American Journal of Botany 89(11): 1697–1708. 2002.

THE PHYLOGENETIC UTILITY OF NUCLEOTIDE SEQUENCES OF SORBITOL 6-PHOSPHATE DEHYDROGENASE IN *PRUNUS* (ROSACEAE)[1]

ESTEBAN BORTIRI,[2] SANG-HUN OH, FANG-YOU GAO, AND DAN POTTER

University of California—Davis, Department of Pomology, One Shields Avenue, Davis, California 95616 USA

Sequences from *s6pdh*, a gene that encodes sorbitol-6-phosphate dehydrogenase in the Rosaceae, are used to reconstruct the phylogeny of 22 species of *Prunus*. The *s6pdh* sequences alone and in combination with previously published sequences of the internal transcribed spacer (ITS) and the cpDNA *trnL-trnF* spacer are analyzed using parsimony and maximum likelihood methods. Both methods reconstructed the same phylogeny when *s6pdh* sequences are used alone and in combination with ITS and *trnL-trnF*, and the topology is in agreement with previous studies that used a larger sample size. The *s6pdh* sequences have about twice as many informative sites as ITS. A molecular clock is rejected for *s6pdh*, most likely due to greater rates of evolution in subgenera *Padus* and *Laurocerasus* than in the rest of the genus. Phylogenetic reconstruction of *Prunus* as determined by analysis of the combined data set suggests an early split into two clades. One is composed of subgenera *Cerasus*, *Laurocerasus*, and *Padus*. The second includes subgenera *Amygdalus*, *Emplectocladus*, and *Prunus*. Species of section *Microcerasus* (formerly in subgenus *Cerasus*) are nested within subgenus *Prunus*. The order of branching and relationships among early diverging lineages is weakly supported, as a result of very short branches that may indicate rapid radiation.

Key words: alcohol dehydrogenase; likelihood; molecular clock; parsimony; *Prunus*; *s6pdh*.

The genus *Prunus* L. (Rosaceae), with over 200 species (Rehder, 1940) of shrubs and trees, is an important component of Northern Hemisphere forest (Fedorov et al., 1941; Elias, 1980; Yü et al., 1986) and desert (Mason, 1913; Browicz and Zohary, 1996) communities. A significant number of species are also found in the tropics worldwide (Kalkmann, 1965; Brako and Zarucchi, 1993). Many species are cultivated worldwide for their fruits, such as cherries (*P. cerasus* and *P. avium*), apricot (*P. armeniaca*), almond (*P. dulcis*), peach (*P. persica*), and plums (several species of subgenus *Prunus*) (Moore and Ballington, 1990). *Prunus serotina*, the North American black cherry, is valued for its timber. Several species are ornamentals (Krüssmann, 1986), in particular flowering cherries of subgenus *Cerasus* (Kuitert, 1999). Wild North American plums (subgenus *Prunus*, section *Prunocerasus*) were an important source of food for Native Americans (Wight, 1915) and are used locally to make tarts and jams. Finally, species of *Prunus* play an important ecological role by providing food for wildlife (Elias, 1980; M. Beck, University of Nevada, personal communication).

In addition to their economic, cultural, and ecological importance, species of *Prunus* display considerable morphological variation in terms of types of inflorescences and fruits and an interesting geographic distribution. The evolution of some characters most likely arose as adaptations to special habitats, such as the presence of dry fruits in desert species.

Prunus is the largest genus in the subfamily Amygdaloideae, which also includes *Exochorda*, *Maddenia*, *Oemleria*, *Prinsepia*, and *Pygeum* in the most recent taxonomic treatment of Rosaceae by Takhtajan (1997). This classification is supported by evidence from wood anatomy (Zhang, 1992) and chromosome number (Goldblatt, 1976). Molecular data have shown that *Maddenia* may be nested within *Prunus* (Lee and Wen, 2001) and *Oemleria*, *Prinsepia*, and *Exochorda* are monophyletic (Morgan, Soltis, and Robertson, 1994; Potter et al., 2002). However, these two clades are not strongly supported as sister to each other. A phylogeny of Rosaceae using *rbcL* weakly supported Amygdaloideae sensu Takhtajan (Morgan, Soltis, and Robertson, 1994), while other molecular data show no relationship between *Prunus* and the clade including *Oemleria*, *Prinsepia*, and *Exochorda* (Potter et al., 2002). *Pygeum* has not been included in the mentioned phylogenetic studies; thus, its relationships to *Prunus* remain to be tested by molecular cladistics.

Historically, *Prunus* has been classified into five subgenera: *Amygdalus*, *Cerasus*, *Laurocerasus*, *Padus*, and *Prunus*. Other authors have chosen to treat those subgenera as separate genera (reviewed by McVaugh, 1951). In this paper we follow Rehder's (1940) classification of the genus into five subgenera with the addition of subgenus *Emplectocladus* (Torr.) Sargent (http://ajbsupp.botany.org/v89/).

Prunus has been the subject of two recent phylogenetic studies (Bortiri et al., 2001; Lee and Wen, 2001). In a maximum parsimony analysis of combined internal transcribed spacer (ITS) and *trnL-trnF* data, we found that most species fall into two groups. Subgenera *Cerasus*, *Padus*, and *Laurocerasus* together are supported as monophyletic and a second group consisting of subgenera *Amygdalus*, *Prunus*, *Emplectocladus*, and the section *Microcerasus* (subgenus *Cerasus*) was found in the strict consensus of most parsimonious trees

[1] Manuscript received 15 January 2002; revision accepted 16 May 2002.

The authors thank Ji-Young Lee for her help with the Southern blots, A. Dandekar and S. Uratsu for generously providing the *Malus domestica* cDNA clone, Terry Lee Harrington for providing the *Sorbus* sp. material, and Clay Weeks for his help to identify accessions at the Davis Clonal Germplasm Repository. The UC Davis Herbaria Society provided funding to EB for collection of plant material. Additional funding to EB came from the Center for Biosystematics at UC Davis and UC Davis Jastro Shield Awards. This research represents a portion of a dissertation submitted by the first author in partial fulfillment of the requirements for the Ph.D. degree at the University of California, Davis.

[2] Author for reprint requests (e-mail: pebortiri@ucdavis.edu).

TABLE 1. Primers used for amplification of *s6pdh*. Refer to Fig. 1 for their annealing sites on *s6pdh*.

Primer name	Sequence
s6pdh-a	5′-ATG CCG GTC ATC GGT CTC GG -3′
s6pdh-b	5′-GCA TAC ACG TCT AAG CCC CA -3′
s6pdh-c	5′-TTT GGA ATT CAG ACC ATG GGC ATG -3′
s6pdh-e	5′-CTT AGT TCG CAG CAT TGG TCT CAG -3′
s6pdh-h	5′-AGA CCA ATG CTG CGA ACT AAG CCC -3′
s6pdh-j	5′-GCA TAC ACG TCT AAG CCC CAA GTC TTG GAA GG -3′
s6pdh-k	5′-AGC TCA TTA CAA GAG TGA AGC AGA CGT TGG -3′
s6pdh-p	5′-AGA GTG GTC CTG GAT TTC TTA TCT A -3′
77L25	5′-TGG CCT TGG AGG TGA TGA AAA GTT C -3′

but lacked significant bootstrap support (Bortiri et al., 2001). Lee and Wen's (2001) results, based on ITS data, were similar in that they also found the same two major groups. When the two data sets, ITS and *trnL-trnF*, were analyzed separately, however, they supported different scenarios for the relationships of subgenus *Cerasus* (Bortiri et al., 2001). According to the chloroplast *trnL-trnF* spacer, *Cerasus* is placed in a group with subgenera *Amygdalus*, *Prunus*, *Emplectocladus*, and the sections *Penarmeniaca* and *Microcerasus*. The nuclear ITS region, on the other hand, supported a clade composed of subgenera *Cerasus*, *Padus*, and *Laurocerasus* (Bortiri et al., 2001).

One of the shortcomings of those previous studies was the lack of support for some deep nodes, in particular those in the group that included the subgenera *Emplectocladus*, *Amygdalus*, and *Prunus*. This outcome was attributed to high homoplasy levels in ITS, perhaps caused by rapid speciation and early hybridization (Bortiri et al., 2001).

In order to test the hypotheses of evolution and relationships in *Prunus* set forth by our previous work with ITS and *trnL-trnF*, we started studies on the nuclear gene *s6pdh*, which encodes $NADP^+$-dependent sorbitol-6-phosphate dehydrogenase (S6PDH [Enzyme Commission 1.1.1.200]) (Yamaki and Ishikawa, 1986). In the Rosaceae, with the exception of Rosoideae sensu stricto (Wallaart, 1980; Morgan, Soltis, and Robertson, 1994), the most important products of photosynthesis found in the phloem are sucrose and the sugar alcohol sorbitol (Webb and Burley, 1962). Sorbitol-6-phosphate is produced as a result of the reduction of glucose-6-phosphate by S6PDH, which uses $NADP^+$ as cofactor. S6PDH has been purified from leaves of loquat (*Eriobotrya japonica*) (Hirai, 1981) and apple seedlings (*Malus domestica*) (Kanayama and Yamaki, 1993). Tao, Uratsu, and Dandekar (1995) transformed tobacco (*Nicotiana tabacum*) plants, which do not produce sorbitol, with the *Malus domestica s6pdh* cDNA sequence and showed that leaves from transformed plants accumulate sorbitol. In addition, the concentration of sorbitol correlated positively with the expression of S6PDH. The genomic sequence of *Malus domestica s6pdh* consists of six exons totaling 3.4 kilobases (kb), with a TC simple repeat in intron 3 (Bains et al., 1998). Studies on polymorphisms of the TC microsatellite in intron 3 of *M. domestica* indicate the presence of two copies of *s6pdh* in that species and one in species of *Prunus* (S. Arulsekar, University of California, Davis, personal communication). *Malus* belongs to subfamily Maloideae, which has chromosome numbers $x = 15, 17$; whereas the rest of the family have chromosome numbers $x = 7, 8, 9$. It has been postulated that subfamily Maloideae is an ancient polyploid (e.g., Sax, 1933; Gladkova, 1972).

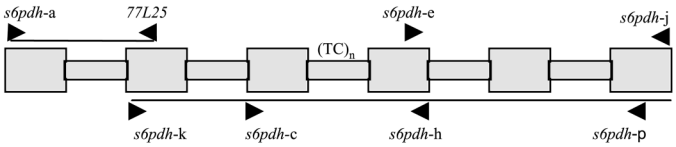

Fig. 1. The *s6pdh* gene in Rosaceae. Exons and introns are not drawn proportionally to their actual length. Primers used in this study and their position in the gene are shown. Intron 3 has a TC repeat. Size of PCR fragments are 0.8 kb (*s6pdh*-a/*77L25*), 0.44 kb (*s6pdh*-c/s6pdh-h), and 1.3 kb (*s6pdh*-k/*s6pdh*-p). Size of PCR products corresponds to *P. dulcis* but is conserved in all of the *Prunus* species.

For that reason we expected *s6pdh* to be a single-copy gene in diploid taxa of Rosaceae other than Maloideae.

In this paper we present a phylogenetic study of *Prunus* using parsimony and maximum likelihood methods as determined by sequence data from the nuclear gene *s6pdh.* We compare the results obtained using the two methods of phylogenetic reconstruction. We also compare rates of evolution of *s6pdh* with those of ITS and the *trnL-trnF* spacer and test whether rates of evolution in *s6pdh* are uniform across all lineages of *Prunus*. Finally, we combine data from all three markers to reconstruct the phylogeny of the genus using parsimony and maximum likelihood methods. The outcomes produced by both methods are discussed in relation to work carried out in previous studies. This paper aims to understand the species phylogeny of *Prunus*. Data from *s6pdh* will complement evidence from the ITS and *trnL-trnF* regions used previously (Bortiri et al., 2001). The main objectives are (a) to examine the phylogenetic relationships of subgenus *Cerasus*, (b) to test the monophyly of the group composed of subgenera *Prunus*, *Amygdalus*, *Emplectocladus*, and section *Microcerasus*, and (c) to generate data that will support a robust phylogeny of *Prunus*.

MATERIALS AND METHODS

Taxon sampling and DNA extractions—A subset of 22 species of *Prunus* were selected to represent all major clades found in a previous study (Bortiri et al., 2001). This sample includes all subgenera and most sections of *Prunus* (Mason, 1913; Rehder, 1940). DNA extractions were performed according to the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987) except the extractions of *Oemleria cerasiformis* (Hook. & Arn.) Landon and *Exochorda racemosa* (Lindl.) Rehder, which followed a different protocol (Dellaporta, Wood, and Hicks, 1983).

All collected vouchers are deposited at the University of California-Davis Herbarium (DAV). Information on the origin of the plant material, as well as GenBank accession numbers, are stored at the *American Journal of Botany* supplementary data web site (http://ajbsupp.botany.org/v89/).

Amplification and sequencing—We obtained *s6pdh* sequences from two different sources: direct polymerase chain reaction (PCR) sequencing and sequencing of cloned PCR products. In addition, we repeated some of the PCR and cloning reactions in order to detect amplification of two or more paralogous genes and to make sure that artifacts during PCR or cloning were not masking or altering the outcome substantially.

The primers used for amplification are shown in Table 1 and their relative positions on *s6pdh* are in Fig. 1. All primers, with the exception of *s6pdh*-p and *77L25*, were based on the cDNA sequence of *Malus domestica* (Kanayama et al., 1992; GenBank accession number D11080). We designed *s6pdh*-p to be used in combination with *s6pdh*-k on species of *Prunus*. A second primer, designated *77L25*, was created to use in combination with *s6pdh*-a for amplification of intron 1 of *Prunus* species (Fig. 1). The *s6pdh*-p and *77L25* primers were based on a cloned sequence of *s6pdh* from *P. dulcis* that had

been obtained with *s6pdh*-k and *s6pdh*-j. All primers were designed with the help of Oligos® 5.0 (National Biosciences, Plymouth, Minnesota, USA) and purchased from Genosys Biotechnologies (The Woodlands, Texas, USA). Sequences from *Pyrus caucasica* Fed., *Oemleria*, *Amelanchier alnifolia* (Nutt.), *Gillenia stipulata* (Torr.) Sarg., *Kageneckia oblonga* Ruiz & Pav., *Vauquelinia californica* (Torr.) Sarg., and *Sorbus* were produced by PCR amplification with primers *s6pdh*-k and *s6pdh*-j. *Sorbaria sorbifolia* (L.) A. Braun and *Spiraea cantoniensis* Lour. were amplified with the primer combinations *s6pdh*-k/*s6pdh*-h and *s6pdh*-c/*s6pdh*-j, which yield overlapping PCR products (Fig. 1). We obtained partial sequences from *Exochorda racemosa*, *Holodiscus microphyllus* Rydb., and *Chamaebatiaria millefolium* (Torr.) Maxim., by amplification with primers *s6pdh*-e and *s6pdh*-j. For the combined analyses of *s6pdh*, ITS, and the *trnL-trnF* spacer we used previously published sequences (Bortiri et al., 2001). We also obtained new ITS and *trnL-trnF* sequences from *Prunus caroliniana*, *P. microcarpa*, and *P. lusitanica* using protocols described previously (Bortiri et al., 2001).

In order to obtain enough DNA for direct sequencing of PCR products we used 100 μL reaction volumes. For the reactions that were later cloned the volume was reduced to 25 μL. All PCR reactions were performed using Applied Biosystems (Foster City, California, USA) reagents and consisted of buffer, 2.5 mmol/L of each dNTP, 0.2 μmol/L of each primer, 1.5 units of AmpliTaq Gold™, and 25 mmol/L $MgCl_2$. The PCR amplification conditions were as follows: an initial 10 min at 95°C, followed by 35 cycles of 30 s at 95°C, 1 min at 54°C, and 2 min at 72°C, and a final extension cycle of 7 min at 72°C.

For direct sequencing we purified the PCR products by separation in 0.8% agarose gels, excised the band of appropriate size, and purified the DNA with a QIAquick gel extraction kit (QIAGEN, Valencia, California, USA) according to the manufacturer's specifications. Direct sequencing of PCR fragments was performed using primers *s6pdh*-k, *s6pdh*-p, and *s6pdh*-c. We used primers *s6pdh*-a and *77L25* for sequencing of intron 1.

For the reactions that were later to be cloned we loaded a 15 μL aliquot of the finished PCR reaction in 0.8% agarose gels for electrophoresis to identify successful PCR amplifications. We cloned the PCR products from the remaining volume of successful reactions using a TOPO TA Cloning® kit (Invitrogen Corporation, Carlsbad, California, USA) according to manufacturer's specifications except that the reaction volumes were halved. We picked 4–10 colonies from each cloning reaction and grew them overnight in 10 mL of Luria-Bertani medium with ampicillin. We extracted the plasmid DNA using a QIAprep® Spin Miniprep kit (QIAGEN) and confirmed the presence of inserts by restricting approximately 2 μg of the mini-prepped plasmid DNA with 1.2 units of EcoRI. We selected clones that had inserts and sequenced them in both directions with primers T7 and T3. All sequencing was performed at the Plant Genetics Facility (UC Davis), which uses an ABI Prism sequencer (Applied Biosystems).

In order to search for proteins related to S6PDH we conducted a BLAST search in GenBank using the deduced amino acid sequence of *Malus domestica* (GenBank accession number GBAN-P28475).

Southern blots—We performed a Southern blot analysis to estimate the number of genes with sequence similarity to *s6pdh* that are present in the genome of several Rosaceae. DNA was extracted from 4 g of fresh leaves using the method of Doyle and Doyle (1987). We restricted 5 μg aliquots of DNA with *Eco*RI, *Eco*RV, *Dra*I, and *Hind*III. The restricted DNA fragments were separated by electrophoresis in 0.8% agarose gels and transferred to membranes according to standard procedures (Sambrook, Fritsch, and Maniatis, 1989). The probe was prepared by PCR amplification of a *Malus domestica* cDNA clone (Kanayama et al., 1992) using primers *s6pdh*-a and *s6pdh*-j, which yield a fragment of 0.93 kb, and labeled by primer extension using ^{32}P dNTPs (Sambrook, Fritsch, and Maniatis, 1989). The membranes containing the restricted fragments were hybridized overnight at 60°C and washed at 60°C in 2× SSPE/0.1% SDS buffer three times. The membranes were then exposed onto an X-ray film for 1 h and developed.

Sequence alignments—The sequences were edited in Sequencher version 4.1.2 (Gene Codes Corporation, Ann Arbor, Michigan, USA) and two data sets were prepared. The first data set included only the coding sequences of *s6pdh*. To prepare this data set we used the View Translation command in Sequencher to find the exon/intron boundaries and edited out the intron sequences using the deduced amino acid sequence of *Malus domestica* as a guide. These DNA coding sequences were imported into Se-Al version 2.0a3 (Rambaut, 2001) and aligned as amino acid sequences using the universal genetic code. The aligned nucleotide sequences were then exported in a NEXUS format. The second data set was prepared by manually aligning the introns from all *Prunus* taxa using the first data set as a scaffold. Intron sequences from outgroup taxa were excluded because they were too divergent and were treated as missing data. Insertions and deletions (indels) introduced by the alignment procedure were treated as missing values.

We used ClustalX (Thompson et al., 1997) to align the ITS, *trnL-trnF*, and the S6PDH (amino acids) data sets. The latter consisted of nine GenBank protein accessions found from the BLAST search, plus the deduced amino acid sequence of S6PDH from *Malus domestica*, *P. ilicifolia*, and *P. dulcis*. Six indels, two from the ITS data set and four from the *trnL-trnF* alignment, were scored as separate characters. We took a conservative approach to scoring indels as additional characters, considering only those that were more than two sites long with unequivocally aligned boundaries, a measure taken by previous authors (Wojciechowski, Sanderson, and Hu, 1999).

Phylogenetic analyses—We used PAUP* 4.0b8 (Swofford, 2001) to find most parsimonious trees (MPT) and trees of maximum likelihood (MLT). Unless otherwise indicated, parsimony searches used a heuristic strategy with 100 replicates of random taxon addition, tree bisection-reconnection (TBR) and no constraints on the maximum numbers of trees to be saved each step. We conducted two different heuristic searches for MLT. In the first place, we used the ''as is'' addition strategy instead of random additions. For the second search we used all trees saved from a parsimony analysis as starting trees. Measures of homoplasy are given as consistency indices excluding uninformative characters (CI*) and retention index (RI). We estimated the statistical support for each node by performing 1000 bootstrap repetitions (Felsenstein, 1985) with simple taxon addition. For the bootstrap analysis of *s6pdh* we saved up to 1000 trees each time. We obtained decay values (Bremer, 1988) for each node by using AutoDecay version 4.0.2 (Eriksson, 1998) and TreeView version 1.6.0 (Page, 1996). For the parsimony analysis of the S6PDHs and related dehydrogenases data set we used a Branch and Bound search strategy and statistical support for each node was calculated by performing 1000 Branch and Bound bootstrap repetitions. We calculated *P* values for incongruence length difference (ILD) tests (Farris et al., 1994a, b) as a measure of incongruence among the three data partitions. The tests were implemented in PAUP* (Swofford, 2001) by running 500 repetitions of the partition homogeneity test, with TBR, and a maximum of 1000 trees saved at each step.

We used Modeltest version 3.04 (Posada and Crandall, 1998, 2001) to select evolution models for the maximum likelihood analyses. Modeltest calculates the operator $\delta = 2(\ln L_1 - \ln L_0)$ (Goldman, 1993) where L_0, the null hypothesis, is a simpler model than L_1. The test evaluates several models of evolution and selects the one that fails to be rejected by an alternative and more complex model. Simpler models, which have fewer parameters, are preferred to more complex ones when there is no significant difference in how well they fit the data because the use of additional parameters introduces a potential source of error (Nei and Kumar, 2000). The models of molecular evolution used in this study are given in the Appendix.

To test the hypothesis that rates of evolution of *s6pdh* are constant in all the lineages of *Prunus* (molecular clock evolution) we used the same operator ($\delta = 2[\ln L_1 - \ln L_0]$), where L_0 is the likelihood when equal rates are forced on all branches and L_1 is the score without such constraint. The outcome was evaluated in a χ^2 distribution with $N - 2$ degrees of freedom, where N = number of taxa, and significance of 0.01 (Felsenstein, 1981; Yang, Goldman, and Friday, 1995; Huelsenbeck and Rannala, 1997). The null hypothesis, i.e., evolution in a clockwise fashion, will fail to be rejected if δ is greater than the value $\chi^2_{0.01}(N - 2)$.

As an estimation of the rate of evolution we calculated the pairwise divergence among *Prunus* species for *s6pdh*, ITS, and the *trnL-trnF* spacer in

PAUP*. We found that different models explain different sets of data significantly better. However, for the purposes of pairwise divergence, we wanted to use a uniform criterion. Thus, we selected a TrN model (Tamura and Nei, 1993), with a gamma shape parameter = 0.5, to calculate the pairwise distances for all three markers.

We carried out the SH (Shimodaira and Hasegawa, 1999) likelihood test of topologies for the trees obtained by ML and MP analyses of the combined data sets. The test was performed in PAUP* (Swofford, 2001) by running 1000 RELL nonparametric bootstrap repetitions using the data that produced the trees being tested. *P* values are then given for each topology and a particular tree topology will fail to be rejected in favor of the ML tree if its *P* value is >0.05 (Shimodaira and Hasegawa, 1999). Likelihood-based tests of alternative phylogenetic hypothesis have been reviewed recently (Goldman, Anderson, and Rodrigo, 2000).

RESULTS

Proteins related to S6PDH—The GenBank protein accessions with the highest similarity to S6PDH are plant mannose $NADP^+$-dependent dehydrogenases. Other accessions with less sequence similarity are fungal and animal proteins described as $NADP^+$-dependent aldo-keto oxydoreductases. We selected sequences from *Apium* (GBAN-AAB97617), *Orobanche* (GBAN-AAG15839), *A. thaliana* (GBAN-AAD23673, GBAN-AAD23674), *Candida* (GBAN-BAA19477), *Aspergillus* (GBAN-AAF61912), *Mus* (GBAN-BAB27586), *Sus* (GBAN-P50578), and *Gallus* (GBAN-CAC40811) plus the deduced amino acid sequence of S6PDH from *Malus domestica*, *P. ilicifolia*, and *P. dulcis* for a phylogenetic analysis. The data set had 345 sites, of which 223 were informative. Twenty-five percent of the amino acids were conserved across all proteins, but the percentage is much higher (49%) among the plant accessions. One most parsimonious tree was recovered (Fig. 2). S6PDH from *P. ilicifolia* and *P. dulcis* are sister to each other and S6PDH from *Malus domestica* is sister to them.

Prunus s6pdh—We attempted to amplify the entire gene in a single PCR from a subset of taxa (*P. dulcis*, *P. ilicifolia*, *P. maritima*) without success. We obtained almost full sequences of *s6pdh* from *P. armeniaca*, *P. ilicifolia*, and *P. dulcis* by using two combinations of primers. The first one (*s6pdh*-a/*77L25*) amplifies intron 1 and part of exons 1 and 2, and the second combination (*s6pdh*-k/*s6pdh*-p) amplifies the region between exons 2 and 6 (Fig. 1). The primer *s6pdh*-p was created to match the sequence of *P. dulcis* because amplifications with *s6pdh*-j were inconsistent. Intron 1 in *Prunus* varies in length between 0.7 kb and 0.8 kb. Amplification of intron 1 from *P. ilicifolia* yielded two bands of different molecular mass. Because intron 1 was highly divergent and difficult to align among those three species, we decided to use the region between exons 2 and 6 to generate data for the phylogenetic analyses. Therefore, all the analyses that follow were performed using sequences from the region spanning exons 2–6 of *s6pdh.*

The PCR amplification with *s6pdh*-k and *s6pdh*-p always yielded a single band, as visualized in agarose gels, of approximately 1.3 kb. The *s6pdh* sequences obtained from *Prunus* and the other Rosaceae genera studied here have the same organization as in *Malus domestica*. The sequences varied in length from 1218 base pairs (bp) in *Oemleria* to 1309 bp in *P. maritima* and *P. salicina*, and the gene is organized into 6 exons and 5 introns (Fig. 1). We also confirmed that the TC repeat in intron 3 is present in all species of *Prunus*.

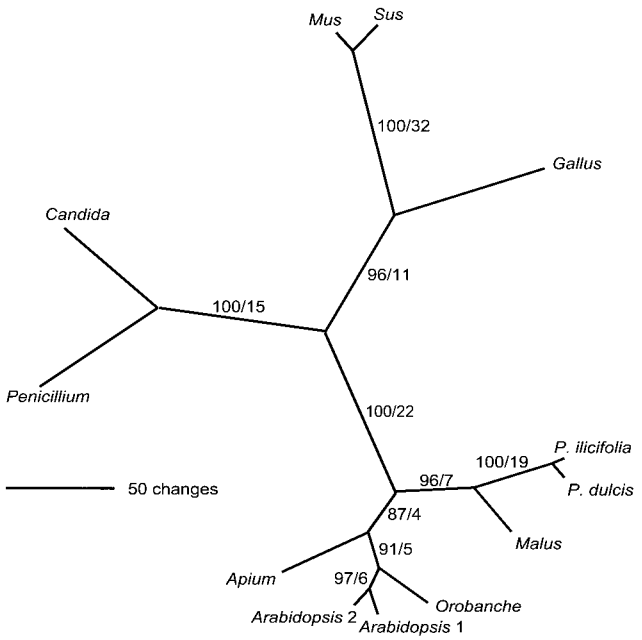

Fig. 2. Most parsimonious tree of length = 809, CI* = 0.83, and RI = 0.82 resulting from analysis of deduced amino acid sequences of S6PDH from *Malus domestica*, *P. dulcis*, *P. ilicifolia*, and nine accessions from GenBank. Branch lengths are proportional to changes. Bootstrap values and decay indexes are shown along the branches. See the *American Journal of Botany* supplementary data web site (http://ajbsupp.botany.org/v89/) for GenBank accession numbers.

Southern blotting of s6pdh—The blots corresponding to *P. persica*, *P. dulcis*, *Spiraea betulifolia* Pall., *Aruncus dioicus* (Walter) Fernald, and *Neillia sinensis* Oliv. (Fig. 3A) indicate that in most cases one region strongly hybridizes to the probe, represented by a band with strong signal in the cases of *Spiraea* (lanes 1–4), *Prunus persica* (5–8), *P. dulcis* (9–10), *Aruncus* (11–14), and *Neillia* (15–18). The lanes corresponding to *P. persica* have one band with very strong signal (Fig. 3A, lanes 5 and 6) or two bands of approximately equal intensity (lanes 7 and 8). One *Eco*RV site was confirmed in *P. dulcis* by sequencing (Fig. 3B) and its presence in *P. persica* is very likely given that both species have the same *Eco*RV restriction pattern (Fig. 3A). *Prunus persica* and *P. dulcis* have one site each for *Eco*RI, *Dra*I, and *Hind*III (Fig. 3B, C), and this is reflected in their restriction patterns with these enzymes. The fragment represented by the band with the strongest signal in lanes 5 and 6 (*P. persica*) and lane 10 (*P. dulcis*) (Fig. 3A) is present as two bands of lower molecular mass in lanes 7 and 8 (*P. persica*) (Fig. 3A) due to a restriction in the middle of *s6pdh* by *Dra*I and *Hind*III (Fig. 3C). The presence of a high molecular mass band in lane 5 (*P. persica*) was unexpected and may be due to a failure to restrict by *Eco*RI.

Maximum parsimony analysis—Introns from accessions other than *Prunus* were excluded because they were too divergent to be aligned. The aligned data set had 1387 characters, of which 390 were parsimony-informative. There were 234 (17%) informative sites within *Prunus*. Parsimony analysis of this data set yielded 273 MPT (Fig. 4A) of length 1198, CI* = 0.58, and RI = 0.81. Among the outgroups, *Gillenia* is sister to Maloideae, and *Oemleria* and *Exochorda* are sister to each other, as are *Sorbaria* and *Chamaebatiaria*. The clade

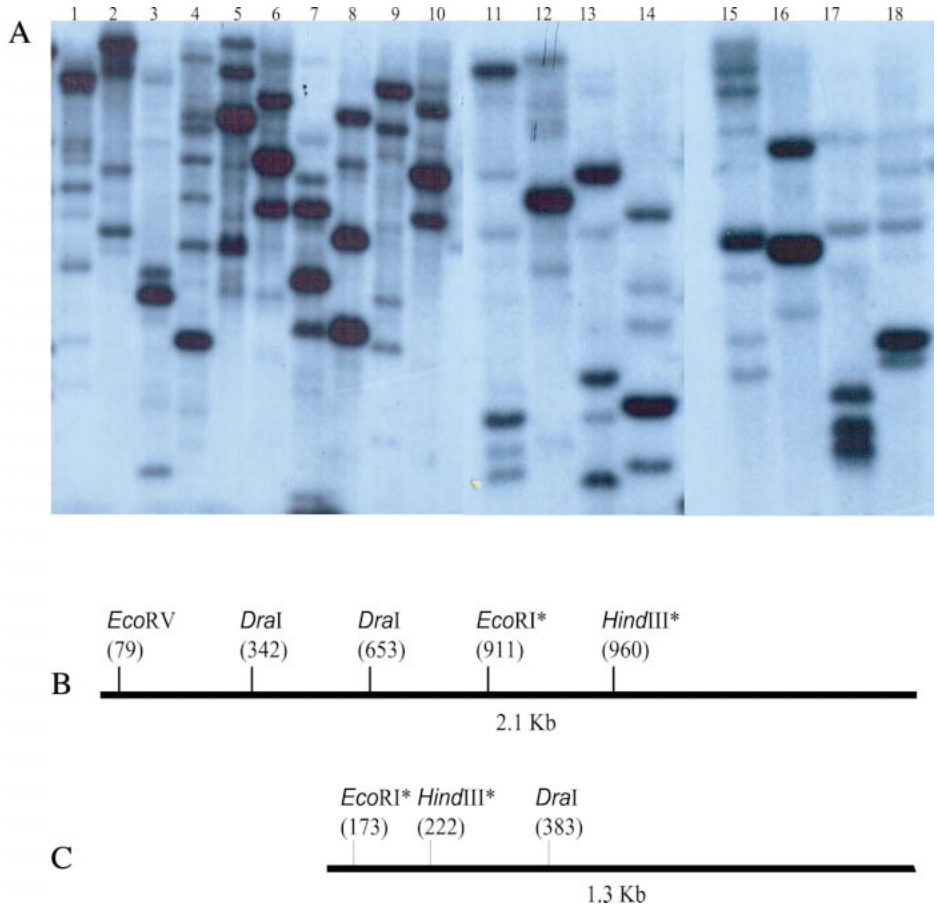

Fig. 3. Analysis of *s6pdh* copy number in Rosaceae. (A) Results from Southern blot hybridization with a radioactively labeled *Malus domestica* cDNA *s6pdh* probe. The taxa are *Spiraea betulifolia* (lanes 1–4), *Prunus persica* (5–8), *P. dulcis* (9–10), *Aruncus dioicus* (11–14), and *Neillia* sp. (15–18). Restriction enzymes are *Eco*RI (lanes 1, 5, 9, 11, and 15), *Eco*RV (2, 6, 10, 12, and 16), *Dra*I (3, 7, 13, and 17), and *Hind*III (4, 8, 14, and 18). (B) Restriction map of *P. dulcis*. (C) Restriction map of *P. persica*. An asterisk (*) marks restriction sites present in all taxa.

Oemleria-Exochorda is not sister to *Prunus*. The genus *Prunus* is strongly supported as monophyletic and the consensus tree shows a monophyletic subgenus *Prunus*, which includes *P. besseyi*, *P. tomentosa*, and *P. microcarpa* (species traditionally classified in section *Microcerasus* of subgenus *Cerasus*). The position of *P. fasciculata*, subgenus *Emplectocladus*, is unresolved. Subgenus *Cerasus*, represented by *P. fruticosa*, *P. takesimensis*, and *P. emarginata*, is monophyletic. Subgenus *Amygdalus*, represented by *P. dulcis* and *P. persica*, is not monophyletic in the consensus tree but had 80% support from the bootstrap analysis.

Sequences from different sources (PCR and different clones) that correspond to the same accession are monophyletic with two exceptions that will be discussed later. The direct PCR sequences from *P. maritima* and *P. armeniaca* are sister to their respective cloned sequences. Among clones 1, 2, 3, 5 and the direct PCR sequence from *P. maritima*, there are eight polymorphic coding sites, and all are synonymous mutations. In clone 4, there is a change from T to C at the GT 5′ boundary of intron 5. The two coding sequences from *P. armeniaca* differ at one site, also a synonymous mutation. Independent PCR reactions from the same accession produced sister sequences in the case of *P. fasciculata* (792-3, 792-D) and *P. andersonii* (796-3, 796-G). The coding sequences of the two *P. andersonii* clones differ in a nonsynonymous mutation (AGU vs. AGA) in exon 1.

We found divergent sequences in *P. caroliniana* and *P. emarginata*. The direct PCR sequence and one of the two sequenced clones of *P. caroliniana* (clone ''A'') are most closely related to *P. ilicifolia* but a second clone (*P. caroliniana*-2) is nested within a clade composed of *P. virginiana*, *P. laurocerasus*, *P. lusitanica*, and a *P. emarginata* clone (Fig. 4A). The two divergent *P. caroliniana* clones differ at 16 out of 251 amino acids. The *P. emarginata* ''direct PCR'' sequence was sister to *P. takesimensis*-*P. fruticosa*. We repeated sequencing of *s6pdh* from a different accession of *P. emarginata* (''994'' in Fig. 4A) and found four clones to form a monophyletic group unrelated to any of the other two *P. emarginata* sequences.

Maximum likelihood analysis—In order to reduce computational time in all analyses of ML we included only one sequence from each species of *Prunus*. For *P. caroliniana* and *P. emarginata* we included one from each group of divergent sequences. We further reduced the number of taxa in this step by excluding *Chamabaetiaria*, *Holodiscus*, and *Exochorda* because their sequences were incomplete (see Materials and methods). After running Modeltest we selected a general time

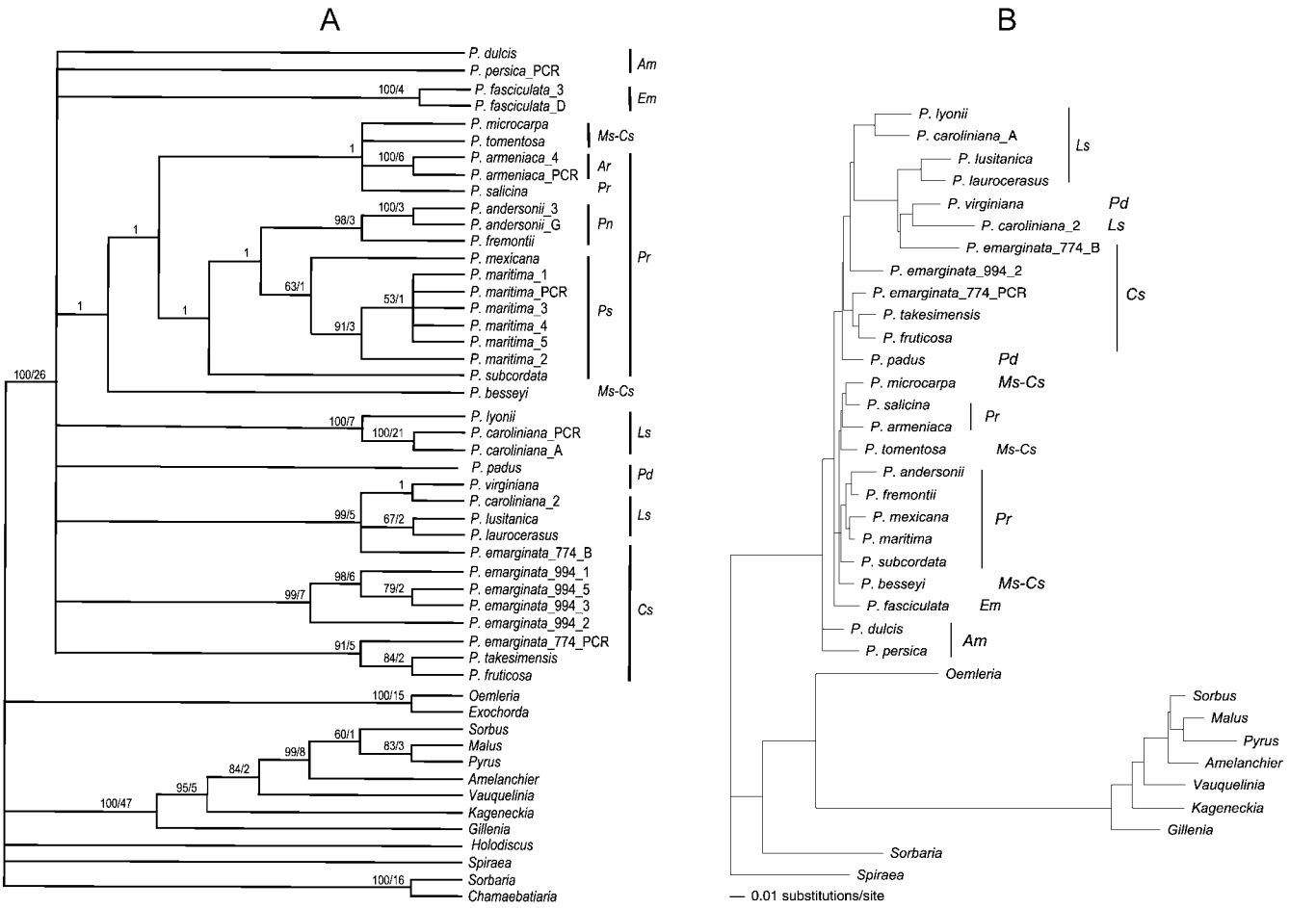

Fig. 4. Results from parsimony and maximum likelihood analyses of *s6pdh* sequences. (A) Strict consensus of 273 MPT (length = 1198, CI* = 0.58, RI = 0.81) resulting from parsimony analysis of *s6pdh* sequence data. Numbers above branches are bootstrap values (when greater than 50%)/decay indexes. (B) Maximum likelihood tree of score −Ln = 7720.96 obtained by analysis of *s6pdh* sequence data. Lengths of branches are proportional to changes. Notice that branches length of outgroups are underestimated because intron sequences were not included. Subgenera and sections are indicated in the right margin.

Figure abbreviations: Am, *Amygdalus*; Cs, *Cerasus*; Em, *Emplectocladus*; Ls, *Laurocerasus*; Ms, *Microcerasus*; Pd, *Padus*; Pn, *Penarmeniaca*; Pr, *Prunus*.

reversible model (Tamura and Nei, 1993) (see Appendix for parameters and models used for all maximum likelihood analyses). The first search, using the ''as is'' addition sequence, recovered one MLT (Fig. 4B) (score −Ln = 7720.96), which shows no topological conflict with the strict consensus of MPT. A maximum parsimony analysis of this reduced data set found 98 MPT. When we swapped all 98 MPT using a maximum likelihood criterion we recovered the same tree as in the first ML search (Fig. 4B). There are no conflicts between the MPT and MLT, except that the MLT is more resolved. *Prunus persica* and *P. dulcis*, subgenus *Amygdalus*, are unresolved in a polytomy with the rest of the genus. The rest of the genus is divided into three lineages joined in a polytomy: *P. fasciculata* (subgenus *Emplectocladus*), subgenus *Prunus* (including *Microcerasus*), and a clade composed of subgenera *Cerasus*, *Padus*, and *Laurocerasus*.

TABLE 2. Rates of evolution of *s6pdh* (exons and introns), ITS1, ITS2, and the *trnL-trnF* spacer measured as the mean number of substitutions per nucleotide site among species of *Prunus*. Distances were calculated using a TrN model (Tamura and Nei, 1993) with gamma shape parameter = 0.5.

Marker	Mean divergence	Maximum
s6pdh exons	0.03	0.068
intron 2	0.067	0.21
intron 3	0.079	0.42
intron 4	0.14	0.7
intron 5	0.076	0.23
ITS1	0.05	0.11
ITS2	0.062	0.12
trnL-trnF spacer	0.021	0.076

Rates of evolution in Prunus—We calculated the means of pairwise divergences for the coding regions of *s6pdh*, all sequenced introns of *s6pdh*, ITS, and *trnL-trnF* among species of *Prunus* as estimations of the rates of evolution (Table 2). We did not attempt to calibrate a molecular clock using those markers because the rates depart significantly from a molecular clock (see below). The highest divergences are found in intron 4 of *s6pdh*, which shows approximately twice as much divergence as the other *s6pdh* introns and both ITS1 and ITS2. The lowest average divergence occurred in the chloroplast *trnL-trnF* spacer. Coding regions of *s6pdh* have a slightly higher rate than *trnL-trnF*, whereas the introns (with the exception of intron 4) have a rate comparable to ITS.

TABLE 3. Results from the molecular clock tests. The scores are given as −Ln (likelihood). See Appendix for models and parameters used for each marker. A constant rate of evolution cannot be rejected for the *trnL-trnF* spacer.

Data set	No molecular clock	Molecular clock enforced	δ	χ^2 0.01[a]
s6pdh	6736.0	6790.8	108	46.64
ITS	3003.4	3027.3	47.8	37.57
trnL-trnF	1317.8	1333.4	31.2	37.57

[a] The degrees of freedom are 26 for *s6pdh* and 20 for ITS and *trnL-trnF*.

Molecular clock test—For the molecular clock test of *s6pdh* we included the same number of ingroup taxa as in the analysis of maximum likelihood trees minus *P. emarginata*-B. As outgroups we included *Oemleria*, *Spiraea*, *Gillenia*, and *Sorbaria*, thus $N = 28$ and the df = 26. The topology of the trees obtained when the molecular clock was enforced is very different from the one with no such constraint in that the subgenera *Padus*, *Laurocerasus*, and *Cerasus* are now paraphyletic and sister to subgenera *Prunus* and *Amygdalus* (tree not shown). The hypothesis that branches are not significantly different across the tree (molecular clock evolution) is rejected with a significance of 0.01 (Table 3) for *s6pdh* and ITS but not for the *trnL-trnF* spacer. The rates are greater in subgenera *Padus* and *Laurocerasus*, and *Cerasus* for all three markers (Fig. 4B; Fig. 4 in Bortiri et al., 2001).

Combined data sets—The ILD test indicated that there is some incongruence in the partitions ITS vs. *trnL-trnF* ($P = 0.018$) and ITS vs. *s6pdh* ($P = 0.02$). However, we decided to combine the three data sets for several reasons. In the first place, the null hypothesis of congruence could not be rejected with a significance level of 0.01 (Cunningham, 1997). In addition, it has been reported that combination of partitions found to be incongruent by the ILD test can still increase phylogenetic accuracy (Joy and Conn, 2001), especially when the conflict is between nodes that show poor support in the separate analyses (Krzywinski, Wilkerson, and Besansky, 2001). In some cases combining incongruent data sets is necessary to recover the true relationships (Yoder, Irwin, and Payseur, 2001). Finally, we advocate analyses of combined partitions in addition to separate analyses on philosophical grounds, since combined analyses maximize the phylogenetic information available and are most consistent with the principle of maximum parsimony (Nixon and Carpenter, 1996).

For the analyses of combined ITS, *trnL-trnF*, and *s6pdh* data sets we included four outgroups (*Oemleria*, *Exochorda*, *Spiraea*, and *Sorbaria*) and 22 species of *Prunus*. We excluded *P. caroliniana*-2 and the two *P. emarginata* sequences (774-B and 994) because we believe they are divergent paralogs (see DISCUSSION). The matrix had 2760 characters, of which 384 characters were parsimony-informative, and 226 were informative among the ingroup (*s6pdh*: 148, ITS: 60, *trnL-trnF*: 18). The heuristic search recovered nine MPT of length = 1592, CI* = 0.58, and RI = 0.61 (Fig. 5A). In the strict consensus *Prunus* is divided into two clades. One composed of subgenera *Cerasus*, *Laurocerasus*, and *Padus*. In the second clade the subgenera *Emplectocladus*, *Prunus* (including *Microcerasus* and *Penarmeniaca*), and *Amygdalus* are joined in a polytomy.

One MLT (Fig. 5B) of score −Ln = 12 056.56 was recovered using a TrN + I + G (Tamura and Nei, 1993) model of evolution on the combined data set (Appendix). The MLT has no conflict with the consensus of the nine MPT and is unresolved for *P. fasciculata* and subgenera *Amygdalus* and *Prunus*. The likelihoods of the nine MPT, evaluated under maximum likelihood settings used to obtain the MLT, ranged from 12 058.78 (tree in Fig. 5A) to 12 061.18. The SH test indicated that there is no significant difference in the likelihood between the trees obtained by parsimony and likelihood analyses, with *P* values ranging between 0.93 and 0.88.

DISCUSSION

Copy number in s6pdh—The use of low-copy number genes for phylogenetic purposes has seen a rapid increase over the last few years. In part this is in response to a need for additional data, especially in cases where existing data sets do not provide enough resolution. Another reason is a need to test phylogenetic hypotheses derived with one marker, such as ITS, with other genes. Some examples of phylogenetic studies using low-copy number genes include the use of *Adh* genes (alcohol dehydrogenases) in Poaceae (Gaut et al., 1999) and *Paeonia* (Sang, Donoghue, and Zhang, 1997), *PgiC* in *Clarkia* (Gottlieb and Ford, 1996), GBSSI, or *waxy*, in the Rosaceae (Evans et al., 2000), and phytochromes A and C in angiosperms (Mathews and Donoghue, 1999).

The *s6pdh* gene belongs to a family of genes that encode $NADP^+$-dependent aldo-keto reductases. Related enzymes are found not only among plants but also in animals and fungi. Mannitol-6-phosphate dehydrogenases (M6PDH) from *Apium*, *Arabidopsis*, and *Orobanche* are the proteins most closely related to S6PDH of Rosaceae. M6PDH catalyzes the reduction of mannose into mannitol, a reaction equivalent to the reduction of glucose into sorbitol by S6PDH. Mannitol is the most common sugar alcohol in angiosperms, while sorbitol is produced by Rosaceae (Zimmerman and Ziegler, 1975; Loescher, 1987), with the exception of Rosoideae sensu stricto (Fig. 1 in Potter et al., 2002).

We do not know whether *m6pdh*, in addition to *s6pdh*, is present in the Rosaceae, but based on the information from the Southern blots it is evident that other loci with sequence similarity to *s6pdh* are found in the genome of several sorbitol-producing Rosaceae. This suggests that *s6pdh* originated via duplication of *m6pdh*-like gene in the common ancestor of the sorbitol-producing Rosaceae. This would explain the presence of more than one region with sequence similarity to *s6pdh* in several Rosaceae (Fig. 3A), whereas *Fragaria* (a genus of Rosoideae sensu stricto) and *Rhamnus* (Rhamnaceae) had only one weak band (data not shown). It is important to point out that the presence of an *m6pdh*-like gene does not mean the plant produces significant amounts of mannitol, i.e., neither Rosoideae sensu stricto (s.s.) nor Rhamnaceae have been found to produce sugar alcohols (Zimmerman and Ziegler, 1975; Wallart, 1980).

We detected more than one region with sequence similarity to *s6pdh* in the genome of several Rosaceae by Southern blotting, and we found divergent sequences in *P. caroliniana* and *P. emarginata*. However, sequencing of several clones of *P. maritima*, as well as repeating PCR and cloning on some accessions (*P. fasciculata*, *P. armeniaca*, *P. andersonii*), failed to yield paralogous genes in those taxa. All of the sequences examined had the characteristic TC repeat in intron 3. Previous experiments involving the use of this microsatellite in *P. dulcis*

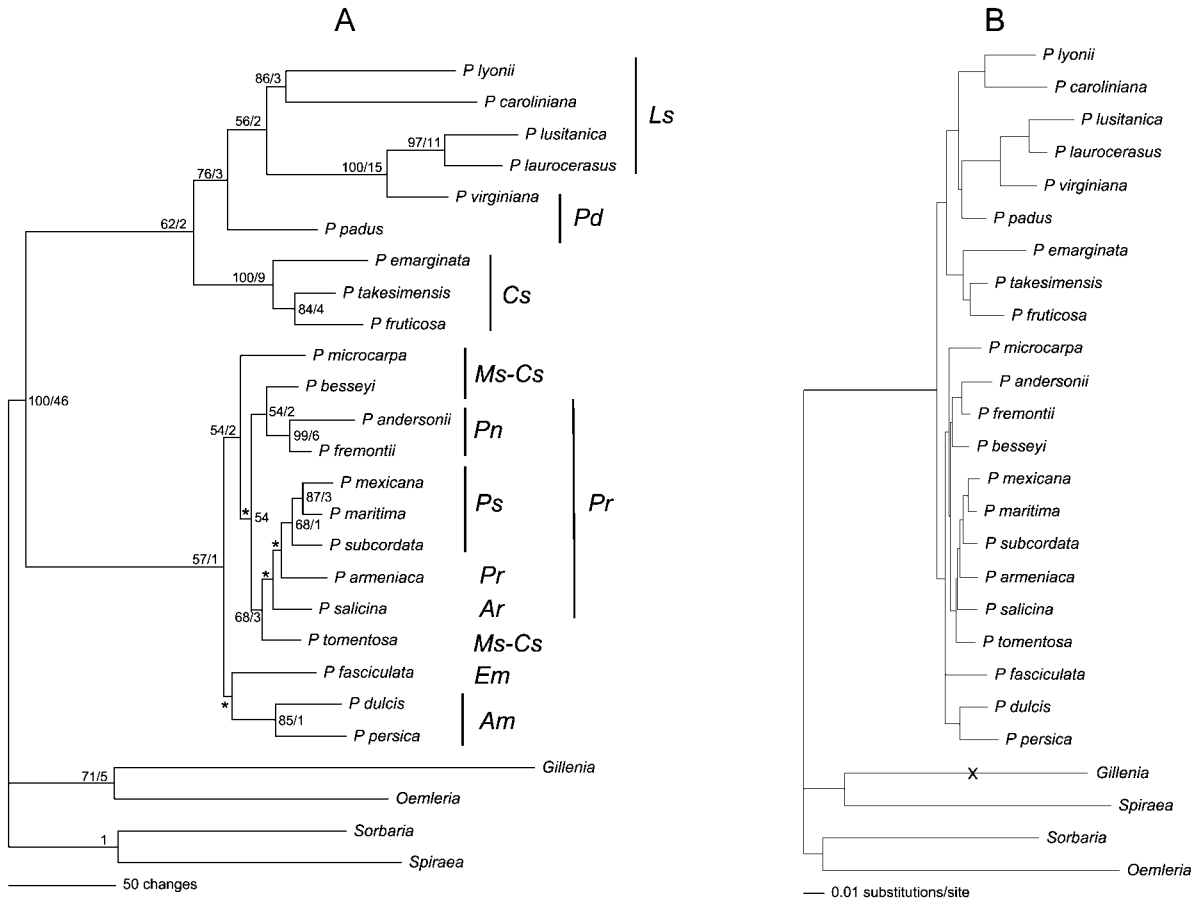

Fig. 5. Results from analyses of ITS, *trnL-trnF*, and *s6pdh* combined data set. (A) One of the nine most parsimonious trees of length = 1611, CI* = 0.58, RI = 0.6, and likelihood score −Ln = 12 058.78. Bootstrap (when >50%) and decay values are shown next to each node. An asterisk indicates branches that collapse in the strict consensus. Subgenera and sections are shown in the right margin. (B) Maximum likelihood tree of score −Ln = 12 056.56. Length of branches of both trees are proportional to changes but the branch leading to *Gillenia* in (B) was reduced to 50% its actual length.

Figure abbreviations: Am, *Amygdalus*; Cs, *Cerasus*; Em, *Emplectocladus*; Ls, *Laurocerasus*; Ms, *Microcerasus*; Pd, *Padus*; Pn, *Penarmeniaca*; Pr, *Prunus*.

and *P. persica* (S. Arulsekar, personal communication) are consistent with *s6pdh* being a single-copy gene in those species.

The presence of divergent paralogs in *P. caroliniana* (subgenus *Laurocerasus*) and *P. emarginata* (subgenus *Cerasus*) could be due to gene duplication by polyploidization. Alternatively, the paralogs could have originated by a duplication of the genomic region containing the gene, as in the case of *m6pdh* of *A. thaliana* (Lin et al., 1999). Several species of *Prunus* are reportedly polyploids (Darlington and Wylie, 1955), in particular all examined species of subgenera *Laurocerasus*, *Padus*, and several crop species. *Prunus caroliniana* has been reported as tetraploid (Robertson, 1974). We did not find information on the chromosome numbers of *P. emarginata* and therefore we cannot rule out the possibility it is a polyploid. The presence of two bands as a result of PCR amplification of intron 1 of *P. ilicifolia*, a tetraploid species (Parfitt et al., 1990), may represent two divergent paralogs resulting from a polyploidization event. We cannot confirm this because we sequenced only one clone of the *s6pdh* PCR product for *P. ilicifolia*.

In summary, the evidence here presented, from southern blotting and sequencing, indicates that *s6pdh* is a low-copy gene. More than one region is likely to be amplified from species of subgenera *Padus* and *Laurocerasus*, which are polyploids. On the other hand, amplification from diploid species, e.g., subgenera *Emplectocladus*, *Prunus*, and *Amygdalus*, should be straightforward. One exception would be polyploid plum species (*P. domestica*, *P. spinosa*), where multiple copies should be found.

The occurrence of an *s6pdh* duplication event in *Prunus* can have its advantages. One of the problems with reconstructing a robust phylogeny of *Prunus* is the genetic distance to other Rosaceae genera, which complicates homology assessment through alignment. The presence of a gene duplication in *Prunus* could avoid this problem altogether by providing the root of the tree at the duplication event (Gottlieb and Ford, 1996; Gaut et al., 1999; Mathews and Donoghue, 1999; Vieira, Vieira, and Charlesworth, 1999; Citerne, Möller, and Cronk, 2000). It is important to note that this will be the case if the duplication is a synapomorphy of the whole genus.

Molecular evolution and utility of s6pdh—Sequences from *s6pdh* have more than twice the number of informative sites as ITS among species of *Prunus* (153 vs. 60) but are slightly less homoplastic (CI is 0.6 in *s6pdh* and 0.57 in ITS). While the coding sequences of *s6pdh* evolve at a rate similar to those in the *trnL-trnF* spacer, the introns have a significantly faster

rate, comparable to ITS or higher, as in the case of intron 4. Despite rates of evolution similar to or higher than ITS and comparable levels of homoplasy, *s6pdh* failed to provide significant support for the branching order of early lineages in the genus. ITS and *trnL-trnF* also show branches of zero or near-zero length in this region of the tree. A similar phenomenon, i.e., short branches corresponding to the early-splitting lineages, has been observed in other groups of angiosperms (Rohwer, 2000; Fishbein et al., 2001; Potter et al., 2002) and attributed to rapid radiation. Nevertheless, *s6pdh* supported the same general topology as ITS and the combined ITS and *trnL-trnF* data sets (Lee and Wen, 2001; Bortiri et al., 2001), which will be discussed below.

Rates of evolution of *s6pdh* are not constant across all branches, as demonstrated by the failure of the data to be explained in terms of a molecular clock. This unequal rate is most likely due to the long branches in the clade including subgenera *Padus* and *Laurocerasus* (Fig. 5A–B). We have not investigated the source of these uneven evolution rates but differences in generation time should not be a reason because all species of *Prunus* are perennials (trees and shrubs). We considered the possibility of pseudogenes being the cause of this acceleration because many species of subgenera *Padus* and *Laurocerasus* are polyploids. However, we think this is unlikely for several reasons. In the first place, all sequences appear to encode functional proteins, i.e., there are no stop codons anywhere in the coding sequences, and their deduced amino acid sequences are highly conserved with respect to the rest of the genus. Secondly, the two divergent copies of *P. caroliniana* show a very high rate of evolution compared to the rest of *Prunus* (Figs. 4B, 5B). We would expect only one, the silenced copy, to have a significantly higher rate as a result of lower selective pressure. Finally, the same pattern, i.e., acceleration of the rate of molecular evolution in *Padus* and *Laurocerasus*, is seen in both ITS and *trnL-trnF* (data not shown, but see Fig. 4 in Bortiri et al., 2001), although to a lesser extent in the cpDNA.

Phylogenetic relationships—In agreement with other molecular data (Bortiri et al., 2001; Potter et al., 2002), *s6pdh* does not support a close relationship between *Prunus* and *Oemleria-Exochorda*. On the other hand, *s6pdh* found the clades *Oemleria-Exochorda* and *Sorbaria-Chamaebatiaria*, which have been strongly supported by other molecular data (Morgan et al., 1994; Potter et al., 2002). Similarly, *Gillenia* is sister to Maloideae according to *s6pdh*, a relationship also obtained by analysis of cpDNA (Potter et al., 2002).

There were no conflicts between the topologies obtained by parsimony and maximum likelihood analyses of *s6pdh*, and the phylogenies of *Prunus* based on this gene were congruent with those based on ITS and *trnL-trnF* in most respects. The consensus is that subgenus *Cerasus* is nested within subgenera *Padus* and *Laurocerasus*, all the species of section *Microcerasus* included in these studies are nested within subgenus *Prunus*, and subgenera *Amygdalus* and *Emplectocladus* are sister to subgenus *Prunus*. In our previous work on *Prunus* we found a discrepancy regarding the phylogenetic relationships of subgenus *Cerasus* (Bortiri et al., 2001). Phylogeny reconstruction with the cpDNA marker *trnL-trnF* strongly supported a clade of subgenera *Padus* and *Laurocerasus* and placed *Cerasus* in a group with the rest of *Prunus*. On the other hand, subgenus *Cerasus* was nested in a clade with subgenera *Padus* and *Laurocerasus* based on data from ITS alone (Lee and Wen, 2001; Bortiri et al., 2001). Both maximum parsimony and likelihood analyses of *s6pdh* agree with the latter scenario and suggest that hybridization may have taken place in the evolution of species of *Prunus*. However, support from *s6pdh* was relatively weak because the node (*Cerasus*, *Padus*, *Laurocerasus*), which received 66% in the bootstrap analysis, collapsed in the strict consensus (Fig. 4A).

In agreement with ITS, *s6pdh* reconstructed a monophyletic subgenus *Prunus*, which includes not only the sections traditionally assigned to it (*Prunus*—Old World plums, *Prunocerasus*—New World plums, and *Armeniaca*—apricots), but also sections *Microcerasus* (formerly in subgenus *Cerasus*) and *Penarmeniaca*. The latter was assigned to subgenus *Prunus* by Mason (1913) but not included in the treatment of *Prunus* by Rehder (1940). Section *Microcerasus*, represented here by *P. besseyi*, *P. microcarpa*, and *P. tomentosa*, a group of shrubs from the Old and New World, has been previously classified in subgenus *Cerasus* (Rehder, 1940; Yü et al., 1986) or as a separate subgenus, *Lithocerasus* (Ingram, 1948). Isozyme data (Mowrey and Werner, 1990) and hybridization studies (Kataoka, Sugiura, and Tomana, 1988), in addition to our previous work on ITS and *trnL-trnF* (Bortiri et al., 2001), have pointed out its closer relationships to subgenus *Prunus* rather than to *Cerasus*. We are presently working on expanding the sampling within this group but the data so far indicates that *Microcerasus* is an artificial group because it is polyphyletic in all analyses, with the exception of a *P. tomentosa-P. bifrons* clade found by the ITS and *trnL-trnF* combined data sets (Bortiri et al., 2001).

Section *Penarmeniaca* includes *P. andersonii* and *P. fremontii*, two dry-fruited species from southwestern USA and Mexico. Sequences from *s6pdh* place them as sister to *P. maritima* and *P. mexicana*, two plum species from eastern North America (Fig. 4A–B), but ITS (Bortiri et al., 2001) and the combined data set suggest a sister relationships with *P. besseyi* (Fig. 5A–B), another North American species. *Prunus besseyi* was found to be sister to the rest of subgenus *Prunus* by *s6pdh* (Fig. 4A–B). Thus, there are some discrepancies between the two nuclear markers (ITS and *s6pdh*) in the topologies they support for these species. This conflict between data sets could be due to lineage sorting, horizontal transfer, or gene duplication and extinction (Page, 1993; Maddison, 1995, 1997; Wendel and Doyle, 1998). It is also possible that the conflict is caused by convergent evolution (''long-branch attraction''). However, we think this is less likely because maximum likelihood, which is less prone to obtain the wrong phylogeny under these conditions of unequal rate of evolution (see below), reconstructed the same phylogeny as parsimony.

Prunus fasciculata, another dry-fruited species from southwestern USA and Mexico, has been considered related to the almonds (subgenus *Amygdalus*) by Rehder (1940). However, results from *s6pdh*, in accordance with our previous study based on ITS and *trnL-trnF* (Bortiri et al., 2001), suggest that *P. fasciculata* diverged early in the evolution of *Prunus* and is not a member of *Amygdalus*. Its phylogenetic position as sister to a *Prunus-Amygdalus-Penarmeniaca-Microcerasus* alliance (Bortiri et al., 2001) is only weakly supported by the combined data set (Fig. 5A–B). In preliminary analyses of *s6pdh*, *P. fasciculata* was found to be sister to *Cerasus*, *Laurocerasus*, and *Padus*. We tested this hypothesis by creating a constrained tree with that topology and found that its likelihood was not significantly lower than the MLT and the nine

MPT ($P = 0.818$). Therefore, the precise phylogenetic position of *P. fasciculata* is still unresolved.

The monophyly of subgenus *Amygdalus* (*P. dulcis* and *P. persica*, almond and peach, respectively) is supported by ITS + *trnL*-trnF + *s6pdh* (Fig. 5A–B), and its position as sister to *Prunus* is in accordance with previous results (Bortiri et al., 2001; Lee and Wen, 2001). Lee and Wen (2001) found *Amygdalus* to be paraphyletic, contrary to our findings (Bortiri et al., 2001). The discrepancy is explained by the inclusion in their study of *P. tenella* Batsch., a species of section *Chamaeamygdalus* Spach. We have recently tested this hypothesis by sequencing ITS and *trnL-trnF* of *P. petunnikovii* Litv., a species closely related to *P. tenella*, and obtained the same results (data not shown) as Lee and Wen (2001); i.e., *Amygdalus*, as treated by Browicz and Zohary (1996) is not monophyletic. Additional sampling would be necessary to test relationships among all sections of subgenus *Amygdalus*.

Effects of phylogenetic methods, amount of data, and number of taxa—The use of maximum likelihood methods for phylogenetic inference has been advocated as a way to avoid inconsistency due to ''long branch attraction'' effect (Felsenstein, 1978; Hendy and Penny, 1989). Several authors have studied the performance of parsimony, likelihood, and other methods under conditions of unequal rates of evolution (Huelsenbeck and Hillis, 1993; Siddall, 1998; Swofford et al., 2001) and the effect of taxon and data addition to improve parsimony accuracy (Kim, 1996; Graybeal, 1998; Poe and Swofford, 1999).

In order to address the concern of reconstructing a misleading phylogeny using only one method, we decided to use parsimony and maximum likelihood and compare the results. Here we show that, despite the presence of unequal rates of evolution in lineages of *Prunus*, both methods reconstruct the same phylogeny and the results are very similar to those obtained previously using a larger number of taxa (Bortiri et al., 2001). *Prunus* is a very large genus with over 200 species, of which we included only 22; however, we took care to select representatives of all groups of species that were previously shown to be monophyletic (Lee and Wen, 2001; Bortiri et al., 2001), thereby reducing the sample size while minimizing the effect on the overall topology. Nevertheless, a large group of species from the tropics is missing here. These species, many of which were formerly classified in genus *Pygeum* (Kalkman, 1965), were found to be sister to genus *Prunus* in a cladistic study of wood anatomy characters (Zhang, 1992). Thus, this group of species is very important to a better understanding of the natural history of *Prunus* because of their geographic distribution (*Prunus* is distributed largely in temperate areas of the Northern Hemisphere) and their putative phylogenetic position. We believe that future studies of the phylogeny of *Prunus* should include these tropical representatives of subgenus *Laurocerasus*, including the species formerly classified in genus *Pygeum*.

To conclude, *s6pdh* appears to be a single copy gene in all diploid species of *Prunus*, with the possible exception in *P. emarginata*. The rates of evolution of exon sequences of *s6pdh* are comparable to *trnL-trnF*, while the introns are closer to ITS, with the exception of intron 4, where the rates are almost twice as high. An acceleration in the rates of all three markers is observed in the *Cerasus-Laurocerasus-Padus* clade. Maximum likelihood and parsimony analyses of *s6pdh* confirmed some findings from previous work with ITS and *trnL-trnF*. An early divergence in the evolution of *Prunus* separated *Laurocerasus* and *Padus* from the rest of genus. Subgenus *Prunus* includes section *Microcerasus*. The two nuclear markers, ITS and *s6pdh*, support slightly different phylogenies within subgenus *Prunus*, which are most likely due to lineage sorting. Future work in this genus should concentrate on species from the tropics, classified in subgenus *Laurocerasus*, and, to a less degree, on additional sampling of section *Microcerasus*.

LITERATURE CITED

BAINS, H. S., R. TAO, S. L. URATSU, AND A. M. DANDEKAR. 1998. Genomic nucleotide sequence of NADP sorbitol-6-phosphate dehydrogenase (NADP-S6PDH) gene from apple (accession No. AF057134) (PGR98-193). *Plant Physiology* 118: 1533.

BORTIRI, E., S.-H. OH, J. JIANG, S. BAGGETT, A. GRANGER, C. WEEKS, M. BUCKINGHAM, D. POTTER, AND D. E. PARFITT. 2001. Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Systematic Botany* 26: 797–807.

BRAKO, L., AND J. L. ZARUCCHI. 1993. Catalogue of the flowering plants and gymnosperms of Peru. Missouri Botanical Garden, St. Louis, Missouri, USA.

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

BROWICZ, K., AND D. ZOHARY. 1996. The genus *Amygdalus* L. (Rosaceae): species relationships, distribution, and evolution under domestication. *Genetic Resources and Crop Evolution* 43: 229–247.

CITERNE, H. L., M. MÖLLER, AND Q. C. B. CRONK. 2000. Diversity of *cycloidea*-like genes in Gesneriaceae in relation to floral symmetry. *Annals of Botany* 86: 167–176.

CUNNINGHAM, C. W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Systematic Biology* 46: 733–740.

DARLINGTON, C. D., AND A. P. WYLIE. 1955. Chromosome atlas of flowering plants. George Allen and Unwin, London, UK.

DELLAPORTA, S. L., J. WOOD, AND J. B. HICKS. 1983. A plant DNA minipreparation: version II. *Plant Molecular Biology Reporter* 1(4): 19–21.

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

ELIAS, T. S. 1980. Trees of North America. Van Nostrand Reinhold, New York, New York, USA.

ERIKSSON, T. 1998. AutoDecay, version 4. Department of Botany, Stockholm University, Stockholm. Available at: http://www.bergianska.se/personal/TorstenE/.

EVANS, R. C., L. A. ALICE, C. S. CAMPBELL, E. A. KELLOGG, AND T. A. DICKINSON. 2000. The granule-bound synthase (GBSSI) gene in the Rosaceae: multiple loci and phylogenetic utility. *Molecular Phylogenetics and Evolution* 17: 388–400.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994a. Permutations. *Cladistics* 10: 65–76.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994b. Testing significance of incongruence. *Cladistics* 10: 315–319.

FEDOROV, A. A., V. L. KOMAROV, K. F. KOSTINA, N. V. KOVALEV, A. N. KRISHTOFOVICH, I. A. LINCHEVSKII, AND A. I. POYARKOVA. 1941. Rosaceae—Rosoideae, Prunoideae. *In* V. L. Komarov [ed.], Flora of the U.S.S.R, vol. 10, 380–448. Botanical Institute of the Academy of Sciences of the USSR, Moscow, Russia.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368–376.

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

FISHBEIN, M., C. HIBSCH-JETTER, D. E. SOLTIS, AND L. HUFFORD. 2001. Phylogeny of Saxifragales (Angiosperms, Eudicots): analysis of a rapid, ancient radiation. *Systematic Biology* 50: 817–847.

GAUT, B. S., A. S. PEEK, B. R. MORTON, AND M. T. CLEGG. 1999. Patterns of genetic diversification within the *Adh* family in the grasses (Poaceae). *Molecular Biology and Evolution* 16: 1086–1097.

GLADKOVA, V. N. 1972. On the origin of subfamily Maloideae. *Botanicheskii Zhurnal* 57: 42–49.

GOLDBLATT, P. 1976. Cytotaxonomic studies in the tribe Quillajeae (Rosaceae). *Annals of the Missouri Botanical Garden* 63: 200–206.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36: 182–198.

GOLDMAN, N., J. P. ANDERSON, AND A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49: 652–670.

GOTTLIEB, L. D., AND V. S. FORD. 1996. Phylogenetic relationships among the sections of *Clarkia* (Onagraceae) inferred from the nucleotide sequences of *PgiC*. *Systematic Botany* 21: 45–62.

GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Systematic Biology* 47: 9–17.

HENDY, M. D., AND D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.

HIRAI, M. 1981. Purification and characteristics of sorbitol-6-phosphate dehydrogenase from loquat leaves. *Plant Physiology* 67: 221–224.

HUELSENBECK, J. P., AND D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. *Systematic Biology* 42: 247–264.

HUELSENBECK, J. P., AND B. RANNALA. 1997. Phylogenetic methods come of age: testing hypothesis in an evolutionary context. *Science* 276: 227–232.

INGRAM, C. 1948. Ornamental cherries. Country Life Limited, London, UK.

JOY, D. A., AND J. E. CONN. 2001. Molecular and morphological phylogenetic analysis of an insular radiation in Pacific black flies (*Simulium*). *Systematic Biology* 50: 18–38.

KALKMAN, C. 1965. The Old World species of *Prunus* subgenus *Laurocerasus* including those formerly referred to *Pygeum*. *Blumea* 13: 1–174.

KANAYAMA, Y., H. MORI, H. IMASEKI, AND S. YAMAKI. 1992. Nucleotide sequence of a cDNA encoding NADP-sorbitol-6-phosphate dehydrogenase from apple. *Plant Physiology* 100: 819–823.

KANAYAMA, Y., AND S. YAMAKI. 1993. Purification and properties of NADP-dependent sorbitol-6-phosphate dehydrogenase from apple seedlings. *Plant Cell Physiology* 34: 819–823.

KATAOKA, I., A. SUGIURA, AND T. TOMANA. 1988. Interspecific hybridization between *Microcerasus* and other *Prunus* spp. *Journal of the Japanese Society for Horticultural Science* 56: 398–407.

KIM, J. 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Systematic Biology* 45: 363–374.

KRÜSSMANN, G. 1986. Manual of cultivated broad-leaved trees and shrubs, vol. 3. Timber Press, Portland, Oregon, USA.

KRZYWINSKI, J., R. C. WILKERSON, AND N. J. BESANSKY. 2001. Towards understanding Anophelinae (Diptera, Culicidae) phylogeny: insights from nuclear single-copy genes and the weight of evidence. *Systematic Biology* 50: 540–556.

KUITERT, W. 1999. Japanese flowering cherries. Timber Press, Portland, Oregon, USA.

LEE, S., AND J. WEN. 2001. A phylogenetic analysis of *Prunus* and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. *American Journal of Botany* 88: 150–160.

LIN, X., ET AL. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402: 761–768.

LOESCHER, W. H. 1987. Physiology and metabolism of sugar alcohols in higher plants. *Physiology Plantarum* 70: 553–557.

MADDISON, W. P. 1995. Phylogenetic histories within and among species. *In* P. C. Hoch and A. G. Stephenson [eds.], Experimental and molecular approaches to plant biosystematics, 273–287. Missouri Botanical Garden, St. Louis, Missouri, USA.

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

MASON, S. C. 1913. The pubescent-fruited species of *Prunus* of the southwestern states. *Journal of Agricultural Research* 1: 147–179.

MATHEWS, S., AND M. J. DONOGHUE. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947–950.

MCVAUGH, R. 1951. A revision of North American black cherries (*Prunus serotina* Ehrh., and relatives). *Brittonia* 7(4): 279–315.

MOORE, J. N., AND J. R. BALLINGTON, JR. 1990. Genetic resources of temperate fruit and nut crops. International Society for Horticultural Science, Wageningen, The Netherlands.

MORGAN, D. R., D. E. SOLTIS, AND K. R. ROBERTSON. 1994. Systematic and evolutionary implications of *rbcL* sequence variation in Rosaceae. *American Journal of Botany* 81: 890–903.

MOWREY, B. D., AND D. J. WERNER. 1990. Phylogenetic relationships among species of *Prunus* as inferred by isozyme markers. *Theoretical and Applied Genetics* 80: 129–133.

NEI, M., AND S. KUMAR. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, New York, USA.

NIXON, K. C., AND J. M. CARPENTER. 1996. On simultaneous analysis. *Cladistics* 12: 221–241.

PAGE, R. D. M. 1993. Genes, organisms, and areas: the problem of multiple lineages. *Systematic Biology* 42: 77–84.

PAGE, R. D. M. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12: 357–358.

PARFITT, B. D., D. J. PINKAWA, D. RICKEL, D. FILLIPI, B. EGGERS, AND D. J. KEIL. 1990. Documented chromosome numbers 1990: 1. Miscellaneous North American vascular plants. *Sida* 14: 305–308.

POE, S., AND D. L. SWOFFORD. 1999. Taxon sampling revisited. *Nature* 398: 299–300.

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

POSADA, D., AND K. A. CRANDALL. 2001. Selecting the best-fit model of nucleotide substitution. *Systematic Biology* 50: 580–601.

POTTER, D., F. GAO, E. BORTIRI, S.-H. OH, AND S. BAGGETT. 2002. Phylogenetic relationships in Rosaceae from chloroplast *matK* and *trnL-trnF* nucleotide sequence data. *Plant Systematics and Evolution* 231: 77–89.

RAMBAUT, A. 2001. Sequence alignment editor version 2.0. Available from the author at: http://evolve.zoo.ox.ac.uk.

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

ROBERTSON, K. R. 1974. The genera of Rosaceae in the southeastern United States. *Journal of the Arnold Arboretum* 55: 611–662.

ROHWER, J. G. 2000. Toward a phylogenetic classification of the Lauraceae: evidence from *matK* sequences. *Systematic Botany* 25: 60–71.

SAMBROOK, J., E. F. FRITSCH, AND T. MANIATIS. 1989. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

SANG, T., M. J. DONOGHUE, AND D. ZHANG. 1997. Evolution of alcohol dehydrogenases genes in peonies (*Paeonia*): phylogenetic relationships of putative nonhybrid species. *Molecular Biology and Evolution* 14: 994–1007.

SAX, K. 1933. The origin of the Pomoideae. *Proceedings of the American Society of Horticultural Sciences* 30: 147–150.

SHIMODAIRA, H., AND M. HASEGAWA. 1999. Multiple comparison of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.

SIDDALL, M. E. 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris zone. *Cladistics* 14: 209–220.

SWOFFORD, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, Massachusetts, USA.

SWOFFORD, D. L., P. J. WADDELL, J. P. HUELSENBECK, P. G. FOSTER, P. O. LEWIS, AND J. S. ROGERS. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology* 50: 525–539.

TAKHTAJAN, A. 1997. Diversity and classification of flowering plants. Columbia University Press, New York, New York, USA.

TAMURA, K., AND M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526.

TAO, R., S. L. URATSU, AND A. M. DANDEKAR. 1995. Sorbitol synthesis in transgenic tobacco with apple cDNA encoding NADP-dependent sorbitol-6-phosphate dehydrogenase. *Plant Cell Physiology* 36: 525–532.

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* 25: 4876–4882.

VIEIRA, C. P., J. VIEIRA, AND D. CHARLESWORTH. 1999. Evolution of the cycloidea gene family in *Antirrhinum* and *Misopates*. *Molecular Biology and Evolution* 16: 1474–1483.

WALLAART, R. B. 1980. Distribution of sorbitol in Rosaceae. *Biochemistry* 19: 2603–2610.

WEBB, K. L., AND J. W. A. BURLEY. 1962. Sorbitol translocation in apple. *Science* 13: 766.

WENDEL, J. F., AND J. J. DOYLE. 1998. Phylogenetic incongruence: window

into genome history and molecular evolution. *In* D. E. Soltis, P. S. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II: DNA sequencing, 265–296. Kluwer, Boston, Massachusetts, USA.

WIGHT, W. F. 1915. Native American species of *Prunus*. *Bulletin of the U.S. Department of Agriculture* 179: 1–75.

WOJCIECHOWSKI, M. F., M. J. SANDERSON, AND J.-M. HU. 1999. Evidence on the monophyly of *Astragalus* (Fabaceae) and its major subgroups based on nuclear ribosomal DNA ITS and chloroplast DNA *trnL* intron data. *Systematic Botany* 24: 409–437.

YAMAKI, S., AND K. ISHIKAWA. 1986. Roles of four sorbitol related enzymes and invertase in the seasonal alteration of sugar metabolism in apple tissue. *Journal of the American Society of Horticultural Sciences* 111: 134–137.

YANG, Z., N. GOLDMAN, AND A. FRIDAY. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Systematic Biology* 44: 384–399.

YODER, A. D., J. A. IRWIN, AND B. A. PAYSEUR. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. *Systematic Biology* 50: 408–424.

YÜ, T. T., L. T. LU, T. C. KU, C. L. LI, AND S. X. CHEN. 1986. Rosaceae (3), Amygdaloideae. *Flora Reipublicae Popularis Sinicae*, vol. 38. Science Press, Beijing, China.

ZHANG, S.-Y. 1992. Systematic wood anatomy of the Rosaceae. *Blumea* 37: 81–158.

ZIMMERMAN, M. H., AND H. ZIEGLER. 1975. List of sugar and sugar alcohols in sieve-tube exudates. *In* M. H. Zimmerman and J. A. Milburn [eds.], Transport in plants. I. Phloem transport. Springer-Verlag, Heidelberg, Germany.

APPENDIX. Models of evolution used for the different markers and their parameters. Abbreviations used—TrN + G: General time reversible with no invariable sites and gamma shape parameter; TrN + I + G: General time reversible with invariable sites and gamma shape parameter (Tamura and Nei, 1993); F81: Felsenstein (1981).

Model	*s6pdh* TrN + G	s6pdh + ITS + *trnL-trnF* TrN + I + G	ITS TrN + I + G	*trnL-trnF* F81
Base frequencies				
A	0.276	0.2638	0.188	0.3227
C	0.226	0.2381	0.3124	0.166
G	0.219	0.2293	0.3086	0.1436
Substitution rates				All rates equal
A–C	1	1	1	
A–G	2.366	2.098	2.3596	
A–T	1	1	1	
C–G	1	1	1	
C–T	3.409	3.553	5.8368	
Invariable sites	0	0.274	0.364	
Gamma parameter	0.562	0.647	0.611	