estimate the model parameters.

Bayesian inference (BI) analyses (Huelsenbeck and Ronquist, 2001; Lewis, 2001) were performed using MrBayes 3.0b (Huelsenbeck and Ronquist, 2001). With this program, it is possible to take into consideration the fact that different parts of the alignment, such as coding and noncoding regions, may evolve differently. For these analyses, introns and exons were cut out and made into two separate data sets, each of which was tested with MrModeltest 1.0b (Nylander, 2002). This program is a simplified version of Modeltest that tests only the models of evolution that are implemented in both PAUP* and MrBayes (24 of the 56 models in Modeltest). For both of these data sets, HKY + G was selected.

Three separate BI analyses were performed to make sure that the Markov chain did not fail to find an optimal tree island. For each of these, a random start tree was used and the Markov chain Monte Carlo analyses were run for 1,000,000 generations, with four simultaneous chains, sampling one tree every 10 generations. Model parameters, such as the shape of the gamma distribution and the transition/transversion ratio, were estimated during the analyses and were allowed to differ between introns and exons. The trees sampled after the likelihoods of the trees had converged on a stable value were used to construct a majority rule consensus tree and to calculate the posterior probabilities of clades. In the three separate analyses, the first 1,990, 1,540, and 1,150 trees, respectively, were removed before the consensus tree was computed.

Support for individual nodes was measured by bootstrap values (Felsenstein, 1985), decay indices (Bremer, 1988; Donoghue et al., 1992), and posterior clade probabilities (Huelsenbeck and Ronquist, 2001; Lewis, 2001). For the bootstrap analyses, PAUP* was set to run 100,000 replicates with one random-addition start tree and TBR branch swapping, saving multiple trees. Decay values were obtained using AutoDecay 4.0 (Eriksson, 1999). In the reverse constraint runs for the decay analyses, PAUP* was set to run 100 replicates of heuristic search with random-addition start trees and TBR branch swapping, saving a single tree for each replicate. The posterior clade probabilities obtained with MrBayes are the fraction of occurrences of clades among the sampled trees (Huelsenbeck and Ronquist, 2001). The data set and trees have been submitted to TreeBase (www.treebase.org, accession numbers S862 for the study and M1397 for the matrix).

Molecular Dating

Age estimates of Colurieae and Geinae were calculated using penalized likelihood, as implemented in the r8s program (Sanderson, 2002b). This semiparametric method does not assume clocklike molecular evolution but allows different evolutionary rates on every branch. As with some other methods (Sanderson, 1997; Thorne et al., 1998), there is an element of rate autocorrelation in penalized likelihood, which is based upon the idea that descendants inherit their rate of evolution from their ancestors. In penalized likelihood, a smoothing value determines the relative importance of the likelihood score and the autocorrelation (roughness) penalty for the optimality score. For clocklike data, a high smoothing value will give equal rates on all branches. Nonclocklike data yield low smoothing values, the roughness penalty has little impact on the optimality score, and rates are permitted to vary independently of one another. The optimal smoothing value is determined through a cross-validation procedure (Sanderson, 2002b). A 95% confidence interval was calculated for two relevant nodes using an algorithm included in r8s (Cutler, 2000; Sanderson, 2002b) with a cutoff value of 4.

At least one dated reference point is needed to calibrate the output from the r8s program into actual age estimates. That calibration point should have a dat e assigned based on other data, such as fossil information. We know of no fossils from Colurieae, and the penalized likelihood analysis was therefore performed on a data set representing Rosaceae as a whole, allowing us to use a calibration point outside Colurieae.

In this analysis, we used an *rbc*L data set of 39 Rosaceae species (Morgan et al., 1994) with the addition of a *Rhamnus cathartica* sequence (accession no. L13189; Chase et al., 1993) to root the tree. ML was used in PAUP* to estimate branch lengths with the general time reversible model, gamma rate variation (shape 0.8393) and a proportion of sites assumed invariable (pinvar 0.6867). For model selection and parameter estimations,

we used Modeltest 3.06. A heuristic search was performed in PAUP* with 10 random-addition sequence starting trees, TBR branch swapping, and MULTREES. The best tree found was input to r8s after pruning of *Rhamnus*. The root branch has to be pruned to get a nonzero branch at the base of the tree. The optimal smoothing level was obtained through cross-validation.

An unconstrained penalized likelihood analysis (Sanderson, 2002a) was conducted with the Powell algorithm and then calibrated by setting the Prunoideae node to 44.3 million years before the present (MYBP) (Magallón et al., 1999).

RESULTS

Phylogenetic Analyses

Twenty-two sequences from nine Geinae species and two from other Rosoideae species were included in the analyses. The aligned sequence length was 1,951 bp, of which 0.2% was scored as missing data. Twenty-two binary-coded indels were also added to the matrix. There were 406 parsimony-informative characters, corresponding to 21% of the total number of characters, and 1,038 constant characters. The exons included 14% informative characters, and the introns 23%. Parsimony analyses resulted in 10 most-parsimonious (MP) trees with a length of 1,358 steps. The strict consensus of these is shown in Figure 1. One of the MP trees is identical to the best tree from the ML analysis (log likelihood [− ln L]-9580.27; Fig. 2).

The three BI analyses resulted in 50% majority rule consensus trees with identical topologies. Three of the nodes were not present in identical proportions among the trees sampled from the three analyses. However, they differed by ≤2%. The posterior clade probabilities from one of the analyses are presented in Figure 1, along with parsimony bootstrap values and decay indices.

In the MP strict consensus (Fig. 1), there are three major clades of Colurieae GBSSI-1 clones, A, B, and C, of which the last two are sister groups. Homoeologues from the six hexaploid species can be found in each of these three clades. Clades E and C contain only copies from the hexaploid species and have identical topologies. Clade B, however, is less resolved and includes a clone from

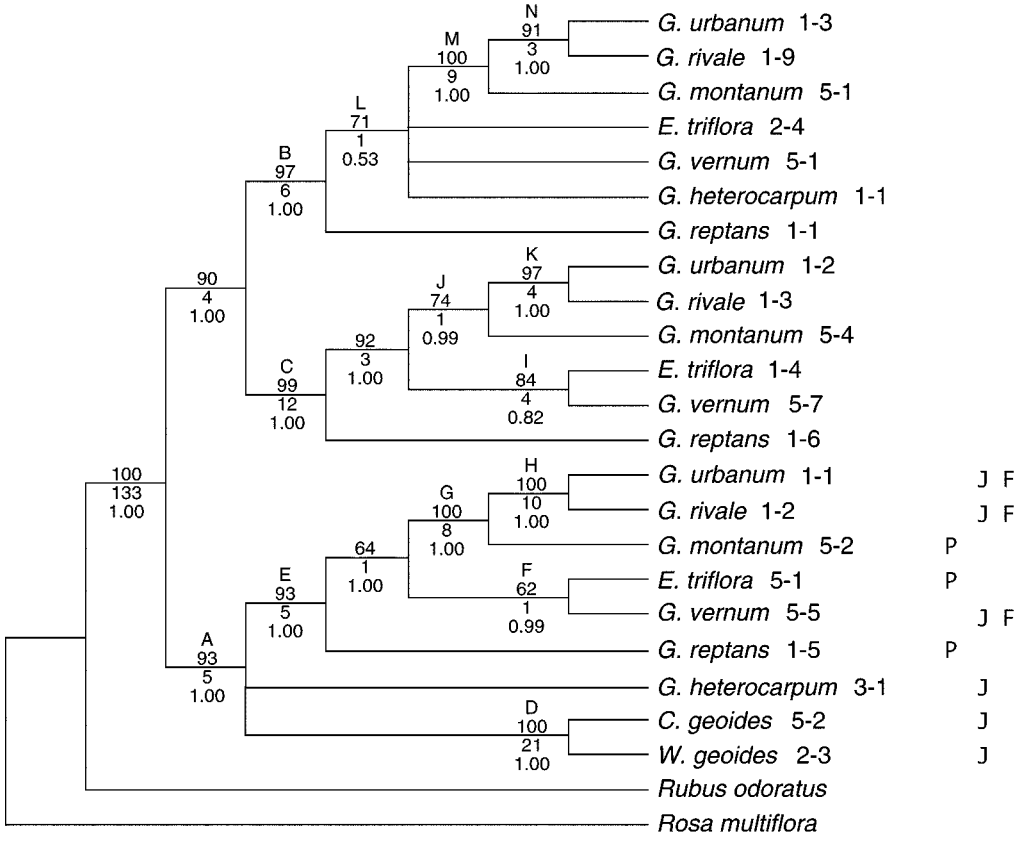

FIGURE 1. Strict consensus of the 10 MP trees (1,358 steps) based on GBSSI-1 exon and intron sequences. Bootstrap values are above branches; decay indices and posterior clade probabilities are below branches. Nodes discussed in the text are indicated by capital letters. *G.* = *Geum*, *E.* = *Erythrocoma*, *W.* = *Waldsteinia*, and *C.* = *Coluria*. Numbers after names are clone identifiers. In clade A, which includes one paralogue from each of the ingroup species, some fruit characters are indicated; P = plumose style; J = jointed style; F = fish-hook fruit type.

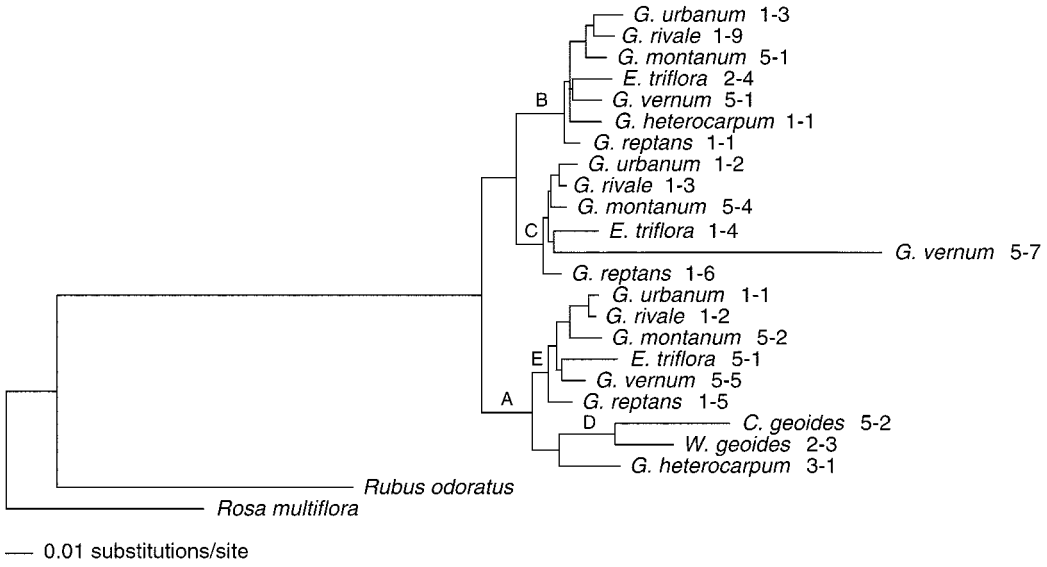

FIGURE 2. Maximum likelihood tree (− ln L-9580.27). The topology is identical to that of 1 of the 10 MP trees. Branch lengths are proportional to inferred substitutions per site. Nodes discussed in the text are indicated by capital letters.

the tetraploid *G. heterocarpum* along with those from the hexaploids. Within clade A, relationships among the hexaploid clade E, the diploid clade D, and a clone from the tetraploid *G. heterocarpum* cannot be resolved except with very low support.

Molecular Dating

The best *rbc*L tree of Rosaceae found (not shown) had a log likelihood likelihood score of 5427.98. This tree is not identical to the parsimony tree of Morgan et al. (1994). However, it is very similar and differs only at a few nodes, which were weakly supported nodes in their analysis. In preparation for the r8s analysis, *Rhamnus* was pruned from the tree. A smoothing value of 0.32 was obtained in the cross-validation run. A low smoothing value indicates departure from clocklike evolution (Sanderson, 2002a). A likelihood ratio test (Felsenstein, 1988; Sanderson, 1998) for the *rbc*L data rejects rate constancy (χ^2-199.67, df-39, $P \ll 0.01$). After calibration of the Prunoideae node at 44.3 MYBP, the age of Colurieae was estimated at 37 MYBP (95% confidence interval [CI], 28–47 MYBP) and that of Geinae at 17 MYBP (95% CI, 10–26 MYBP). The age of Rosoideae was estimated at 54 MYBP and Rosaceae as a whole to 69 MYBP.

DISCUSSION

Allopolyploidy

Several species within Geinae may be allopolyploids. Clones from the same polyploid individual appear in different well-supported clades in the GBSSI-1 gene tree (Fig. 1), and some of these clones are more closely related to clones in other species than they are to each other, indicating that the homoeologues were donated by different ancestral taxa. This situation occurs for all the polyploids included in the study: *G. urbanum*, *G. rivale*, *G. montanum*, *E. triflora*, *G. vernum*, and *G. heterocarpum*. The pattern of relationships among genes indicated in Figure 1 may be explained by the hypothetical historical course of events outlined in Figure 3a. According to this hypothesis, two diploids hybridized to form an allotetraploid species. This new lineage subsequently backcrossed with one of the original diploid parent lineages and a hexaploid was formed.

A gene tree congruent with the topology of the GBSSI-1 tree (Fig. 1) would be obtained if these reticulations were followed by extinction of one of the original diploid species or if we failed to include any representative of this lineage in the analysis and several speciation events took place within the hexaploid and diploid lineages (Figs. 3b, 3c). According to this hypothesis, the hexaploids (Fig. 1, clades E and C and clade B excluding *G. heterocarpum*) would have evolved through two consecutive hybridization events (Fig. 3a). Whether these were single or recurrent formations (Soltis and Soltis, 1995, 1999) cannot be determined from these data because only a single individual of each species is represented. The first hybridization would have involved the ancestral lineage of clade D or perhaps another ancestral diploid of clade A and an unknown diploid species and would have given rise to a tetraploid ancestor of *G. heterocarpum*. The second hybridization would have been the backcrossing of the *G. heterocarpum* lineage with a species from the lineage containing its unknown diploid parent, resulting in a hexaploid lineage. This hexaploid lineage appears as three different clades in the gene tree (Figs. 3b, 3c) because of its double hybrid origin; it seems to have radiated considerably since its formation. In the GBSSI-1 tree (Fig. 1), it corresponds to clades E and C and the hexaploids in clade B. The topologies of clades E and C are identical, but there is no support in these analysis

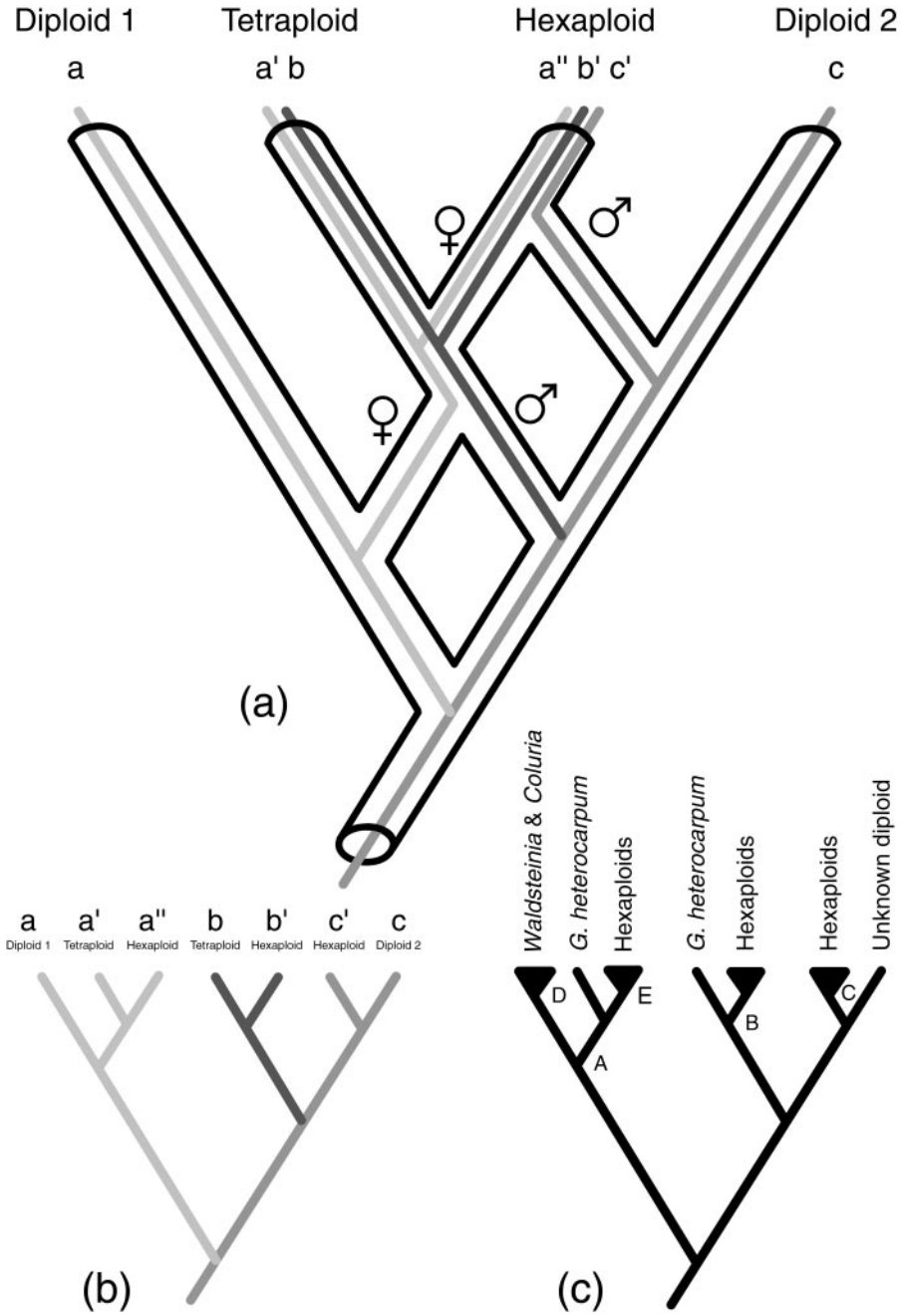

FIGURE 3. A hypothesis about reticulate evolution in Geinae. (a) Taxon tree of four species corresponding to ancestral lineages of extant species in Geinae. A nuclear single-copy gene is marked as shaded lines. The different paralogues are defined by lowercase letters. The phylogeny includes two instances where new lineages have arisen as a result of allopolyploidy. Maternal and paternal lineages were defined by comparing the GBSSI-1 tree (Fig. 1) to the cpDNA tree (Fig. 4). (b) Gene tree of the four species in Figure 3a, based on the nuclear single-copy gene. Homoeologous copies from the allopolyploids are more closely related to orthologous copies in their respective ancestral lineages than they are to each other. (c) GBSSI-1 tree (Fig. 1) adjusted for comparison to the gene tree. Capital letters refer to clades in Figure 1.

for a clade including only the hexaploids in clade B. Here, a paralogue from the tetraploid *G. heterocarpum* is nested among the hexaploids. However, this topology received low support (Fig. 1, node L: bootstrap = 71, decay index = 1, posterior probability = 0.53) and may be the result of recombination.

The topology of the bootstrap tree based on the *trn*L–*trn*F region (Smedmark and Eriksson, 2002; Fig. 4) is congruent with clade A in the GBSSI-1 tree (Fig. 1), i.e., the clade containing one paralogue from each species. Because the *trn*L–*trn*F region is located in the chloroplast, which is maternally inherited in Rosaceae (Corriveau

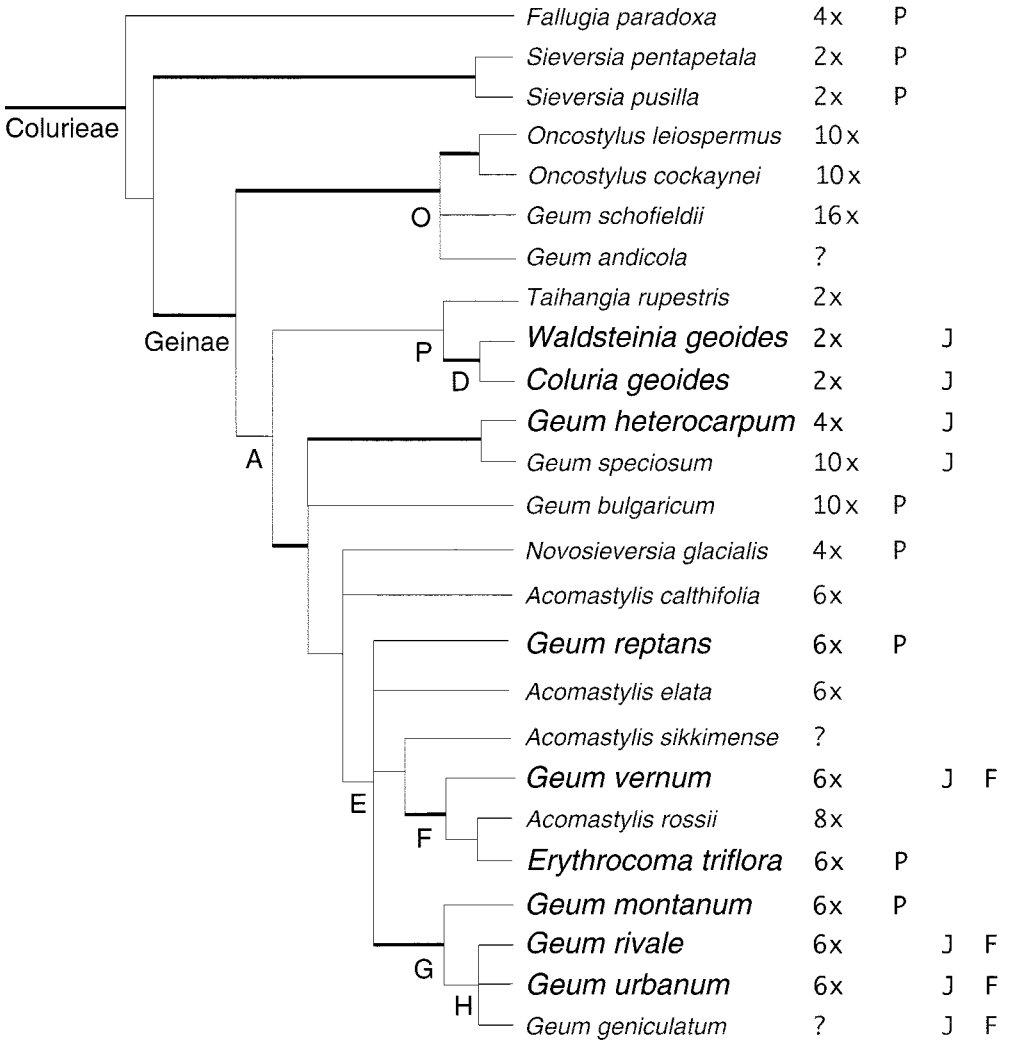

FIGURE 4. Parsimony bootstrap tree of Colurieae, based on the chloroplast *trn*L–*trn*F region (Smedmark et al., 2002). Branches with bootstrap support >90% are shown with thicker lines. Capital letters refer to clades in Figure 1. Clade O is a potential paternal lineage involved in the two hypothesized allopolyploidizations (corresponding to Diploid 2 and unknown diploid in Fig. 3). Species included in the present study are printed in larger font. Ploidy level and fruit type (where applicable; see Figure 1) are given after the species name. One species, *Coluria elegans*, which was included in the original analysis (Smedmark et al., 2002), has been excluded from this figure, because we later found out that the specimen was incorrectly determined. It is not a *Coluria* at all.

and Coleman, 1988), this bootstrap tree represents the maternal phylogeny. Thus, clade A (Fig. 1) shows the phylogeny of the maternal lineage and its descendant homoeologues in the polyploids, which indicates that a member of the ancestral lineage of *C. geoides* and *W. geoides* (Fig. 4, clade P) was the ovulate parent in the hybridization giving rise to *G. heterocarpum*. The *G. heterocarpum* lineage should also have been the maternal component in the second hybridization, for clade A in the GBSSI-1 tree to be congruent with the *trn*L–*trn*F tree. The b homoeologues 2in the gene tree (Fig. 3b) do not appear in the *trn*L–*trn*F tree, because they are derived from the pollen parents. According to this hypothesis, the unknown diploid paternal lineage should be sought outside the clade corresponding to the GBSSI-1 clade A in the *trn*L–*trn*F phylogeny. One clade in the *trn*L–*trn*F tree (Fig. 4, clade O) is the sister to a clade corresponding to clade A in the GBSSI-1 tree (Fig. 2). This potential paternal lineage includes *G. schofieldii* and the southern hemisphere *Oncostylus cockaynei*, *O. leiospermus*, and *G. andicola*.

The position of *G. heterocarpum* clone 3-1 is not resolved in the GBSSI-1 bootstrap tree (Fig. 1). In the ML tree (Fig. 2) and 5 of the 10 MP trees, it is the sister of *Waldsteinia* and *Coluria*. This result is supported by BI analyses with a posterior probability of 1.00. In the other five MP trees, *G. heterocarpum* 3-1 is sister to the hexaploid clade E, a relationship that also is supported by the *trn*L–*trn*F region (bootstrap = 99, decay index = 6, posterior probability = 1.00; Smedmark and Eriksson, 2002). This topology agrees with the hypothesis presented above.

Alternative evolutionary scenarios are always possible. Although we favor the one described above because it is simple, the limited sampling in this study (because we have been unable to amplify the desired DNA region from herbarium specimens) might make this scenario potentially sensitive to the addition of more taxa. Ideally, all diploids that might represent ancestral lineages of the polyploids should be sampled. However, at present the only other known diploid in Geinae besides a few species in *Coluria* and *Waldsteinia*, is *Taihangia rupestris*, their sister according to chloroplast data (Fig. 4). Some taxa with higher ploidy levels that could not be sampled for this study are indicated by cpDNA and nrDNA data (Smedmark and Eriksson, 2002) to be highly relevant for an expanded study, e.g., the tetraploid *Novosieversia glacialis*, which is placed between the diploid clade D and the hexaploid clade E (Fig. 4), *G. andicola* of unknown ploidy level that is indicated to have different ancestries in the ITS and *trn*L–*trn*F trees (Smedmark and Eriksson, 2002), and a representative of the potential paternal lineage, clade O (Fig. 4), e.g., an *Oncostylus* species.

Alternative interpretations of our results, not involving allopolyploidy, could be based on the presence of three GBSSI-1 paralogues before the diversification of Geinae. Either GBSSI-1 underwent two duplications in the diploid progenitor of Geinae or the ancestral lineage of Geinae was hexaploid. The first of these explanations implies either that lineage sorting has taken place in the diploids and the tetraploid or that we have not succeeded in discovering all three paralogues in these taxa. However, three paralogues in a diploid progenitor of Geinae would result in the tetraploids having 6 paralogues each and the hexaploids 12. It seems unlikely that we happened to sample one each of the original paralogues, deriving from the diploid, and missed the additional copies present in the polyploids.

If, however, Geinae were originally hexaploid the polyploidizations must have taken place before the divergence of clade D, because the diploid clade D (Fig. 1) is in clade A rather than being sister to all three paralogues. Although superficially parsimonious, this interpretation implies a reduction of the ploidy level and thus loss of GBSSI-1 loci in two lineages: one from hexaploidy to tetraploidy (*G. heterocarpum*) and one from hexaploidy to diploidy (*W. geoides* and *C. geoides*). Furthermore, comparison with the *trn*L–*trn*F tree (Fig. 4), with wider taxon sampling, indicates that at least one more reduction to the tetraploid level in *N. glacialis* would be required. We are not aware of any cases where the ploidy level of a lineage has been shown to have undergone reduction by the loss of entire genomes. Both of these alternative explanations seem more complicated than the one we propose here.

Geum montanum has sometimes been reported to be tetraploid (Gajewski, 1957; Moore, 1973). The position in the tree of the three distinct copies of GBSSI-1 indicates that there are three loci of this gene, which means that at least the specimen we sampled is hexaploid, in agreement with chromosome counts by other workers (e.g., Goldblatt, 1985). In a cytogeographic study of *G. montanum*, Krahulcová (1994) did not find any tetraploids, even in the localities where they previously had been reported. This absence is remarkable because Gajewski studied the group for decades, and *G. montanum* played an important part in his research and his conclusions about the evolution in the group. For example, he crossed this species with at least 15 other species, made numerous chromosome counts, and studied and made drawings of the chromosome configurations during meiosis; yet he did not report ever having encountered hexaploid individuals. Given our hypothesis, the contradicting reports of the ploidy level of *G. montanum* imply that the species evolved within the allohexaploid clade and later underwent a reduction of the ploidy level in a geographically limited population. Although unlikely, perhaps two morphologically similar but phylogenetically distinct species with different ploidy levels exist.

Age Estimate of Geinae

The age of neither Colurieae nor Geinae is known. The Rosaceae are represented in the fossil record by several specimens, e.g., fruits and wood of Prunoideae from the Eocene, 44.3 MYBP (Magallón et al., 1999). Wikström et al. (2001) estimated that the Rosaceae originated 76 MYBP, using nonparametric rate smoothing (Sanderson, 1997) and a three-gene data set (Soltis and Soltis, 1999; Soltis et al., 2000). Our results from molecular dating analyses using the penalized likelihood method (Sanderson, 2002a) and sequence data from *rbc*L indicate that Colurieae may have appeared at least 37 MYBP and Geinae at 17 MYBP. The proposed hybridizations must have taken place early in the history of Geinae. If the age estimate of Geinae is anywhere near correct, the time span seems to be reasonable for the substantial speciation that has taken place since the formation of the allopolyploid lineage and for these species to reach the wide geographic distribution they have today.

Fruit Morphology

The fruits in Colurieae are achenes, some with persistent styles and some with partly or entirely deciduous styles. The plesiomorphic type of fruit within the group is wind dispersed with a strongly elongating, feathery hairy style. This type of fruit is also present in several groups within Geinae (Smedmark and Eriksson, 2002; see also Fig. 4). The fruits with deciduous styles have a layer of thin-walled cells, a joint, where the outer part is abscised when the fruit is mature.

According to the bifurcating phylogenies presented here (e.g., Fig. 1), the joint has evolved in at least three different groups; among the species sampled for this study, it is present in *C. geoides, W. geoides, G. heterocarpum, G. vernum, G. urbanum*, and *G. rivale*. A possibility that agrees with the reticulate taxon tree hypothesis (Fig. 3a) is that the character *jointed style* is associated with genome a in Figure 3. Thus, the joint would have been passed from an ancestor of *Waldsteinia* and *Coluria*, with styles jointed at the base, to *G. heterocarpum* in the first hybridization. All of the species in the paternal lineage suggested above (Fig. 4, clade O) have persistent styles. The fruits of *G. heterocarpum* are harpoonlike, with an outer deciduous

segment of the style whose removal exposes a stiff inner part with reflexed bristles at the apex. In the second hybridization, the joint would then be passed on from *G. heterocarpum* to the hexaploid clade, where various types of fruits have evolved, some with plumose styles and some jointed. The geographically most widespread group of the Geinae containing the largest number of species is *Geum* subgenus *Geum* ("*Eugeum*," Bolle, 1933), represented here by *G. urbanum* and *G. rivale* (Fig. 1, nodes H, K, and N). They have fruits with an outer deciduous segment and a stiff persistent inner segment that is hooked at the apex. These fruits, of the fish-hook fruit type, are dispersed by animals. Both nuclear and chloroplast data indicate that the fish-hook fruit type formed twice within the hexaploid clade; it also is present in *G. vernum* (Figs. 1, 4; nodes F, I, and L).

Comparison with Earlier Cytogenetic Results

Although GBSSI sequence data largely support hypotheses about genome homology based on cytogenetic studies in Triticeae (Mason-Gamer, 2001), this is not the case in Geinae. The results presented here give strong support for Gajewski's conclusion (1957, 1958) that allopolyploidy has played an important role in the history of Geinae. However, his interpretations, based on cytogenetic studies, of homology of genomes in different species seem to be largely incorrect. *Geum urbanum* and *G. rivale* do share three homoeologous genomes corresponding to the three nodes B, C, and E in Figure 1, but they are not the product of a hybridization between *G. montanum* and an ancestor of *Waldsteinia* or *Coluria*, as Gajewski concluded. First, the *G. montanum* specimen used in the present study is hexaploid rather than tetraploid, and second, all its three subgenomes are homologous with those of *G. urbanum* and *G. rivale*, because its clones reside in the same clades (Fig. 1, clades G, J, and M). Thus, the *Waldsteinia* and *Coluria* lineage may be involved in the origin of the clade in which all three of these species belong (Fig. 3). Gajewski also postulated (1957, 1958) that *G. reptans* was formed through allopolyploidy between *G. montanum* and an unknown diploid species. *Erythrocoma triflora* was interpreted to have two subgenomes in common with *G. urbanum* and *G. rivale* and only one with *G. montanum*, indicating a separate polyploid origin. Neither of these hypotheses is supported by our results. Rather, *G. reptans* and *E. triflora* both seem to be descendants of the same hexaploid that was ancestral to *G. montanum*, *G. urbanum*, and *G. rivale* (Fig. 1, nodes E and C). Gajewski noted that very few bivalents were formed during meiosis in the pollen mother cells of hybrids between subgenus *Geum* species, such as *G. urbanum* and *G. rivale*, and *G. vernum* of subgenus *Stylipus*. Because *G. vernum* resembles subgenus *Geum* species in the fish-hook type of fruit, Gajewski concluded that they are closely related. However, both nrDNA (Fig. 1, nodes F and I) and cpDNA (Fig. 4, node F) sequences indicate that *G. vernum* is more closely related to, e.g., *Erythrocoma triflora* than to species of the subgenus *Geum*.

Gajewski's hypothesis (1957, 1958) about the origin of the fish-hook fruit of subgenus *Geum* is rejected by these results. He concluded that this type of fruit is the product of a hybridization between a diploid species with a style that is jointed at the base and deciduous in its entirety and a tetraploid species with a persistent, elongating, and feathery hairy style. He suggested these parental lineages to be ancestors of *Waldsteinia* or *Coluria* and *G. montanum*, respectively. Our data do not support this hypothesis.

Extensive and rapid reorganization of polyploid genomes is to be expected (Song et al., 1995; Ozkan et al., 2001; Shaked et al., 2001). The cytogenetic analyses (Gajewski, 1957, 1958) indicate that genome restructuring has been more profound in some species of the hexaploid clade than in others. *Geum vernum* appears to have undergone the most thorough reorganization of its genome, judging from the relatively low number of bivalents formed at meiosis in hybrids with the other species in this clade.

Gene Level Processes in Polyploids

Genes duplicated by polyploidy may retain their original function or develop a similar one (Cronn et al., 1999). Duplicates may be preserved through loss or reduction of expression for different subfunctions so that both copies are necessary for normal functioning (Force et al., 1999). Duplication may also lead to relaxation of selection, allowing one copy to undergo functional diversification or become a pseudogene (Wendel, 2000). Duplicated genes may interact through interlocus recombination, gene conversion, or concerted evolution (Wendel, 2000).

Of the 22 cloned GBSSI-1 sequences in this study, 9 contain stop codons and may have become pseudogenes. Most of these nine sequences have deletions or insertions in the exon regions, leading to the formation of several stop codons. However, in three of them, single substitutions have caused the formation of stop codons. The amino acid codons downstream of a stop codon would be expected to accumulate nonsense mutations if they were not translated into a polypeptide. Because this is not the case in these three sequences, the stop codons may either be PCR artifacts or they may be read through. Evidence from *Escherichia coli* and *Drosophila* indicates that internal stop codons may be partially read through, perhaps by a type of tRNA that will insert an appropriate amino acid (Klagges et al., 1996).

The most divergent Geinae GBSSI-1 sequence is that of *G. vernum* 5-7, which has a very long terminal branch in the ML/MP tree (Fig. 2). It starts to deviate from the other sequences in the eighth exon. Despite this considerable divergence, it is still alignable with the other sequences. This sequence does not contain any stop codons. However, some of the intron splice sites are different from those of all other GBSSI-1 clones. The third, fifth, and seventh introns all start with GG and the eighth starts with AG instead of the highly conserved GT. The seventh intron also ends with AA, instead of the conserved AG. These alterations in splice sites may have rendered this gene copy inactive. However, a GT to TT mutation at

the 5′ slice site of an intron in one allele of the *waxy* locus in rice only results in a 10-fold reduction of translation (Isshiki et al., 1998).

Thus, at least six of the sequences may have become pseudogenes, but the inactivation of these paralogues must be quite recent because it has not yet led to long branches. However, we found at least one functional GBSSI-1 copy in each species. In some species, two or even three homoeologous copies seem to have persisted in a functional state. For example, in *G. reptans* and *G. montanum* all three GBSSI-1 copies seem to be functional. The occurrence of homogenizing processes, such as recombination, gene conversion, or concerted evolution, among homoeologues in these species has not yet been studied. The presence of three types of sequences in all the hexaploid species provides evidence against total homogenization of homoeologues in any of the species. There is however an indication that *G. reptans* 1-1 might have been affected by partial homogenization or perhaps PCR recombination. This clone is sister of the rest of clade B, which includes *G. heterocarpum* and the hexaploids, in all the MP trees and in the ML tree (Fig. 2), although this relationship was not well supported (Fig. 1). This topology is different from that of the other two clades containing hexaploids (Fig. 1, clades E and C). Here, the *G. reptans* paralogues are sisters of topologically identical clades containing only hexaploids. A closer examination of the paralogues from this species reveals that the first 204 positions of the alignment are identical in clones 1-1 (Fig. 1, clade B) and 1-6 (clade C). A comparison of clones 1-1 and 1-5 (clade E) reveals a single position in the first 204 where the two sequences differ. In contrast, the other five hexaploid species differ in this region by 13–17 positions for clones in clade B and those in clade C and 29–42 positions for clones in clades B and A. These differences may be an indication that recombination between all three *G. reptans* paralogues at the 5′ end of the amplified region has occurred.

A more complex explanation would be needed if one of the loci in the hexaploids had a history different from that of the other two versus all sharing the same history. Shared history would indicate that a hexaploid lineage formed through polyploidy had experienced a number of speciations. Therefore, partial homogenization of *G. reptans* 1-1 probably has "pulled" it down to a basal position in clade B. Whether this recombination is a result of PCR recombination or homogenizing processes in the cell is not possible to deduce from this analysis.

CONCLUSIONS

Phylogenetic analyses of GBSSI-1 sequences indicate that hybridizations in combination with polyploidy have been important mechanisms for generating diversity in Geinae. Clones from the same polyploid individuals occur in different well-supported clades in the GBSSI-1 tree (Fig. 1). Some of these clades are more closely related to paralogues in other species than they are to each other. Our data are consistent with two relatively ancient allopolyploidization events in the history of the group. Although this study supports the conclusions based on cytogenetic data of the occurrence of allopolyploid speciation within Geinae, specific hypotheses about the origin of individual species (Gajewski, 1957, 1958) are not supported. The six hexaploid species included in the present study belong to the same clade, which seems to have been formed through allopolyploidy. The tetraploid *G. heterocarpum* also appears to be of allopolyploid origin. The exact ancestry of these polyploid lineages cannot be deduced with confidence from these results because of limited sampling of diploids. Despite the low number of taxa included in this analysis, results based on chloroplast data from the *trn*L–*trn*F region (Smedmark and Eriksson, 2002) are corroborated. One implication for the evolution of fruits within the group, supported by both nuclear and chloroplast data, is that the intricate fruit type of subgenus *Geum* has evolved twice, because *G. vernum* does not form a clade with *G. urbanum* and *G. rivale*.

ACKNOWLEDGMENTS

We are very grateful to Zack Murrell for providing material of *Geum vernum*, to Neva Hax and Wes Wright for help in the lab, to Bengt Oxelman, Niklas Wikström, Tao Sang, and Jeff Doyle for helpful and stimulating discussion, and to Birgitta Bremer, Steven J. Brunsfeld, and an anonymous reviewer for comments on a previous version of the manuscript. This study was supported by a National Science Foundation (USA) grant (DEB-9806945) to Christopher Campbell.

REFERENCES

ALICE, L. A. 1997. Molecular phylogenetic systematics of *Rubus* (Rosaceae). Ph.D. Dissertation, Univ. of Maine, Orono.

BOLLE, F. 1933. Eine Übersicht über die Gattung *Geum* L. und die ihr nahestehenden Gattungen. Feddes Repert. Beih. 72:1–119.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution. 42:795–803.

CHASE, M. W., D. E. SOLTIS, R. G. OLMSTEAD, D. MORGAN, D. H. LES, B. D. MISHLER, M. R. DUVALL, R. A. PRICE, H. G. HILLS, Y.-L. QIU, K. A. KRON, J. H. RETTIG, E. CONTI, J. D. PALMER, J. R. MANHART, K. J. SYTSMA, H. J. MICHAELS, W. J. KRESS, K. G. KAROL, W. D. CLARK, M. HEDRÉN, B. S. GAUT, R. K. JANSEN, K.-J. KIM, C. F. WIMPEE, J. F. SMITH, G. R. FURNIER, S. H. STRAUSS, Q.-Y. XIANG, G. M. PLUNKETT, P. M. SOLTIS, S. M. SWENSEN, S. E. WILLIAMS, P. A. GADEK, C. J. QUINN, L. E. EGUIARTE, E. GOLENBERG, G. H. LEARN, JR., S. W. GRAHAM, S. C. H. BARRETT, S. DAYANANDAN, AND V. A. ALBERT. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Mo. Bot. Gard. 80:528–580.

CORRIVEAU, J. L., AND A. W. COLEMAN. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. Am. J. Bot. 75:1443–1458.

CRONN, R. C., R. L. SMALL, AND J. F. WENDEL. 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. USA 96:14406–14411.

CUTLER, D. J. 2000. Estimating divergence times in the presence of an overdispersed molecular clock. Mol. Biol. Evol. 17:1647–1660.

DONOGHUE, M. J., R. G. OLMSTEAD, J. F. SMITH, AND J. D. PALMER. 1992. Phylogenetic relationships of Dipsacales based on *rbc*L sequences. Ann. Mo. Bot. Gard. 79:333–345.

DOYLE, J. J., AND J. L. DOYLE. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13–15.

ERIKSSON, T. 1999. *AutoDecay*. Hypercard stack distributed by the author, Royal Swedish Academy of Sciences, Stockholm.

ERIKSSON, T., M. J. DONOGHUE, AND M. S. HIBBS. 1998. Phylogenetic analysis of *Potentilla* using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of Rosoideae (Rosaceae). Plant Syst. Evol. 211:155–179.

EVANS, R. C., L. A. ALICE, C. S. CAMPBELL, E. A. KELLOGG, AND T. A. DICKINSON. 2000. The granule-bound starch synthase (GBSSI) gene in the Rosaceae: Multiple loci and phylogenetic utility. Mol. Phylogenet. Evol. 17:388–400.

EVANS, R. C., AND C. S. CAMPBELL. 2002. The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. Am. J. Bot. 89:1478–1484.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. 22:521–565.

FORCE, A., M. LYNCH, F. B. PICKETT, A. AMORES, Y. YAN, AND J. POSTLETHWAIT. 1999. Preservation of duplicate genes by complementary degenerative mutations. Genetics 151:1531–1545.

GAJEWSKI, W. 1957. A cytogenetic study on the genus *Geum*. Monogr. Bot. 4:3–414.

GAJEWSKI, W. 1958. Evolution in the genus *Geum*. Evolution 13:378–388.

GOLDBLATT, P. (ed.). 1985. Index to plant chromosome numbers. 1982–1983. Mo. Syst. Bot. Mo. Bot. Gard. 13.

GREENE, E. L. 1906. Segregates from *Sieversia*. Pages 174–179 *in* Leaflets of botanical observation and criticism, Volume 1. Washington, D.C.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 21:160–174.

HUELSENBECK, J. P., AND F. R. RONQUIST. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

ILTIS, H. 1913. Über das gynophor und die Fruchtausbildung bei der Gattung *Geum*. Sitzungsber. Königl. Preuss. Akad. Wiss. Berl. 122:1–36.

ISSHIKI, M., K. MORINO, M. NAKAJIMA, R. J. OKAGAKI, S. R. WESSLER, T. IZAWA, AND K. SHIMAMOTO. 1998. A naturally occurring functional allele of the rice *waxy* locus has a GT to TT mutation at the 5′ splice site of the first intron. Plant J. 15:133–138.

KLAGGES, B. R. E., G. HEIMBECK, T. A. GODENSCHWEGE, A. HOFBAUER, G. O. PFUGFELDER, R. REIFEGERSTE, D. REISCH, M. SCHAUPP, S. BUCHNER, AND E. BUCHNER. 1996. Invertebrate synapsins: A single gene codes for several isoforms in *Drosophila*. J. Neurosci. 16:3154–3165.

KRAHULCOVÁ, A. 1994. Cytogeography of *Geum montanum* (Rosaceae). Folia Geobot. Phytotaxon. 29:85–90.

LEWIS, P. O. 2001. Phylogenetic systematics turns over a new leaf. Trends Ecol. Evol. 16:30–37.

MAGALLÓN, S., P. R. CRANE, AND P. S. HERENDEEN. 1999. Phylogenetic pattern, diversity, and diversification of eudicots. Ann. Mo. Bot. Gard. 86:297–372.

MASON-GAMER, R. J. 2001. Origin of North American *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. Syst. Bot. 26:757–768.

MASON-GAMER, R. J., C. F. WEIL, AND E. A. KELLOGG. 1998. Granule-bound starch synthase: Structure, function, and phylogenetic utility. Mol. Biol. Evol. 15:1658–1673.

MILLER, R. E., M. D. RAUSHER, AND P. S. MANOS. 1999. Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and *waxy* sequences. Syst. Bot. 24:209–227.

MOORE, R. J. (ed.). 1973. Index to plant chromosome numbers. 1967–1971. Regnum Veg. 90.

MORGAN, D. R., D. E. SOLTIS, AND K. R. ROBERTSON. 1994. Systematic and evolutionary implications of *rbc*L sequence variation in Rosaceae. Am. J. Bot. 81:890–903.

NYLANDER, J. A. A. 2002. MrModeltest, version 1.0b. Computer program distributed by the author, Department of Systematic Zoology, Uppsala University, Sweden.

OZKAN, H., A. A. LEVY, AND M. FELDMAN. 2001. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops–Triticum*) group. Plant Cell 13:1735–1747.

PERALTA, I. E., AND D. M. SPOONER. 2001. Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section Lycopersicon [Mill.] Wettst. subsection Lycopersicon). Am. J. Bot. 88:1888–1992.

POSADA, D., AND K. A. CRANDALL. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817–818.

RAMBAUT, A. 1996. Se-Al; Sequence alignment editor. Computer program, Department of Zoology, Oxford Univ. Oxford, U.K.

SANDERSON, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.

SANDERSON, M. J. 1998. Estimating rates and time in molecular phylogenies: Beyond the molecular clock? Pages 242–264 *in* Molecular systematics of plants. II: DNA sequencing (D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds.). Kluwer AP, Boston.

SANDERSON, M. J. 2002a. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

SANDERSON, M. J. 2002b. r8s: Analysis of rates ("r8s") of evolution (and other stuff), version 1.5. Section of Evolution and Ecology, Univ. of California, Davis. Distributed at http://ginger.ucdavis.edu/r8s/.

SANG, T., AND D. ZHANG. 1999. Reconstructing hybrid speciation using sequences of low copy nuclear genes: Hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. Syst. Bot. 24:148–163.

SHAKED, H., K. KASHKUSH, H. OZKAN, M. FELDMAN, AND A. A. LEVY. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridisation and allopolyploidy in wheat. Plant Cell 13:1749–1759.

SMALL, R. L. 2002. The GBSSI gene family in Malvaceae. (Abstr.) Available at www.2002.botanyconference.org/section12/abstracts/126.shtml.

SMEDMARK, J. E. E., AND T. ERIKSSON. 2002. Phylogenetic relationships of *Geum* (Rosaceae) and relatives inferred from the nrITS and *trn*L–*trn*F regions. Syst. Bot. 27:303–317.

SOLTIS, D. E., AND P. S. SOLTIS. 1995. The dynamic nature of polyploid genomes. Proc. Natl. Acad. Sci. USA 92:8089–8091.

SOLTIS, D. E., AND P. S. SOLTIS. 1999. Polyploidy: Recurrent formation and genome evolution. Trends Ecol. Evol. 14:348–352.

SOLTIS, D. E., P. S. SOLTIS, M. W. CHASE, M. E. MORT, D. C. ALBACH, M. ZANIS, V. SAVOLAINEN, W. H. HAHN, S. B. HOOT, M. F. FAY, M. AXTELL, S. M. SWENSEN, K. C. NIXON, AND J. S. FARRIS. 2000. Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbc*L, and *atp*B sequences. Bot. J. Linn. Soc. 133:381–461.

SONG, K., P. LU, K. TANG, AND T. C. OSBORN. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA 92:7719–7723.

STADEN, R. 1996. The Staden sequence analysis package. Molec. Biotech. 5:233–241.

SWOFFORD, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, Massachusetts.

TANK, D. C., AND T. SANG. 2001. Phylogenetic utility of the glycerol-3-phosphate acyltransferase gene: Evolution and implications in *Paeonia* (Paeoniaceae). Mol. Phylogenet. Evol. 19:421–429.

THORNE, J. L., H. KISHINO, AND I. S. PAINTER. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

WENDEL, J. F. 2000. Genome evolution in polyploids. Plant. Mol. Biol. 42:225–249.

WIKSTRÖM, N., V. SAVOLAINEN, AND M. W. CHASE. 2001. Evolution of the angiosperms: Calibrating the family tree. Proc. R. Soc. Lond. B Biol. Sci. 268:2211–2220.

YANG, Z. 1996. Among-site variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11:367–371.

First submitted 23 August 2002; reviews returned 3 November 2002; final acceptance 9 February 2003
Associate Editor: Roberta Mason-Gamer