American Journal of Botany 93(3): 412–425. 2006.

POLYPLOID AND HYBRID EVOLUTION IN ROSES EAST OF THE ROCKY MOUNTAINS[1]

SIMON JOLY,[2,5] JULIAN R. STARR,[3] WALTER H. LEWIS,[4] AND ANNE BRUNEAU[2]

[2]Institut de Recherche en Biologie Végétale, Université de Montréal, 4101 Sherbrooke est, Montréal, Québec H1X 2B2 Canada; [3]Department of Biology, 214 Shoemaker Hall, University of Mississippi, University, Mississippi 38677 USA; and [4]Biology Department, Washington University, Campus Box 1137, St. Louis, Missouri 63130 USA

This study investigates the impact of hybridization and polyploidy in the evolution of eastern North American roses. We explore these processes in the *Rosa carolina* complex (section *Cinnamomeae*), which consists of five diploid and three tetraploid species. To clarify the status and origins of polyploids, a haplotype network (statistical parsimony) of the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) nuclear gene was estimated for polyploids of the complex and for diploids of section *Cinnamomeae* in North America. A genealogical approach helped to decipher the evolutionary history of polyploids from noise created by hybridization, incomplete lineage sorting, and allelic segregation. At the diploid level, species west of the Rocky Mountains are distinct from eastern species. In the east, two groups of diploids were found: one consists of *R. blanda* and *R. woodsii* and the other of *R. foliolosa*, *R. nitida*, and *R. palustris*. Only eastern diploids are involved in the origins of the polyploids. *Rosa arkansana* is derived from the *blanda–woodsii* group, *R. virginiana* originated from the *foliolosa–nitida–palustris* group, and *R. carolina* is derived from a hybrid between the two diploid groups. The distinct origins of these polyploid taxa support the hypothesis that the three polyploids are separate species.

Key words: haplotype network; incomplete lineage sorting; multiple origins; polyploidy; reticulate evolution; *Rosa carolina* complex; statistical parsimony.

Wild species of roses are characterized by extensive morphological variation, which has resulted in a notoriously complex taxonomy. For instance, Linnaeus (Stearn, 1957, p. 158) wrote in *Species Plantarum*, "The species of *Rosa* are with difficulty to be distinguished, with even greater difficulty to be defined; nature seems to me to have blended several or by way of sport to have formed several from one." North American roses are no exception; Crépin (1896), Watson (1885), Rydberg (1920), and Erlanson MacFarlane (1966) described 13, 18, 129, and 22 *Rosa* species on this continent, respectively. Hybridization has long been considered to be one of the major causes of taxonomic confusion (Linnaeus, 1753; Crépin, 1894, 1896), and artificial crosses have shown that in fact most diploids are interfertile (Erlanson, 1934; Ratsek et al., 1939, 1940; Lewis and Basye, 1961). Cytological studies during the early 20th century demonstrated that polyploidy is frequent in *Rosa* (Täckholm, 1922; Hurst, 1925) and that it could represent another source of variation. The present research explores issues related to hybridization and polyploidy, two important processes in plant evolution (Arnold, 1997; Otto and Whitton, 2000), that may explain the difficulty of recognizing species in wild roses.

This study focuses on the North American *Rosa carolina* L. complex of section *Cinnamomeae*, a group that epitomizes the complexity of the genus. Indeed, Lewis (1957c, p. 126) considered the group to be "... the most difficult taxonomic problem in our North American *Rosa*." The complex consists of five diploid and three tetraploid species, almost entirely located east of the Rocky Mountains. The diploids *R. blanda* Ait., *R. foliolosa* Nutt., *R. nitida* Wild., *R. palustris* Marsh., and *R. woodsii* Lindl. (the sole species of the complex also found west of the Rocky Mountains) are relatively well circumscribed (Lewis, 1957c; Erlanson MacFarlane, 1966), but natural interspecific hybrids have been reported (Erlanson, 1929, 1934; Lewis, 1962), and some have been given species status (Rydberg, 1920; Erlanson, 1934). In contrast, the tetraploid taxa *R. arkansana* Porter, *R. carolina* L., and *R. virginiana* Mill. are characterized by extensive continuous morphological variation that blurs their limits with each other and with their putative diploid ancestors in the *R. carolina* complex (Erlanson, 1934; Lewis, 1957b). Despite the important biosystematic investigations involving cytology and morphology in this complex (Erlanson, 1929, 1934; Lewis, 1957b), the limits and origins of the polyploid taxa are still unclear. The broad polymorphism of polyploid species may be caused by hybridization given that it frequently has been reported in areas of contact between *R. carolina* and *R. arkansana* in the west (e.g., *R.* × *rudiuscula* Greene: Erlanson MacFarlane, 1966; Lewis, 1957b; A. Fishbein and W. H. Lewis, Washington University, unpublished manuscript) and between *R. carolina* and *R. virginiana* in the east (Fernald, 1922; Lewis, 1957b) (Fig. 1). Yet, it is also possible that these taxa represent a single polymorphic species rather than three distinct taxa. Therefore, reconstructing the origins of the polyploids is a logical first step toward a global understanding of the *R. carolina* complex because it could be relevant to solving the species status of the polyploids if these are shown to have evolved independently.

Several factors can impair our ability to determine from which species the polyploids evolved and whether they have evolved by autopolyploidy (from a single species) or by allopolyploidy (from more than one species: Grant, 1981; Ramsey and Schemske, 1998). For example, introgression of foreign alleles in an autopolyploid can hide its real origins by making it look like an allopolyploid. Other problems can result from irregularities in chromosome segregation. Allopolyploids

[1] Manuscript received 13 June 2005; revision accepted 23 November 2005.

The authors thank L. Brouillet, E. Dickson, B. Ertter, A. Meilleur, and J. Saarela for providing plant material. Financial help for this study came from research grants (A.B.) and fellowships (S.J. and J.S.) from the National Sciences and Engineering Research Council of Canada and from the Fonds québécois de la recherche sur la nature et les technologies.

[5] Author for correspondence (e-mail: simon.joly@umontreal.ca)

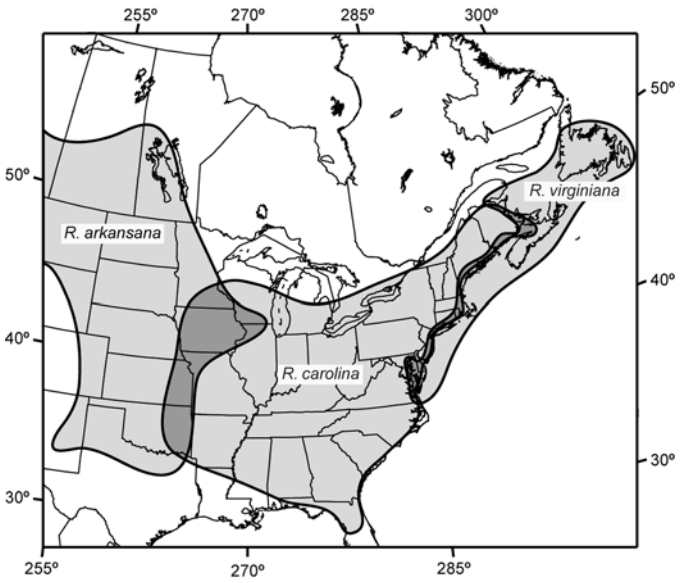

Fig. 1. Approximate distributions of the polyploid taxa *Rosa arkansana*, *R. carolina*, and *R. virginiana*. Areas where species overlap are in dark gray. The distributions are based upon Lewis (1957a) and personal collections.

are expected to have disomic segregation where chromosomes only pair with their homologues (bivalent formation) (Stebbins, 1950, 1971; Levin, 2002), thus guaranteeing the preservation of homologous loci inherited from the parental species. However, these predictions are not always met, and allopolyploids may have occasional polysomic segregation via multivalent formation. This could lead to the fixation of alleles from a single parental species in the genome of the allopolyploid and hide its reticulate origin. The challenge when investigating polyploid evolution is thus to extract the true signal from the noise created by these confounding events in order to adequately reconstruct the evolutionary history of polyploids.

Investigation of polyploid origin must be done within a sound phylogenetic framework. To date, phylogenetic studies of *Rosa* have not included a good sampling of North American roses (e.g., Millan et al., 1996; Matsumoto et al., 1998; Wissemann and Ritz, 2005), leaving their relationships obscure. Reconstruction of the diploid relationships could be further complicated by the recent origin of the complex, which is suggested by the low variation of ribosomal (Ritz et al., 2005) and chloroplastic markers (Wissemann and Ritz, 2005). Recent origin of species may result in incomplete lineage sorting of several molecular markers for the diploids (Pamilo and Nei, 1988; Rosenberg, 2002, 2003), which in turn could hamper our ability to accurately identify the species that were involved in the origins of polyploids. These potential problems need to be addressed prior to investigating polyploid evolution.

A genealogical approach using a single-copy nuclear gene is used to address the relationship of diploids and to investigate the origins of the polyploids. A genealogical approach has major advantages over a genotyping method (e.g., microsatellites, amplified fragment length polymorphisms [ALFPs], isozymes) because it places the data in a historical perspective: it relates who is ancestral to whom rather than who is similar to whom. This is particularly important in order to discern some of the confounding events mentioned earlier from our principal goal—reconstructing polyploid evolution. The use of nuclear genes is particularly useful in this regard because non-haploid organisms (except for clonal and apomict taxa) receive one chromosome copy from each parent. Thus, nuclear genes can retain information about the reticulate history of organisms, which is impossible for maternally or paternally transmitted markers. Such an approach has been successful in reconstructing the polyploid origins of other taxa (Doyle et al., 2002; Senchina et al., 2003; Smedmark et al., 2003; Helfgott and Mason-Gamer, 2004; Joly and Bruneau, 2004; Mason-Gamer, 2004; Petersen and Seberg, 2004; Evans et al., 2005).

MATERIALS AND METHODS

Sampling—Because it was more important to assess the extent of genetic variation within species rather than within populations, a single individual per population was investigated. Populations were sampled to represent the geographical range of each species of the complex (Table 1). Diploid roses of section *Cinnamomeae* west of the Rocky Mountains, *R. gymnocarpa* Nutt. and *R. pisocarpa* Gray, were included because they could be involved in the origins of the eastern polyploids. Diploid roses of section *Synstylae* found in North America, *R. setigera* Michx. (native to North America) and *R. multiflora* Thunb. (introduced from China and now a noxious invasive in eastern North America [Meiners et al., 2001; Hunter and Mattice, 2002]), were included as outgroup taxa. Only one species of *Rosa* section *Cinnamomeae* occurring east of the Rocky Mountains was not investigated here: *R. acicularis* Lindl., a circumboreal species that has both hexaploid and octoploid populations (Lewis, 1959). Investigation of its origin would require a broader taxonomic sampling at the diploid level, which is beyond the scope of the present study.

Ploidy level determination—Lewis (1957b) showed that the length of the stomatal guard cells can discriminate diploid and tetraploid roses of the complex. Twenty-five guard cells per individual were measured for all specimens of eastern species for which we had material. Nail polish was used to fingerprint the abaxial surface of one dried terminal leaflet. The length of stomatal guard cells was measured with a Leitz microscope (type: 307–107.002; Leica Microsystems, Richmond Hill, Ontario, Canada) using a 63× objective. A *K*-mean analysis for two groups was performed to see if diploids and polyploids could be differentiated without previous knowledge of the ploidy level of individuals.

Molecular methods—DNA was extracted using a modified version of the cetyltrimethylammonium bromide (CTAB) extraction of Doyle and Doyle (1987). Modifications involved scaling the protocol for a total CTAB volume of 600 μL; adding 12 μL of 0.5 mol/L ethylenediamine tetra-acetic acid (EDTA) pH 8.0 per 600 μL of CTAB and 1% polyvinylpyrrolidone (PVP) to the extraction buffer prior to extraction; adding 20 μg of RNAse A to the CTAB buffer prior to incubation at 65°C; performing two chloroform-isoamyl alcohol (24:1) extractions and precipitating the DNA with 1.5 volumes of 100% ethanol.

Gene selection—North American roses are particularly uniform at the DNA level. For example, sequences of the internal transcribed spacer of the 18S-5.8S-26S ribosomal gene family showed few variations among North American rose species sampled by Ritz et al. (2005), even though this marker is generally considered to be highly variable in many plant taxa (Baldwin et al., 1995). Similarly, only five variable characters were found between *R. woodsii*, *R. blanda*, and *R. palustris* among 4318 base pairs (bp) from seven chloroplast gene spacers or introns (S. Joly and J. R. Starr, unpublished data). Because of this, introns of single-copy nuclear genes became the alternative for providing sufficient variation. Initial screening (data not shown) of several nuclear genes (*LEAFY* [e.g., Frohlich and Meyerowitz, 1997; Archambault and Bruneau, 2004]; *GBSSI* [e.g., Evans et al., 2000]; *RPB2* [e.g., Denton et al., 1998; Pfeil et al., 2004]; *GAPDH* [e.g., Strand et al., 1997; Olsen and Schaal, 1999]) identified *GAPDH* as the most variable region.

Gene amplification—The cytosolic glyceraldehyde 3-phosphaste dehydrogenase (*GAPDH*) gene was amplified from the end of exon 7 (according to the *Arabidopsis thaliana* sequence; GenBank locus tag: At3g04120) to the beginning of exon 11 (which is exon 9 in *A. thaliana*; Fig. 2). The 5′ end of

TABLE 1. Accessions included in this study of eastern North American roses (*Rosa*). For each accession, voucher information, locality data, the number of alleles at the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) locus found and the number of clones sequenced are indicated. The stomatal guard cell length is given for eastern species.

Species	Accession	Collector(s)	Province/ State[a]	Lat., Long.	Clones sequenced	No. alleles	Guard cell length (μm)
R. arkansana	345	Joly and Starr 601 (MT)	Iowa	43°00′22.7″N, 89°58′44.6″W	12	4	27.72
R. arkansana	406	Joly and Starr 663 (MT)	Minn.	43°43′37.1″N, 95°03′53.1″W	12	2	23.35
R. arkansana	416	Joly and Starr 673 (MT)	Minn.	46°33′27.5″N, 96°13′07.9″W	12	4	22.40
R. arkansana	470	Joly and Starr 730 (MT)	Man.	50°04′55.7″N, 97°08′17.1″W	13	4	23.65
R. arkansana	503	Joly and Starr 763 (MT)	N.Dak.	47°58′09.9″N, 97°46′35.3″W	19	4	23.16
R. arkansana	665	Lewis 15837-1 (MO)	Kansas	38°32′N, 94°55′W	11	3	—
R. arkansana	692	Lewis 15792-2 (MO)	Kansas	—	14	4	23.10
R. arkansana	848	Ryan 3 (MT)	Sask.	52°05.479″N, 106°41.857″W	11	4	26.78
R. blanda	160	Joly and Starr 409 (MT)	N.B.	45°57′43.7″N, 67°22′26.1″W	3	2	14.65
R. blanda	326	Joly and Starr 582 (MT)	Ont.	42°15′29.7″N, 83°02′58.8″W	3	2	16.15
R. blanda	365	Joly and Starr 622 (MT)	Wis.	42°39′07.5″N, 89°43′32.4″W	4	2	15.99
R. blanda	421	Joly and Starr 678 (MT)	Minn.	48°06′36.3″N, 96°09′16.0″W	4	2	14.71
R. blanda	462	Joly and Starr 722 (MT)	Man.	50°00′59.3″N, 96°55′35.2″W	4	2	17.54
R. blanda	528	Joly and Starr 788 (MT)	Ont.	46°28′15.4″N, 80°29′27.2″W	—	1	16.96
R. blanda	567	Joly 921 (MT)	N.Y.	—	4	2	—
R. blanda	621	Joly et al. 962 (MT)	N.B.	47°22′32.5″N, 66°04′30.5″W	4	2	15.63
R. blanda	652	Joly et al. 993 (MT)	Que.	48°02′58.8″N, 65°28′43.6″W	4	2	14.87
R. blanda	1214	Bruneau et al. 1214 (MT)	Que.	45°31′N, 73°50′W [b]	—	1	16.60
R. blanda	1219	Bruneau et al. 1219 (MT)	Que.	45°30′N, 73°50′W [b]	—	1	16.14
R. blanda	1236	Bruneau et al. 1236 (MT)	Que.	48°21′N, 68°45′W [b]	4	2	16.26
R. blanda	1239	Bruneau et al. 1239 (MT)	Que.	48°21′N, 68°45′W [b]	4	2	17.02
R. blanda	98016	Drouin 98-016 (MT)	Que.	47°26′N, 70°30′W [b]	—	2	15.44
R. carolina	268	Joly and Starr 523 (MT)	Va.	38°21′29.8″N, 79°04′54.1″W	13	4	27.54
R. carolina	289	Joly and Starr 545 (MT)	W.Va.	38°41′N, 80°00′W [b]	10	4	22.07
R. carolina	320	Joly and Starr 576 (MT)	Ont.	42°15′29.7″N, 83°02′58.8″W	13	4	27.97
R. carolina	395	Joly and Starr 651 (MT)	Minn.	43°48′03.6″N, 92°29′21.6″W	13	4	23.16
R. carolina	553	Lewis 15783-3 (MO)	Mo.	38°31′23″N, 90°40′36″W	14	4	22.80
R. carolina	576	Joly 906 (MT)	N.Y.	—	12	4	—
R. carolina	626	Joly et al. 967 (MT)	N.B.	47°22′36.2″N, 66°04′42.2″W	11	2	19.30
R. carolina	671	Lewis 15844 (MO)	Okla.	36°92′N, 94°88′W	10	4	24.47
R. foliolosa	699	Lewis 15846-3 (MO)	Okla.	34°24′N, 96°00′W	3	2	19.18
R. foliolosa	795	O′Kennon and McLemore 19069A (MT)	Tex.	33°24′32.2″N, 97°30′22.0″W	—	2	17.02
R. gymnocarpa	543	Ertter 18001 (JEPS)	Idaho	—	—	1	—
R. gymnocarpa	751	Lewis 15852-1 (MO)	B.C.	49°02′N, 118°13′W	3	2	—
R. gymnocarpa	767	Ertter 18293a (JEPS)	Idaho	—	—	1	—
R. multiflora	302	Joly and Starr 558 (MT)	Pa.	42°08′48.4″N, 80°08′00.1″W	4	2	—
R. nitida	570	Meilleur s.n. (MT)	Que.	—	4	2	14.68
R. nitida	604	Joly et al. 941 (MT)	N.B.	45°56′29.2″N, 64°52′07.3″W	3	2	17.97
R. nitida	675	Brouillet 03-55-1 (MT)	Nfld.	—	—	2	19.09
R. nitida	812	Joly 1010-1 (MT)	Que.	46°22′45.3″N, 75°00′20.6″W	4	2	17.05
R. palustris	168	Joly and Starr 417 (MT)	N.B.	45°33′43.2″N, 67°25′31.2″W	4	2	15.63
R. palustris	255	Joly and Starr 510 (MT)	N.J.	38°56′02.8″N, 74°57′29.5″W	4	2	15.96
R. palustris	304	Joly and Starr 560 (MT)	Pa.	42°09′32.9″N, 80°07′10.7″W	4	2	16.81
R. palustris	317	Joly and Starr 573 (MT)	Ont.	42°19′41.0″N, 82°18′49.0″W	4	2	17.24
R. palustris	331	Joly and Starr 587 (MT)	Mich.	42°19′32.0″N, 84°29′51.2″W	—	1	16.14
R. palustris	386	Joly and Starr 644 (MT)	Wis.	44°01′30.6″N, 89°43′13.1″W	—	1	17.12
R. palustris	581	Joly 912 (MT)	N.Y.	—	—	1	—
R. pisocarpa	774	Ertter 18303a (JEPS)	Calif.	—	4	2	—
R. pisocarpa	784	Ertter 18305c (JEPS)	Oreg.	42°05.7′N, 123°41.0′W	4	2	—
R. pisocarpa	847	Ertter 18428 (JEPS)	Calif.	41°09.2′N, 123°49.2′W	4	2	—
R. setigera	298	Joly and Starr 554 (MT)	Pa.	42°08′48.4″N, 80°08′00.1″W	—	1	—
R. virginiana	182	Joly and Starr 431 (MT)	N.B.	45°05′00.4″N, 67°03′01.1″W	14	3	21.92
R. virginiana	195	Joly and Starr 444 (MT)	Maine	44°30′56.7″N, 68°11′14.6″W	10	4	21.83
R. virginiana	225	Joly and Starr 474 (MT)	Conn.	41°20′43.0″N, 71°54′14.2″W	13	3	23.29
R. virginiana	246	Joly and Starr 496 (MT)	N.J.	38°55′57.4″N, 74°57′28.5″W	12	4	20.16
R. virginiana	262	Joly and Starr 517 (MT)	Md.	38°14′08.2″N, 75°08′15.7″W	13	4	21.49
R. virginiana	587	Joly and Edelist 924 (MT)	N.S.	45°43′09.7″N, 61°53′56.3″W	11	4	23.44
R. virginiana	656	Joly et al. 997 (MT)	Que.	48°02′58.8″N, 65°28′43.6″W	10	3	21.16
R. virginiana	684	Brouillet 03-60-1 (MT)	Nfld.	—	14	3	23.68
R. woodsii	4	Spellenberg 12555 (MT)	N.Mex	—	—	1	16.54
R. woodsii	492	Joly and Starr 752 (MT)	Sask.	49°12′35.3″N, 101°50′46.1″W	—	1	15.20
R. woodsii	498	Joly and Starr 758 (MT)	N.Dak.	48°21′09.6″N, 99°47′07.5″W	—	2	18.00
R. woodsii	700	Saarela 266-1 (MT)	Alta.	—	—	2	—
R. woodsii	733	Dickson 2017 (MT)	Alta.	—	—	2	—
R. woodsii	741	Lewis 15848-1 (MO)	B.C.	49°45′N, 120°50′W	3	2	—
R. woodsii	800	Joly 1005-1 (MT)	Colo.	40°12′23.4″N, 104°49′54.0″W	—	1	14.59
R. woodsii	807	Joly 1008-1 (MT)	Colo.	40°38′36.8″N, 104°20′32.0″W	—	1	15.72

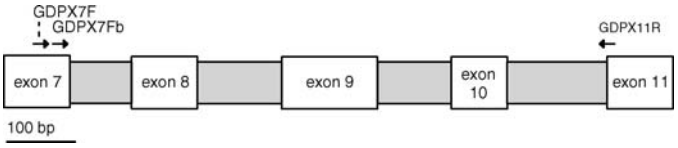

Fig. 2. The glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) locus in North American *Rosa*. Primers are not to scale and their positions are approximate. The first exon (7) is numbered according to *Arabidopsis thaliana*, but the amplified region between primers GDPX7F and GDPX11R contains two introns not present in *A. thaliana*.

the forward primer GPDX7F (5′-GATAGATTTGGAATTGTTGAGG-3′; Strand et al., 1997) starts 52 bp upstream of the intron in the seventh exon, whereas the GPDX11R primer (5′-GACattgaatgagataaacc-3′; lowercase letters represent intron nucleotides) spans the junction between exon 11 and the previous intron. Polymerase chain reactions in final volumes of 50 μL contained 1× PCR reaction buffer (Roche Diagnostics, Laval, Québec, Canada; for a total $MgCl_2$ concentration of 1.5 mmol/L), 0.05% Tween 20, 5 μg bovine serum albumin, 1 mmol/L of each primer, 200 μmol/L of each dNTP, two units *Taq* polymerase, and ca. 300 ng genomic DNA. The PCR conditions included an initial denaturation step of 3 min at 94°C, followed by 40 cycles of denaturation (30 s at 95°C), annealing (30 s at 48°C), and elongation (2 min at 72°C), with a final extension step of 10 min at 72°C. A long elongation time was used, and reactions were performed in triplicate to reduce the potential for PCR recombinants (Judo et al., 1998; Cronn et al., 2002). The triplicate reactions also reduced the possibility of finding the same *Taq*-induced mutation in many different clones. The PCR products were purified with polyethylene glycol (PEG; molecular mass, 8000) according to the following procedure. The PCR reactions were mixed with an equal volume of PEG solution (20% PEG, 2.5 mol/L NaCl), incubated 15 min at 37°C, and centrifuged 15 min at 12 000 × g. The supernatant was removed and the pellet was washed twice with 80% ethanol (spinning 5 min at 12 000 × g before ethanol removal). The pellet was dried 2 min in a vacuum centrifuge (no heat) and was resuspended in $TE_{0.1}$ (20 mmol/L Tris-HCl, 0.1 mmol/L EDTA, pH 8.0).

Sequencing was performed using the reverse primer GPDX11R and the forward primer GPDX7Fb (5′-cttatgactaccgtgcactc-3′; Fig. 2). The 5′ end of GPDX7Fb is located 28 bp upstream of the intron in exon 7. Sequencing reactions were performed with BigDye terminator chemistry (version 1.1; Applied Biosystems, Foster City, California, USA) following the manufacturer's protocols and were run on a 3100-*avant* automated sequencer (Applied Biosystems). Sequences were assembled and edited in Sequencher (version 4.1; GeneCodes, Ann Arbor, Michigan, USA).

Allele sampling—In order to derive firm conclusions on the origins of polyploids, it is important to sample all alleles in every individual. The approach used to achieve this objective differed for diploids and polyploids. Diploids that did not show polymorphic nucleotides in direct sequencing (from the total PCR reaction) were assumed to be homozygous and were not cloned. Such an assumption is valid because two equally frequent templates should be equally visible on chromatograms if there is no strong PCR bias in the reactions (Rauscher et al., 2002). When only one polymorphic nucleotide was found for one individual, no cloning was necessary because the alleles can easily be distinguished. In contrast, individuals that showed more than one polymorphic site or that had indels among its alleles were cloned. In these cases, 3–4 clones were sequenced to retrieve allelic sequences. More than one clone was sequenced to eliminate the possibility of sampling a PCR recombinant with a single clone.

All tetraploids were cloned because it is easier to miss polymorphic sites on direct sequences when four alleles may be present in the genome. Assuming no PCR bias between alleles (but see Wagner et al., 1994), the binomial distribution predicts that the probability of sampling all alleles in an individual is

$$P = \left[1 - \left(\frac{t-1}{t}\right)^n\right]^t,$$

where t is the number of alleles in the individual and n is the number of clones sequenced. If there were four alleles in a tetraploid, 15 clones would be required in order to obtain a 95% probability that all alleles have been sampled. With three alleles, 11 clones are needed. On average, 11–15 clones were sequenced per individual (Table 1), with additional clones sequenced in all cases where the alleles resulting in polymorphisms detected in direct sequencing were not recovered.

For both diploids and tetraploids, *Taq*-induced PCR errors were identified and removed from analyses by comparing the sequence of cloned amplicons to one another and to the initial sequences obtained from direct sequencing. Henceforth, it will be assumed that all alleles were retrieved from each individual even if there is a non-zero probability that some alleles were not sampled in some individuals. The PCR products were cloned with the TOPO TA cloning kit (Invitrogen, Burlington, Ontario, Canada). Plasmids containing the gene were extracted from *E. coli* using the QIAprep miniprep kit (Qiagen, Mississauga, Ontario, Canada) and were sequenced as described earlier. Alleles from both diploids and tetraploids were aligned with ClustalX (Thompson et al., 1994, 1997) with a gap opening penalty of 25 and a gap extension penalty of 6. The resulting alignment did not need further manual corrections.

Testing recombination—Two different methods were used to detect recombination: the homoplasy test (Maynard Smith and Smith, 1998), which works best when divergence between sequences is low (less than 5%; Maynard Smith and Smith, 1998; Posada and Crandall, 2001), and a parsimony network approach (Templeton et al., 1992). The homoplasy test was performed using datin and exph programs (Maynard Smith and Smith, 1998) under conservative ($S_E = 0.6S$) and liberal ($S_E = S$) conditions, where S_E is the effective number of sites and S is the total number of sites in the data set. First and second codon positions in exons were removed from the analysis because they are evolutionarily constrained (Maynard Smith and Smith, 1998), and the analysis was performed only on ingroup taxa. With the parsimony network approach, recombination was inferred only when it could explain at least two homoplasies and when the homoplasies corresponding to the parental alleles were physically clustered on the recombinant allele (Aquadro et al., 1986; Templeton et al., 1992).

Network construction—GapCoder (Young and Healy, 2003) was used to code indels under the simple gap coding method of Simmons and Ochoterena (2000). The resulting matrix was used to estimate the gene genealogy of the *GAPDH* locus by statistical parsimony (Templeton et al., 1992) as implemented in the TCS program (version 1.18; Clement et al., 2000). The statistical limit of parsimony was evaluated on the matrix with the gaps recoded (although estimating it without the gaps gave the same result), and the final network was constructed so that all the haplotypes could be united in a single network.

Statistical distinction of diploid species—Diploid species boundaries were tested by permutations using an analysis of molecular variance (AMOVAs; Excoffier et al., 1992). An uncorrected P distance matrix among haplotypes was calculated in PAUP* version 4.10b (Swofford, 2002), and the partitioning of haplotype variance in different groups (species) was tested in Arlequin version 2 (10 000 permutations; Schneider et al., 2000).

Origins of the polyploids—To reconstruct the evolutionary history of the polyploid taxa, the closest diploid haplotype ancestor for each allele of each polyploid individual was identified to determine which diploid species contributed to polyploids. Because alleles can mutate in polyploids, simply counting the number of haplotypes in a polyploid species will overestimate the number of origins (Doyle et al., 2004). A conservative way of evaluating the likelihood that the polyploid species evolved recurrently is to estimate the number of "polyploid haplotype groups" that comprise all polyploid haplotypes that have a most recent common diploid haplotype (or expected diploid haplotype) ancestor (Fig. 3; see also Doyle et al., 2004). At formation, a tetraploid can acquire up to four different alleles from diploids. Independent

←

[a] Abbreviations for states and provinces follow the nomenclature of Flora of North America (Flora of North America Editorial Committee, 1993): Alta. = Alberta, B.C. = British Columbia, Calif. = California, Colo. = Colorado, Conn. = Connecticut, Man. = Manitoba, Md. = Maryland, Mich. = Michigan, Minn. = Minnesota, Mo. = Missouri, N.B. = New Brunswick, Nfld. = Newfoundland, N.J. = New Jersey, N.Mex. = New Mexico, N.Y. = New York, N.Dak. = North Dakota, N.S. = Nova Scotia, Okla. = Oklahoma, Ont. = Ontario, Oreg. = Oregon, Pa. = Pennsylvania, Que. = Québec, Sask. = Saskatchewan, Tex. = Texas, Va. = Virginia, W.Va. = West Virginia, Wis. = Wisconsin.

[b] Approximate coordinates not determined by GPS.

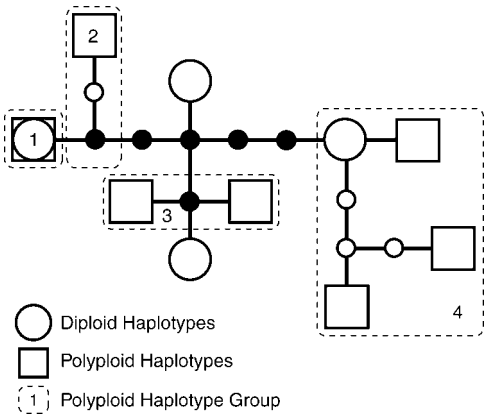

Fig. 3. Illustration of a network used to evaluate the minimum number of haplotypes contributed by diploids to polyploids when estimating the number of independent origins of a polyploid species. Large open circles (diploids) and squares (polyploids) represent sampled haplotypes. Small filled circles represent unsampled interior haplotypes inferred to have occurred in diploids, whereas small open circles represent interior unsampled haplotypes that cannot be inferred to have occurred in the diploids. Polyploid haplotype groups (dashed lines) are comprised of polyploid haplotypes that have a most recent common diploid or expected diploid haplotype ancestor.

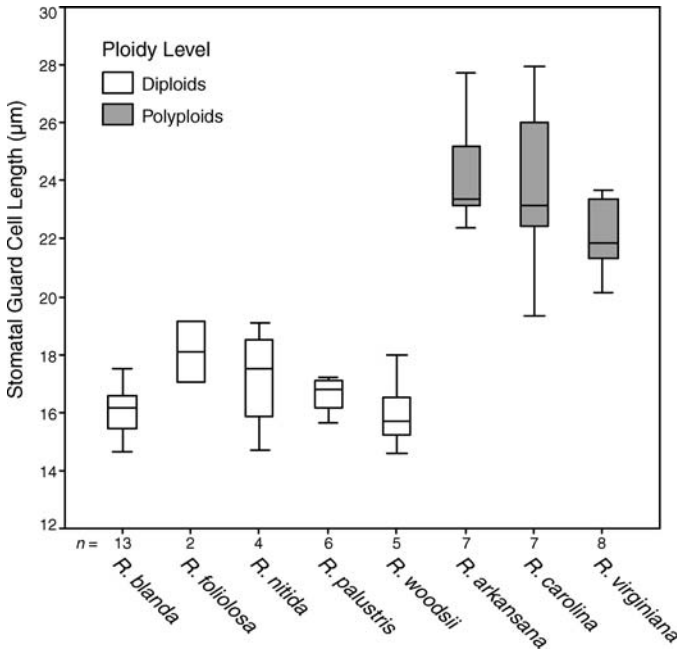

Fig. 4. Boxplots of length variation of the guard cells for each species of *Rosa*, contrasting results from diploids and polyploids. The boxplots were constructed using the mean length per individual, and the number of individuals assessed per species is indicated (*n*). The mean length of guard cells for each individual can be found in Table 1. Each boxplot shows the median, the interquartile range and the range of all observations.

polyploid origins can involve one or more identical diploid alleles, yet it is impossible to detect this if there is segregation in polyploid populations. To be conservative, it was therefore assumed that for one polyploid species, a polyploid haplotype group can only be involved in one origin and that each origin always involved four polyploid haplotype groups. So if there are *n* polyploid haplotype groups in one polyploid species ($n = 4$ in Fig. 3), there needs to be at least $n/4$ (rounded to the upper unit) distinct origins to account for this variability (one distinct origin in the simplified example given in Fig. 3).

RESULTS

Sequences and alleles—The number of alleles found and the number of clones sequenced for each individual is indicated in Table 1. The phylogenetic analysis used the portion of the *GAPDH* gene that starts immediately after exon 7 and that stops at the GPDX11R primer, 17 bp downstream of exon 11. The length of this aligned region is 759 bp and it includes 15 indels. Multiple alleles in an individual were distinguished by a letter (i.e., A, B, etc.) following the species name and accession number. GenBank accession numbers (DQ091014–DQ091057, DQ091060–DQ091174) are given for each allele of each individual in Appendix S1 (see Supplemental Data accompanying the online version of this article).

Of all alleles recovered, one was obviously a pseudogene: the carolina289.A allele. This allele has a deletion of 1 bp in exon 10, that causes a frame shift and introduces a stop codon. Because the indel was visible in the direct sequences, and therefore present in relatively high proportions in the PCR products (Rauscher et al., 2002) and because the reactions were performed in triplicate, it is unlikely that this mutation is the result of a PCR error.

Length of stomatal guard cells—Based on the taxonomic identifications, diploids and polyploids had disjoint distributions for their mean stomatal guard cell length (Fig. 4), and the difference between the two groups is statistically significant (two-way Student *t* test: $\nu = 50$, $t = -14.061$, $P < 0.001$; homoscedasticity hypothesis accepted: Levene $F = 3.949$, $P = 0.53$). The mean lengths of diploids and polyploids were under 19.18 µm and over 19.30 µm, respectively (Fig. 4, Table 1). The gap is more important when making abstraction of the *carolina* 626 individual, without which all polyploids would have a mean length over 20.16 µm.

The mean lengths of the two clusters recovered by a *K*-mean analysis were 16.60 µm and 23.75 µm. Only two assignments (of 52) disagreed with taxonomic identifications: individuals *carolina* 626 (19.30 µm) and *virginiana* 246 (20.16 µm) fell in the shorter cluster otherwise constituted of only diploid species. To confirm the ploidy level of these individuals, it is helpful to consider the number of alleles found. For example, *virginiana* 246 has four alleles (Table 1), which is strong evidence of polyploidy, and it will hereafter be treated as a polyploid. In contrast, *carolina* 626 only has two alleles, which is inconclusive as to its ploidy level. This latter individual will be treated as a polyploid based on its morphology and on its stomatal guard cell length that is longer than that of any diploid (Table 1).

The stomatal cell lengths reported are about 1.3 times smaller than those obtained by Lewis (1957b, 1958, 1959) for both diploids and polyploids. These discrepancies are simply caused by differences in methodology.

Network—One of the premises of tree-like phylogenetic methods is that all characters have the same evolutionary history. Recombination can violate this assumption for nuclear loci, and it is important to test for its presence when using such markers. The homoplasy test was significant under both the conservative and liberal conditions ($P < 0.001$), suggesting that recombination is present in the data set. In contrast, no clear recombinants were detected using the network approach.

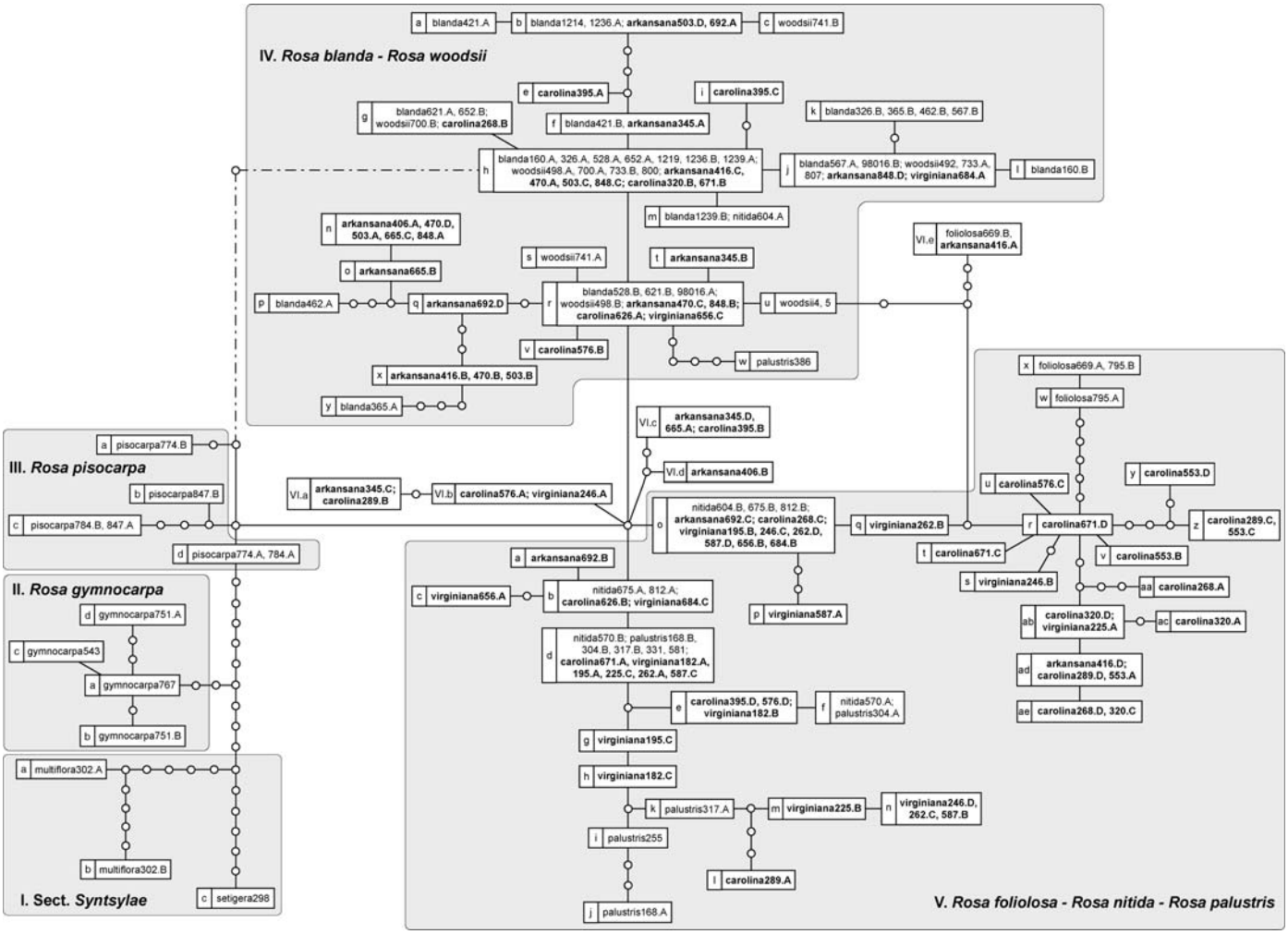

Fig. 5. Network obtained by statistical parsimony analysis of the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene. The limit of parsimony is 12 steps. Individuals within species are identified by their specific epithet and accession number, and letters following accession numbers are used to differentiate multiple alleles within an individual. Diploids are in lightface type, whereas polyploids appear in boldface type. Each box represents a haplotype; small open circles between boxes represent unsampled (i.e., inferred) haplotypes. The shaded boxes represent the principal diploid groups that are discussed in the text. The broken line indicates an alternative branching scenario that is less likely for *Rosa pisocarpa* (see section Results, subsection Diploids).

Even within the loops, there was always one alternative that required only one homoplasy. The discrepancy between these results could be due to the presence of homoplasious sites in the data set: a standard parsimony analysis gave a consistency index of 0.83. Even if allelic variation ranges from 0 to 3.4% of variation among ingroup taxa, this level of homoplasy may be high enough to violate the homoplasy test's assumption of low levels of variation, which could bias the test towards a conclusion for recombination. Such behavior of the homoplasy test has previously been reported (Posada and Crandall, 2001; Posada, 2002). Because no clear recombination events were identified on the network, the evidence for recombination in the data is equivocal at best and the data set was analyzed as if there were no recombination.

Haplotypes with a distance of more than 12 steps (parsimony limit) from all other haplotypes were not statistically supported and their relationship to the rest of the haplotypes should be viewed as if estimated by standard parsimony procedures (Fig. 5). However, only section *Synstylae* was not connected to the rest of the network with this limit; the two sub-networks were 13 steps away. Henceforth, haplotypes will be referred to by the number of the box in which they occur on the network and by their specific letter (e.g., I-a represents the haplotype of allele multiflora302.A of section *Synstylae*; Fig. 5).

Diploids—Relative to the outgroup species *R. multiflora* and *R. setigera* (section *Synstylae*), alleles of *R. gymnocarpa* are monophyletic. The other western species, *R. pisocarpa*, is either paraphyletic or polyphyletic depending upon how the loop involving the *R. pisocarpa* haplotypes is resolved on the network (Fig. 5). In the presence of ambiguity, one hypothesis can be favored over others because a loop is more likely to be broken beside the most recent haplotype of the loop (Crandall and Templeton, 1993). Coalescent theory predicts that an old haplotype is more frequent (Donnelly and Tavaré, 1986) and that more lineages are related to it (Crandall and Templeton,

1993; Castelloe and Templeton, 1994). These predictions can be used to determine "outgroup weights" that are correlated to the age of the haplotype (Castelloe and Templeton, 1994). According to these outgroup weights and the predictions of Crandall and Templeton (1993), the most likely hypothesis is the one that links the *R. pisocarpa* III-a haplotype to the other *R. pisocarpa* haplotypes (solid line on the network; the alternative solution is shown by a broken line; Fig. 5). Consequently, the paraphyletic option for *R. pisocarpa* is more likely than the polyphyletic one. This also suggests a division between western (boxes II, III) and eastern (IV, V, VI) diploid species of section *Cinnamomeae* (Fig. 5).

Regarding the diploid species east of the Rocky Mountains, two main groups can be distinguished on the network (Fig. 5). The first group includes all alleles of diploid species *R. blanda* and *R. woodsii* (the *blanda–woodsii* or BW group, box IV in Fig. 5), whereas the other contains most alleles of *R. foliolosa*, *R. nitida*, and *R. palustris* (the *foliolosa–nitida–palustris* or FNP group, box V). These groups are not monophyletic, but they are nevertheless almost exclusive. There are two exceptions: one allele of *R. palustris* and one of *R. nitida* occur in the BW group. Even with these, the AMOVAs showed that the distinction between the BW and the FNP groups is significant ($P < 0.001$; Table 2). Neither the AMOVAs nor the network found a distinction between *R. blanda* and *R. woodsii*. Within the FNP group, AMOVAs suggest that *R. foliolosa* is significantly distinct from *R. nitida* and *R. palustris* ($P < 0.001$) and also that the differentiation between *R. nitida* and *R. palustris* is marginally significant ($P < 0.05$; Table 2). The network is ambiguous regarding these distinctions, however, and *R. nitida* and *R. palustris* do not clearly form distinct groups (Fig. 5). Moreover, only two individuals of *R. foliolosa* were investigated, limiting the significance of the distinction found with AMOVAs. In addition, the *R. foliolosa* alleles have *R. nitida* alleles as ancestors. Therefore, *R. foliolosa*, *R. nitida*, and *R. palustris* are considered to form a single group in the following analyses.

Polyploids—Polyploid haplotypes on the network are exclusively related to eastern diploids. Most polyploid alleles can be clearly attributed to either the BW or the FNP diploid groups, and only a limited number of alleles have an ambiguous relationship (those that could not be placed in either group; VI, a–e in Fig. 5). All polyploid species have haplotypes that belong to both the FNP and the BW group (Fig. 5, 6), but not all individuals of each species have alleles from both diploid groups. All eight *R. arkansana* individuals studied have alleles that belong to the BW diploid group (Fig. 6a). Three of them have exclusively such alleles, two also have one allele that has an ambiguous relationship, and three have one allele from the FNP group. Almost all eight *R. carolina* individuals have alleles that are from both the BW and FNP diploid groups (Fig. 6b). There are only two exceptions, and one of these has an allele of unknown relationship. Finally, five individuals of *R. virginiana* have exclusively FNP-related alleles, two have haplotypes related to both eastern diploid groups, and one has haplotypes from the FNP group and of unknown origin (Fig. 6c).

The number of polyploid haplotype groups was 13, 11, and 12 in *R. arkansana*, *R. carolina*, and *R. virginiana*, respectively (Appendix S2-S4, see Supplemental Data accompanying the online version of this article for the circumscription of these haplotype groups for each polyploid). This requires a minimum of three distinct polyploid origins to explain the observed genetic diversity in all three polyploid taxa.

DISCUSSION

Diploid species boundaries—Three evolutionary processes can result in nonmonophyletic species in a genealogical framework: hybridization, incomplete lineage sorting (or deep coalescence), and gene duplication (Maddison, 1997; Funk and Omland, 2003). Among these processes, gene duplication is the least likely problem at low phylogenetic levels. Because no evidence of gene duplication was found, this process will not be discussed further.

Incomplete lineage sorting and hybridization—Attempts have been made to distinguish between incongruence due to incomplete lineage sorting and incongruence due to hybridization in gene trees (Sang and Zhong, 2000), but these mostly have been unfruitful (Holder et al., 2001). However, it is possible, in some circumstances, to discriminate between the two processes by using the full amount of information contained in branch lengths (Holder et al., 2001). Take a hypothetical example of a lineage that splits into two distinct species at time T_S, where one incongruent haplotype happens to be more closely related to the haplotypes of its sister species than it is to its own haplotypes (Fig. 7). Note that the time of speciation is independent of the gene lineages and corresponds to the time when gene flow ceased among sibling species (see Holder et al., 2001). With incomplete lineage sorting, the most recent common ancestor of the incongruent haplotype and the haplotypes of the sister species must have been present in the common lineage before the speciation event (Fig. 7A). Therefore, the time since the divergence of the incongruent allele and the alleles of the sister species (T_{LS}) must be at least as old as the time of divergence of the two species ($T_{LS} \geq T_S$). On a hypothetic genealogy, the incongruent allele should branch near the split between the two species relative to an outgroup taxon, and it should be quite divergent from the alleles of the sister species because it has evolved independently from the other sister species alleles for a time T_{LS} (Fig. 7B).

In contrast, the time of divergence between an incongruent haplotype caused by hybridization and haplotypes of its sister species (T_H) can be younger that the speciation event (e.g., Fig. 7C), which would result in an incongruent allele connected on the network far from the root and similar to the contemporary alleles of its sister species (Fig. 7D). However, because the incongruent allele could also coalesce with alleles of the other species before the speciation event, hybridization could result in a pattern identical to that expected from incomplete lineage sorting (e.g., Fig. 7B). Therefore, it should be possible to identify an hybridization event when the pattern observed is similar to the one in Fig. 7D, but in the presence of a pattern such as that of Fig. 7B, it is impossible to discriminate between both hypotheses (see also Holder et al., 2001).

The *GAPDH* haplotype network may give us examples of both hybridization and incomplete lineage sorting between the *blanda–woodsii* and the *foliolosa–nitida–palustris* diploid groups. First, a hybridization event is probably the cause of the position of the nitida604.A allele (haplotype IV-m) in the *blanda–woodsii* group (Fig. 5). The hybridization hypothesis is supported because the haplotype connects to the network three steps away from the node separating the two diploid groups on

TABLE 2. Partition of variance (AMOVAs) within and among different species or groups of species of *Rosa*.

Groups tested	df	Percentage of variance
R. blanda, R. woodsii vs. *R. nitida, R. palustris* vs. *R. foliolosa*		
Among groups	2	43.26***
Within group	57	56.74
R. blanda, R. woodsii vs.. *R. nitida, R. palustris, R. foliolosa*		
Among groups	1	32.66***
Within group	58	67.34
R. blanda, R. woodsii vs. *R. nitida, R. palustris*		
Between groups	1	37.14***
Within group	54	62.86
R. blanda, R. woodsii vs. *R. foliolosa*		
Between groups	1	60.06***
Within group	40	39.94
R. nitida, R. palustris vs. *R. foliolosa*		
Between groups	1	49.54***
Within group	20	50.46
R. blanda vs. *R. woodsii*		
Between groups	1	-3.06^{ns}
Within group	36	103.06
R. nitida vs. *R. palustris*		
Between groups	1	11.45*
Within group	16	88.55

Note: *** P (observed value $\geq$ random value) ≤ 0.0001; * P (obs. $\geq$ rand.) ≤ 0.05; ns, not significant.

the network and also because it is found in a contemporary *R. blanda* individual. This shows that the divergence between the incongruent haplotype IV-m and the other species' allele is recent relative to the separation between the two diploid groups. The other incongruent allele, palustris386 (haplotype IV-w), is more likely to be caused by incomplete lineage sorting because it diverges from its ancestor one step away from the node delimiting the two diploid groups on the network (i.e., the split is relatively old) and because it is five steps away from the closest contemporary alleles of the *blanda–woodsii* group, which is plausible if it has evolved independently from these alleles for some time. As discussed above, however, it is impossible to completely reject the hypothesis of hybridization for this incongruence. It is also plausible that contemporary *blanda–woodsii* alleles closer to this allele exist but were not sampled.

Testing species boundaries—Hybridization is more frequent among closely related species. The same is true of incomplete lineage sorting, which is particularly important for nuclear genes because their effective population sizes are greater than for chloroplast or mitochondrial genes (Moore, 1995; Wollenberg and Avise, 1999; Rosenberg, 2003). If we consider that species are ecologically, morphologically, and (or) genetically cohesive groups of populations that evolve independently from other such groups, then nuclear genes may fail to identify recently derived species if a criterion of monophyly (e.g., the genealogical species concept, Baum and Shaw [1995]; the monophyletic species concept, Mishler and Theriot [2000], Wheeler and Platnick [2000]) is applied (Hudson and Coyne, 2002). Templeton (2001) has proposed using nested clade analysis as a way to test "cohesive" species boundaries (i.e., Templeton, 1989), therefore allowing some incongruence between the species tree and the gene tree. Unfortunately, this method requires extensive population sampling, which is a laborious task for single-copy nuclear genes because of the extensive cloning effort necessary to properly sample alleles. As an alternative, AMOVAs were used to evaluate the genetic variation due to within-species (or groups of species) variation as compared to among-species variation and to test if this latter variance is greater than that expected by chance. This method also allows some alleles to be incongruent with the species tree.

The network suggests that *R. gymnocarpa* is sister to all other North American *Rosa* species of section *Cinnamomeae*. The distinctiveness of this species has already been reported based on morphological characters (Watson, 1885; Crépin, 1896), but its phylogenetic position was uncertain. *Rosa pisocarpa*, although non monophyletic, is distinct from diploid species of the *R. carolina* complex on the network, and its position suggests that eastern diploid species are monophyletic.

Among the largely eastern taxa of the complex, AMOVAs identified two major groups of diploids: *blanda–woodsii* and *foliolosa–nitida–palustris*. This shows that the incongruence found among groups (and discussed earlier) is not significant and that these groups could be considered as distinct. In the *blanda–woodsii* group, no distinction was found between *R. blanda* and *R. woodsii*. Indeed, these species cannot be distinguished using morphological and molecular (AFLP) characters (J. R. Starr, S. Joly, and A. Bruneau, unpublished data). Moreover, hybrids between *R. blanda* and *R. woodsii* have been shown to be highly fertile (Erlanson, 1934; Ratsek et al., 939), and a hybrid zone appears to exist in the area where the two species overlap (Lewis, 1962). Given this, the status of these species certainly needs to be addressed. In the *foliolosa–nitida–palustris* group, analyses of molecular variance suggested that *R. foliolosa* was distinct, although no strong conclusions regarding this species are drawn because of limited sampling. Yet, the distinction of *R. foliolosa* from other eastern diploid species is supported by morphology, this species being peculiar for its narrow leaflets and short pedicels, among other characters (Lewis, 1957b, 1958). The AMOVAs also suggest a weak distinction between *R. nitida* and *R. palustris* even if the network clearly shows that they do not form distinct groups. The species status for these two taxa is different from that of *R. blanda* and *R. woodsii* because they are clearly distinct morphologically (Lewis, 1957a, b). *Rosa nitida* has numerous red bristles, is generally less than 1 m tall, and has no distinct infrastipular thorns, whereas *R. palustris* lacks bristles, is greater than 1 m tall, and almost always has curved infrastipular thorns. Therefore, the absence of reciprocal monophyly between *R. nitida* and *R. palustris* for the *GAPDH* marker may be a consequence of their recent divergence.

Origins of the polyploids—The identification of genetically distinct groups of diploids in section *Cinnamomeae* in North America allows the evaluation of different evolutionary hypotheses concerning the origin of the polyploids. Yet it can be difficult to determine whether a polyploid is an autopolyploid or an allopolyploid in the event of conflicting signals produced by hybridization among polyploid species, gene flow between diploids and polyploids, or allelic segregation in polyploids. Both homoploid hybridization among polyploid species and gene flow from diploids to polyploids can introduce haplotypes in a polyploid that were

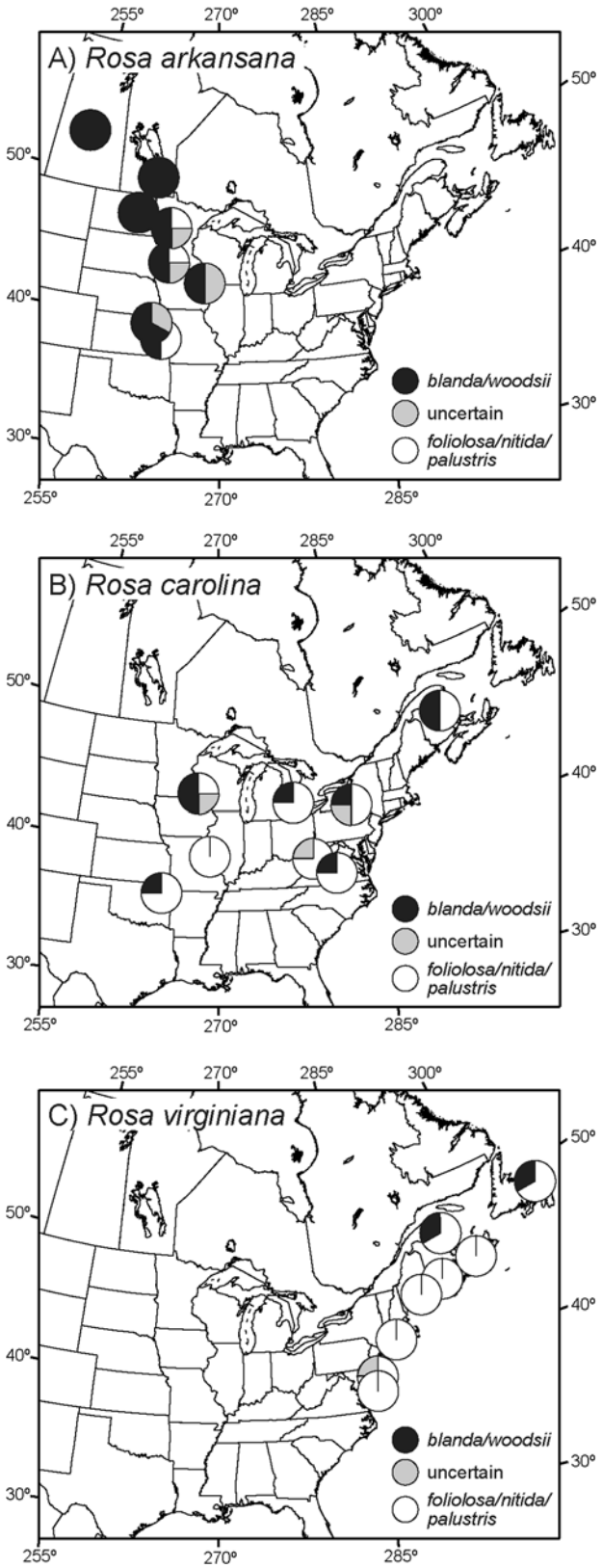

Fig. 6. Genetic constitutions of individuals sampled for the tetraploids, (A) *Rosa arkansana*, (B) *R. carolina*, and (C) *R. virginiana*. The genetic constitution of each individual is represented by a pie chart; shades of gray represent the proportion of alleles from each of the diploid groups (Fig. 5). Total number of alleles for each accession is given in Table 1.

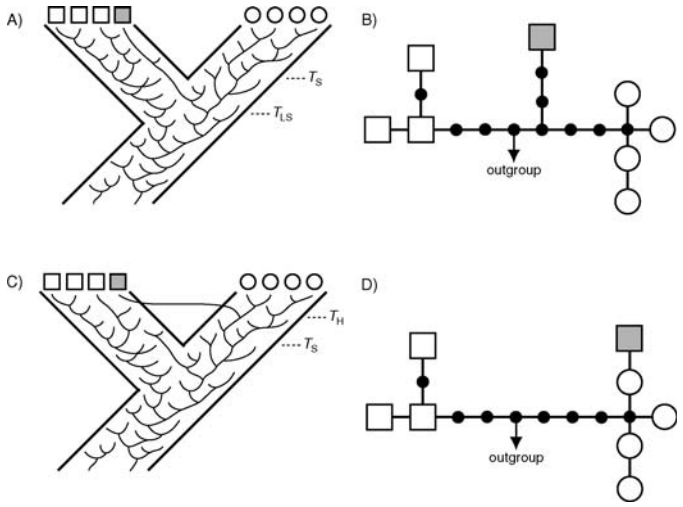

Fig. 7. Hypothetical evolutionary networks that illustrate expected patterns of incongruence due to incomplete lineage sorting or hybridization. (A) Two hypothetical species with incomplete lineage sorting between them and (B) expected network for this scenario. (C) Hypothetical species with a hybridization event between them and (D) expected network for this scenario. The gray box represents the incongruent allele. T_S = time to speciation; T_{LS} = time to divergence of the incongruent allele and the other species alleles in the situation of lineage sorting; T_H = time to divergence of the incongruent allele and the other species' alleles during interspecific hybridization.

not originally involved in its formation and can cause an autopolyploid to look like an allopolyploid. However, gene flow also can cause an allopolyploid to look like an autopolyploid if alleles from a diploid species are fixed in the allopolyploid due to recurrent gene flow. A further confounding factor is allelic segregation. Allopolyploids are expected to maintain alleles from both parental species in their genomes by disomic segregation due to bivalent formation at meiosis. This is to be expected in northeastern American polyploid *Rosa* species because individuals from the three polyploid species investigated show bivalent formation (Erlanson, 1929; Lewis, 1957b). Nonetheless, occasional pairing between homologous chromosomes (from the different diploid species) at meiosis could cause tri- or tetravalent formation. Indeed, trivalents and tetravalents have been observed in these polyploids (W. H. Lewis, unpublished data), but these and other meiotic irregularities such as lagging chromosomes and interlocked ring bivalents are rare and are only known of individuals from the zone of sympatry between *R. arkansana* and *R. carolina* (Lewis, 1966). Such multivalent formation leads to multisomic segregation that could bias the expected 1:1 ratio of parental alleles in an individual. Eventually this could lead to the fixation of alleles that come from a single diploid parent, resulting in a situation in which an allopolyploid might look like an autopolyploid.

Inspection of the *GAPDH* network shows that polyploids are of recent origin because many polyploid haplotypes are also found in contemporary diploids. The presence of shared haplotypes among diploids and polyploids makes the determination of the type of polyploid formation more difficult for each species. This is because it is harder to eliminate hypotheses of hybridization among polyploid species and of gene flow between diploids and polyploids when diploids and

polyploids share the same haplotypes. Of these confounding processes, gene flow between ploidy levels seems unlikely for many reasons. First, very few triploids have been reported in wild roses (Erlanson, 1929), and crosses between diploids and tetraploids give triploids that are highly sterile (Erlanson, 1934). Second, diploid and tetraploid species of *Rosa* are often separated both in space and in time of flowering, with diploids flowering before the tetraploids, except for *R. palustris*, that flowers after all other species (Erlanson, 1930). Polyploids more often grow in dry soils, either in sandy soils (*R. carolina* and *R. virginiana*; although *R. virginiana* also grows in salt marshes) or in upland prairies (*R. arkansana*), whereas diploids grow in bogs (*R. nitida* and *R. palustris*) or in mesic soils along woods and rivers (*R. blanda* and *R. woodsii*). Therefore, we consider that the probability of gene flow between ploidy levels is low. For the other conflicting processes, hybridization at the polyploid level and allele segregation in the polyploids, the recent origin of the polyploids allows us to make some assumptions about the expected results.

Given that each polyploid species has evolved recurrently (discussed later), the recent origin of polyploids gives little time for between-population genetic homogenization within polyploid species. Thus, if we have many recent formations of the polyploid species, we expect that individuals from several separate populations retain information of their origin. In other words, hybridization and allele segregation should only affect a limited number of populations in each species. Therefore, the expectations for an autopolyploid species is that most individuals will have alleles from a single diploid species even if a few can have acquired alleles from another diploid species via introgression. Moreover, individuals bearing introgressed alleles should be geographically close to individuals (or species) from which the allele is derived (Rieseberg, 1998). In a similar way, it is unlikely that parental alleles in allopolyploid individuals will segregate in all populations and even less likely that the segregation will always be toward the same parental alleles (unless there is selection). Therefore, we expect that most individuals of an allopolyploid species will possess alleles from two diploid species even if some individuals could have fixed alleles from a single diploid species or have segregated toward a ratio of parental diploid alleles that deviates from the expected 1:1 ratio. In a further attempt to limit the potential impact of hybridization on the conclusions regarding polyploid origins, we avoided sampling individuals in areas where the distribution of polyploid species overlapped. The only exception is for *R. arkansana*, for which a few individuals were sampled from the zone of sympatry with *R. carolina*; potential impacts on the conclusions will be discussed later.

Of the eight *R. arkansana* individuals sampled, all have alleles in the *blanda–woodsii* group with five lacking alleles from the *foliolosa–nitida–palustris* diploid group. Moreover, the three individuals with alleles from this latter group come from the region of sympatry between *R. arkansana* and *R. carolina* (Figs. 1 and 6). This suggests that *R. arkansana* evolved from within the *blanda–woodsii* group and that the presence of alleles from the *foliolosa–nitida–palustris* group in some individuals could be the result of introgression from *R. carolina*. Indeed, a hypothesis of introgression from *R. carolina* to *R. arkansana* is supported by cytological (Lewis, 1966) and morphological (A. Fishbein and W. H. Lewis, Washington University, unpublished manuscript) evidence, suggesting hybridization between these species. Because the relationships within the *blanda-woodsii* group are unresolved using the *GAPDH* marker, it cannot be stated whether *R. arkansana* is an auto- or an allopolyploid using a taxonomic definition (Grant, 1981; Ramsey and Schemske, 1998). Yet, some prefer to define autopolyploidy in a cytological context (Stebbins, 1980; Levin, 2002) according to which autopolyploids evolve from parents that are interfertile at the diploid level, whereas allopolyploids are formed from a hybrid that has reduced fertility. This definition predicts multivalent formation in autopolyploids and bivalent formation in allopolyploids, at least in the first stages of their evolution. According to the cytological definition, *R. arkansana* would probably be an autopolyploid because *R. blanda* and *R. woodsii* produce highly fertile hybrids and because they are morphologically and genetically similar.

Rosa carolina is different from *R. arkansana* in that all except two individuals investigated have alleles from both the *blanda–woodsii* and the *foliolosa–nitida–palustris* diploid groups. Given the wide geographic distribution of the individuals sampled, we can affirm that *R. carolina* is an allopolyploid with one parent from the *blanda–woodsii* diploid group and the other from the *foliolosa–nitida–palustris* group. The deviation from a 1:1 ratio of parental alleles expected for allopolyploids observed in some individuals is probably the result of either segregation of homologous chromosomes or introgression.

Finally, individuals of *R. virginiana* were found to possess only alleles that were exclusive to the *foliolosa–nitida–palustris* diploid group, except for two individuals that also have a *blanda–woodsii* allele and one that has an allele of ambiguous origin. Therefore, the most likely hypothesis for the origin of this polyploid species is that it originated from within the *foliolosa–nitida–palustris* diploid group. Again, we cannot be certain that *R. virginiana* is an auto- or an allopolyploid due to the lack of resolution within the *foliolosa–nitida–palustris* group. It is highly likely that *R. foliolosa* was not involved in the evolution of this species, however, because no *R. virginiana* alleles were closely related to the alleles sampled from *R. foliolosa*. The situation is also different from that for *R. arkansana* because we have no information on the fertility of hybrids between *R. palustris* and *R. nitida*. Hence, any conclusions regarding the type of polyploid origin of *R. virginiana* must await further data.

To summarize, *R. arkansana* evolved from the *blanda-woodsii* group, *R. viriginiana* from the *foliolosa-nitida-palustris* group, and *R. carolina* from a cross between these two eastern diploid groups. These results allow an evaluation of different hypotheses that have been proposed concerning the origins of eastern polyploids. Erlanson (1929) proposed that *R. arkansana* originated from a cross between *R. blanda* and either *R. macounii* Greene or *R. fendleri* Crépin, two species now considered synonymous with *R. woodsii* (Erlanson, 1934). This hypothesis is compatible with the present findings, although our results cannot confirm that two taxonomic species were involved. For *R. carolina*, Erlanson (1929) first proposed that *R. virginiana* would have crossed with *R. palustris* and that the hybrid eventually would have given a tetraploid that would have backcrossed to *R. virginiana* to give *R. carolina*. This hypothesis is improbable according to the present results because it would imply that the genetic diversity of *R. carolina* is a subset of *R. virginiana*. Because several *R. carolina* haplotypes do not have *R. virginiana* haplotypes as ancestors, our data disagree with such an evolutionary scenario. A few

years later, Erlanson (1938) suggested that *R. blanda* and *R. woodsii* gave rise to all three eastern tetraploid species as well as to *R. foliolosa*, *R. nitida*, and *R. palustris*. Her hypothesis regarding the evolution of *R. foliolosa*, *R. nitida*, and *R. palustris* seems improbable in light of the present data because these species do not appear to be derived from *R. blanda* and *R. woodsii*. Her hypothesis regarding the evolution of *R. carolina* and *R. virginiana* from *R. blanda* and *R. woodsii* alone is also likely inaccurate because the *foliolosa-nitida-palustris* diploid group was certainly involved in the origin of these two tetraploid species.

The results clearly show that the western diploid species were not involved in the origins of the eastern polyploid species. It is indeed improbable that a western species would have been involved in the origin of the polyploids without leaving a trace, given that several polyploid individuals from a wide geographic range were sampled. A general pattern of evolution within section *Cinnamomeae* in North America thus emerges from these results: diploids west and east of the Rocky Mountains seem to form distinct groups and eastern polyploids evolved from eastern diploids following the diversification of diploids.

Multiple origins of polyploidy—The number of polyploid origins was estimated using "polyploid haplotype groups" (Fig. 3), which estimates the genetic diversity of polyploids that is contributed by diploids. When working with haploid markers, each polyploid haplotype group can be interpreted as a distinct polyploid origin (e.g., Soltis et al., 1989; Doyle et al., 1990; Segraves et al., 1999; Sharbel and Mitchell-Olds, 2001). Similarly for autosomal markers, a specific combination of polyploid haplotype groups in individuals can sometimes be considered to represent a distinct origin. This is true of selfing allopolyploids that are homozygous at each homologous locus (as in *Glycine*; Doyle et al., 2004) and of clonal taxa (Joly and Bruneau, 2004). However, more often alleles at nuclear loci will segregate in polyploids, and this can create any possible combination of alleles. Hence, interpreting each genotype as an independent origin would seem to overestimate the true number of polyploid origins. For this reason, it was assumed that for each species, each tetraploid formation involved four distinct polyploid haplotype groups and that each independent formation always involved polyploid haplotype groups that were not involved in other polyploid origins. These assumptions are clearly overly conservative. For example, there may be unsampled diploid haplotypes that would increase the number of polyploid haplotype groups and a tetraploid formation can involve less than four alleles. Yet, the approach is legitimate if the objective is to evaluate the likelihood that species evolved recurrently rather than to estimate the true number of polyploid origins.

According to these conservative assumptions, all polyploid species must have evolved at least three times to explain the observed diversity of polyploids. This estimate makes many simplifications such as an absence of gene flow between ploidy levels that would tend to overestimate the number of independent origins. Yet, the impact of gene flow between ploidy levels is probably limited in North American roses (discussed earlier). Hybridization between polyploid species is another way by which polyploids acquire genetic variability that is not due to multiple origins. It is harder to account for hybridization because polyploids are known to hybridize and because they have a recent origin; this is why individuals mostly were sampled from outside the zones of sympatry between polyploids. The only exception is *R. arkansana*, from which we sampled five individuals that are considered near or in the sympatric zone with *R. carolina* (Figs. 1, 6). But even with these individuals removed (accessions 345, 406, 416, 665, and 692), there are still seven polyploid haplotype groups represented, and two independent origins of *R. arkansana* are needed to explain such a diversity.

Interestingly, polyploids have been able to acquire most of the available genetic diversity at the diploid level; almost all diploid haplotypes were also found in one or more polyploid species (Fig. 5). This further supports the hypothesis of independent origins of polyploid species, but above all it shows that polyploids possess a high degree of genetic variation. In the end, it is this genetic diversity that is most important, not how it was acquired. This variability, coupled with recombination and mutation in polyploid species, is likely to allow polyploid species to create adaptive genotypes that will be fitter and have more evolutionary potential in certain environments.

Taxonomic consequences—The rose species investigated here have sometimes been divided into sections *Cinnamomeae* (*R. arkansana*, *R. blanda*, *R. woodsii*) and *Carolinae* (*R. carolina*, *R. foliolosa*, *R. nitida*, *R. palustris*, *R. virginiana*) based on strictly basal placentation (*Carolinae*) vs. basiloparietal placentation (*Cinnamomeae*), presence (*Carolinae*) vs. absence (*Cinnamomeae*) of hypanthium glands, and deciduous (*Carolinae*) vs. persistent (*Cinnamomeae*) sepals after fruit maturation (Crépin, 1889). The present data suggest that the separation of these two sections is artificial. First, it makes section *Cinnamomeae* paraphyletic, and second, the reticulate origin of *R. carolina* also renders section *Carolinae* unnatural. Therefore, the best solution would be to treat section *Carolinae* as synonymous with section *Cinnamomeae*. This was previously proposed by Erlanson (1934) and Lewis (1957a) based on the unreliability of the morphological characters that were used to separate these sections and also supports investigations of biochemical (Grossi et al., 1998) and molecular characters (Wissemann and Ritz, 2005). Yet, this taxonomy still is used in the most recent comprehensive flora treatments in the United States (generic flora of the southeastern United States, Robertson, 1974) and in Europe (Tutin et al., 1968), perhaps because Rehder's (1940) classification, which uses section *Carolinae*, is still the most widely cited taxonomic treatment of *Rosa*. We suggest that section *Carolinae* be completely removed from further taxonomic treatments.

The present study also sheds light upon the species status of the three polyploid taxa of the *R. carolina* complex. The results suggest that *R. arkansana*, *R. carolina*, and *R. virginiana* have distinct evolutionary histories, although it will certainly be important to confirm this with more markers. Consequently, this also suggests that these polyploids should be considered distinct species. These species are, of course, highly polymorphic probably in part owing to their recurrent origins, and their identification will remain difficult especially in regions of sympatry where the extensive variation is best explained by hybrid zones. Yet, results suggest that these are secondary hybrid zones (Endler, 1977; Barton and Hewitt, 1985) that were formed after polyploid speciation. Of course, distinct evolutionary histories do not guarantee that species will always remain distinct, and the extent of gene flow in these secondary hybrid zones will determine the future of these polyploids.

Conclusion—This study shows that both hybridization and polyploidy have been important in the evolution of the *Rosa carolina* complex. Three species are the result of polyploid speciation, and hybridization has occurred among diploid species and has been involved in the formation of the polyploids. In addition, hybridization further complicates the picture of the polyploids and may lead to the extensive morphological variation observed in these taxa. Finally, this study of wild rose species gives a conceptual framework that may be used to unveil the evolutionary history of other species complexes where hybridization and polyploidy are important.

LITERATURE CITED

AQUADRO, C. F., S. F. DESSE, M. M. BLAND, C. H. LANGLEY, AND C. C. LAURIE-AHLBERG. 1986. Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* 114: 1165–1190.

ARCHAMBAULT, A., AND A. BRUNEAU. 2004. Phylogenetic utility of the *LEAFY/FLORICAULA* gene in the Caesalpinioideae (Leguminosae): gene duplication and a novel insertion. *Systematic Botany* 29: 609–626.

ARNOLD, M. L. 1997. Natural hybidization and evolution. Oxford University Press, New York, New York, USA.

BALDWIN, B. G., M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL, AND M. J. DONOGHUE. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.

BARTON, N. H., AND G. M. HEWITT. 1985. Analysis of hybrid zones. *Annual Reviews in Ecology and Systematics* 16: 113–148.

BAUM, D. A., AND K. L. SHAW. 1995. Genealogical perspectives on the species problem. *In* P. C. Hoch and A. G. Stephenson [eds.], Experimental and molecular approaches to plant biosystematics, 289–303. Monographs in Systematic Botany. Missouri Botanical Garden, St. Louis, Missouri, USA.

CASTELLOE, J., AND A. R. TEMPLETON. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution* 3: 102–113.

CLEMENT, M., D. POSADA, AND K. A. CRANDALL. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.

CRANDALL, K. A., AND A. R. TEMPLETON. 1993. Empirical test of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134: 959–969.

CRÉPIN, F. 1889. Sketch of a new classification of roses. *Journal of the Royal Horticultural Society* 11: 217–228.

CRÉPIN, F. 1894. *Rosae hybridae*: études sur les roses hybrides. *Bulletin de la Société Royale de Botanique de Belgique, 1ière partie (Mémoires)* 33: 52–55.

CRÉPIN, F. 1896. *Rosae Americanae*. I. Observations upon the genus *Rosa* in North America. *Botanical Gazette* 22: 1–34.

CRONN, R., M. CEDRONI, T. HASELKORN, C. GROVER, AND J. F. WENDEL. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics* 104: 482–489.

DENTON, A. L., B. L. MCCONAUGHY, AND B. D. HALL. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Molecular Biology and Evolution* 15: 1082–1085.

DONNELLY, P., AND S. TAVARÉ. 1986. The ages of alleles and a coalescent. *Advances in Applied Probability* 18: 1–19.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

DOYLE, J. J., J. L. DOYLE, A. H. D. BROWN, AND J. P. GRACE. 1990. Multiple origins of polyploids in the *Glycine tabacina* complex inferred from chloropast DNA polymorphism. *Proceedings of the National Academy of Sciences, USA* 87: 714–717.

DOYLE, J. J., J. L. DOYLE, A. H. D. BROWN, AND R. G. PALMER. 2002. Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: histone H3-D gene sequences. *Evolution* 56: 1388–1402.

DOYLE, J. J., J. L. DOYLE, J. T. RAUSCHER, AND A. H. D. BROWN. 2004. Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): a study of contrasts. *Biological Journal of the Linnean Society* 82: 583–597.

ENDLER, J. A. 1977. Geographic variation, speciation, and clines. Princeton University Press, Princeton, New Jersey, USA.

ERLANSON, E. W. 1929. Cytological conditions and evidences for hybridity in North America wild roses. *Botanical Gazette* 87: 443–506.

ERLANSON, E. W. 1930. The phenological procession in North American wild roses in relation to the polyploid series. *Papers of the Michigan Academy of Sciences, Arts and Letters* 11: 137–150.

ERLANSON, E. W. 1934. Experimental data for a revision of the North American wild roses. *Botanical Gazette* 96: 197–259.

ERLANSON, E. W. 1938. Phylogeny and polyploidy in *Rosa*. *New Phytologist* 37: 72–81.

ERLANSON MACFARLANE, E. W. 1966. The old problem of species in *Rosa* with special reference to North America. *American Rose Annual* 51: 150–160.

EVANS, B. J., D. B. KELLEY, D. J. MELNICK, AND D. C. CANNATELLA. 2005. Evolution of RAG-1 in polyploid clawed frogs. *Molecular Biology and Evolution* 22: 1193–1207.

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. *Molecular Phylogenetics and Evolution* 17: 388–400.

EXCOFFIER, L., P. SMOUSE, AND J. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.

FERNALD, M. L. 1922. Notes on the flora of western Nova Scotia 1921. *Rhodora* 24: 157–180.

FLORA OF NORTH AMERICA EDITORIAL COMMITTEE. 1993. Flora of North America, vol. 1. Oxford University Press, New York, New York, USA.

FROHLICH, M. W., AND E. M. MEYEROWITZ. 1997. The search for flower homeotic gene homologs in basal angiosperms and Gnetales: a potential new source for data on the evolutionary origin of flowers. *International Journal of Plant Sciences* 158: 131–142.

FUNK, D. J., AND K. E. OMLAND. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Reviews in Ecology, Evolution, and Systematics* 34: 397–423.

GRANT, V. 1981. Plant speciation, 2nd ed. Columbia University Press, New York, New York, USA.

GROSSI, C., O. RAYMOND, AND M. JAY. 1998. Flavonoid and enzyme polymorphisms and taxonomic organisation of *Rosa* sections: Carolinae, Cinnamomeae, Pimpinellifoliae and Synstylae. *Biochemical Systematics and Ecology* 26: 857–871.

HELFGOTT, D. M., AND R. J. MASON-GAMER. 2004. The evolution of North American *Elymus* (Triticeae: Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. *Systematic Botany* 29: 850–861.

HOLDER, M. T., J. A. ANDERSON, AND A. K. HOLLOWAY. 2001. Difficulties in detecting hybridization. *Systematic Biology* 50: 978–982.

HUDSON, R. R., AND J. A. COYNE. 2002. Mathematical consequences of the genealogical species concept. *Evolution* 56: 1557–1565.

HUNTER, J. C., AND J. A. MATTICE. 2002. The spread of woody exotics into the forests of a northeastern landscape, 1938–1999. *Journal of the Torrey Botanical Society* 129: 220–227.

HURST, C. C. 1925. Chromosomes and characters in *Rosa* and their significance in the origin of species. *Experiments in Genetics* 37: 534–550.

JOLY, S., AND A. BRUNEAU. 2004. Evolution of triploidy in *Apios americana* (Leguminosae) revealed by the genealogical analysis of the histone H3-D gene. *Evolution* 58: 284–295.

JUDO, M. S. B., A. B. WEDEL, AND C. WILSON. 1998. Stimulation and

suppression of PCR-mediated recombination. *Nucleic Acids Research* 26: 1819–1825.

LEVIN, D. A. 2002. The role of chromosomal change in plant evolution. Oxford University Press, Oxford, UK.

LEWIS, W. H. 1957a. An introduction to the genus *Rosa* with special reference to *R. acicularis*. *Virginia Journal of Science* 8: 197–202.

LEWIS, W. H. 1957b. A monograph of the genus *Rosa* in North America east of the Rocky Mountains. Ph.D. dissertation, University of Virginia, Charlottesville, Virginia, USA.

LEWIS, W. H. 1957c. Revision of the genus *Rosa* in eastern North America: a review. *American Rose Annual* 42: 116–126.

LEWIS, W. H. 1958. A monograph of the genus *Rosa* in North America. II. *R. foliolosa*. *Southwestern Naturalist* 3: 145–153.

LEWIS, W. H. 1959. A monograph of the genus *Rosa* in North America. I. *R. acicularis*. *Brittonia* 11: 1–24.

LEWIS, W. H. 1962. Monograph of the genus *Rosa* in North America. IV. *R.* × *dulcissima*. *Brittonia* 14: 65–71.

LEWIS, W. H. 1966. Chromosome numbers of phanerogams. 1. *Annals of the Missouri Botanical Garden* 53: 100–101.

LEWIS, W. H., AND R. E. BASYE. 1961. Analysis of nine crosses between diploid *Rosa* species. *Proceedings of the American Society for Horticultural Science* 78: 573–579.

LINNAEUS, C. 1753. Species plantarum. Impensis Laurentii Salvii, Stockholm, Sweden.

MADDISON, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

MASON-GAMER, R. J. 2004. Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Systematic Biology* 53: 25–37.

MATSUMOTO, S., M. KOUCHI, J. YABUKI, M. KUSUNOKI, Y. UEDA, AND H. FUKUI. 1998. Phylogenetic analyses of the genus *Rosa* using the *mat*K sequence: molecular evidence for the narrow genetic background of modern roses. *Scientia Horticulturae* 77: 73–82.

MAYNARD SMITH, J., AND N. H. SMITH. 1998. Detecting recombination from gene trees. *Molecular Biology and Evolution* 15: 590–599.

MEINERS, S. J., S. T. A. PICKETT, AND M. L. CADENASSO. 2001. Effect of plant invasions on the species richness of abandoned agricultural land. *Ecography* 24: 633–644.

MILLAN, T., F. OSUNA, S. BOBOS, A. M. TORRES, AND J. I. CUBERO. 1996. Using RAPDs to study phylogenetic relationships in *Rosa*. *Theoretical and Applied Genetics* 92: 273–277.

MISHLER, B. D., AND E. C. THERIOT. 2000. The phylogenetic species concept (*sensu* Mishler and Theriot): monophyly, apomorphy, and phylogenetic species concepts. *In* Q. D. Wheeler and R. Meier [eds.], Species concepts and phylogenetic theory, 44–54. Columbia University Press, New York, New York, USA.

MOORE, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49: 718–726.

OLSEN, K. M., AND B. A. SCHAAL. 1999. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proceedings of the National Academy of Sciences, USA* 96: 5586–5591.

OTTO, S. P., AND J. WHITTON. 2000. Polyploid incidence and evolution. *Annual Reviews of Genetics* 34: 401–437.

PAMILO, P., AND M. NEI. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568–583.

PETERSEN, G., AND O. SEBERG. 2004. On the origin of the tetraploid species *Hordeum capense* and *H. secalinum* (Poaceae). *Systematic Botany* 29: 862–873.

PFEIL, B. E., C. L. BRUBAKER, L. A. CRAVEN, AND M. D. CRISP. 2004. Paralogy and orthology in the Malvaceae *rpb*2 gene family: investigation of gene duplication in *Hibiscus*. *Molecular Biology and Evolution* 21: 1428–1437.

POSADA, D. 2002. Evaluating methods for detecting recombination from DNA sequences: empirical data. *Molecular Biology and Evolution* 19: 708–717.

POSADA, D., AND K. A. CRANDALL. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences, USA* 98: 13757–13762.

RAMSEY, J., AND D. W. SCHEMSKE. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Reviews in Ecology and Systematics* 29: 467–501.

RATSEK, J. C., W. S.. FLORY, JR., AND S. H. YARNELL. 1940. Crossing relations of some diploid and polyploid species of roses. *Proceedings of the American Society for Horticultural Science* 38: 637–654.

RATSEK, J. C., S. H. YARNELL, AND W. S. FLORY JR. 1939. Crossing relations of some diploid species of roses. *Proceedings of the American Society for Horticultural Science* 37: 983–992.

RAUSCHER, J. T., J. J. DOYLE, AND A. H. D. BROWN. 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. *Molecular Ecology* 11: 2691–2702.

REHDER, A. 1940. Manual of cultivated trees and shrubs hardy in North America, 2nd ed. Macmillan, New York, New York, USA.

RIESEBERG, L. H. 1998. Molecular ecology of hybridization. *In* G. R. Carvalho [ed.], Advances in molecular ecology, 243–265. IOS Press, Burke, Virginia, USA.

RITZ, C. M., H. SCHMUTHS, AND V. WISSEMANN. 2005. Evolution by reticulation: European dogroses originated by multiple hybridization across the genus *Rosa*. *Journal of Heredity* 96: 4–14.

ROBERTSON, K. R. 1974. The genera of Rosaceae in the southeastern United States: tribe Roseae. *Journal of the Arnold Arboretum* 55: 611–621.

ROSENBERG, N. A. 2002. The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology* 61: 225–247.

ROSENBERG, N. A. 2003. The shape of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution* 57: 1465–1477.

RYDBERG, A. 1920. Roses of northeastern America. *Bulletin of the Torrey Botanical Club* 47: 45–66.

SANG, T., AND Y. ZHONG. 2000. Testing hybridization hypotheses based on incongruent gene trees. *Systematic Biology* 49: 422–434.

SCHNEIDER, S., D. ROESSLI , AND L. EXCOFFIER. 2000. Arlequin: a software for population genetics data analysis, version 2. Genetic and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

SEGRAVES, K. A., J. N. THOMPSON, P. S. SOLTIS, AND D. E. SOLTIS. 1999. Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. *Molecular Ecology* 8: 253–262.

SENCHINA, D. S., I. ALVAREZ, R. C. CRONN, B. LIU, J. RONG, R. D. NOYES, A. H. PATERSON, R. A. WING, T. A. WILKINS, AND J. F. WENDEL. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution* 20: 633–643.

SHARBEL, T. F., AND T. MITCHELL-OLDS. 2001. Recurrent polyploid origins and chloroplast phylogeography in the *Arabis holboellii* complex (Brassicaceae). *Heredity* 87: 59–68.

SIMMONS, M. P., AND H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

SMEDMARK, J. E. E., T. ERIKSSON, R. C. EVANS, AND C. S. CAMPBELL. 2003. Ancient allopolyploid speciation in Geinae (Rosaceae): evidence from nuclear granule-bound starch synthase (GBSSI) gene sequences. *Systematic Biology* 52: 374–385.

SOLTIS, D. E., P. S. SOLTIS, AND B. D. NESS. 1989. Chloroplast-DNA variation and multiple origins of autopolyploidy in *Heuchera micrantha* (Saxifragaceae). *Evolution* 43: 650–656.

STEARN, W. T. 1957. An introduction to the *Species Plantarum* and cognate botanical works of Carl Linnaeus, Carl Linnaeus—*Species Plantarum*—A facsimile of the 1st ed. 1753. The Ray Society, London, UK.

STEBBINS, G. L. 1950. Variation and evolution in plants. Columbia University Press, New York, New York, USA.

STEBBINS, G. L. 1971. Chromosomal evolution in higher plants. Edward Arnold, London, UK.

STEBBINS, G. L. 1980. Polyploidy in plants: unsolved problems and

prospects. *In* W. H. Lewis [ed.], Polyploidy: biological relevance, 495–520. Plenum Press, New York, New York, USA.

STRAND, A. E., J. LEEBENS-MACK, AND B. G. MILLIGAN. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology* 6: 113–118.

SWOFFORD, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, Massachusetts, USA.

TÄCKHOLM, G. 1922. Zytologische studien über die Gattung *Rosa*. *Acta Horti Bergiana* 7: 97–381.

TEMPLETON, A. R. 1989. The meaning of species and speciation: a genetic perspective. *In* D. Otte and J. A. Endler [eds.], Speciation and its consequences, 3–27. Sinauer, Sunderland, Massachusetts, USA.

TEMPLETON, A. R. 2001. Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology* 10: 779–791.

TEMPLETON, A. R., K. A. CRANDALL, AND C. F. SING. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619–633.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876–4882.

THOMPSON, J. D., D. G. HIGGINS, AND T. J. GIBSON. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.

TUTIN, T. G., V. H. HEYWOOD, N. A. BURGES, D. M. MOORE, D. H. VALENTINE, S. M. WALTERS, AND D. A. WEBB. 1968. Flora europaea. Cambridge University Press, London, UK.

WAGNER, A., N. BLACKSTONE, P. CARTWRIGHT, M. DICK, B. MISOF, P. SNOW, G. P. WAGNER, J. BARTELS, M. MURTHA, AND J. PENDLETON. 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Systematic Biology* 43: 250–261.

WATSON, S. 1885. Contributions to American botany. I. A history and revision of the roses of North America. *Proceedings of the American Academy of Arts and Sciences* 20: 324–352.

WHEELER, Q. D., AND N. I. PLATNICK. 2000. The phylogenetic species concept. *In* Q. D. Wheeler and R. Meier [eds.], Species concepts and phylogenetic theory, 55–69. Columbia University Press, New York, New York, USA.

WISSEMANN, V., AND C. M. RITZ. 2005. The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and *atp*B-*rbc*L intergenic spacer (IGS) versus conventional taxonomy. *Botanical Journal of the Linnean Society* 147: 275–290.

WOLLENBERG, K., AND J. C. AVISE. 1999. Sampling properties of genealogical pathways underlying population pedigrees. *Evolution* 52: 957–966.

YOUNG, N. D., AND J. HEALY. 2003. GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* 4: 6.