Systematic Botany (2013), 38(3): pp. 850–857

DOI 10.1600/036364413X670449

Molecular Phylogeny of *Daucus* (Apiaceae)

David Spooner,[1,5] Percy Rojas,[1] Meredith Bonierbale,[2] Lukas A. Mueller,[3] Manish Srivastav,[4] Douglas Senalik,[1] and Philipp Simon[1]

[1]United States Department of Agriculture-Agricultural Research Service, Vegetable Crops Research Unit, University of Wisconsin, 1575 Linden Drive, Madison, Wisconsin 53706-1590, U. S. A.
[2]International Potato Center, PostOffice Box 1558, Lima, 12, Peru.
[3]Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, New York 14853, U. S. A.
[4]Division of Fruits and Horticultural Technology, Indian Agricultural Research Institute, New Delhi, 110 012, India.
[5]Author for correspondence (david.spooner@ars.usda.gov)

Communicating Editor: Fernando Zuloaga

Abstract—We studied the phylogeny of 22 accessions of *Daucus* and seven accessions in related genera, with DNA sequences from eight nuclear orthologs and one plastid (*psbA/trnH*) region. Maximum parsimony and maximum likelihood analyses of the concatenated data matrix of 7,212 aligned nucleotides provided excellent bootstrap support for many clades. Concordant with prior molecular results *Pseudorlaya pumila* was firmly imbedded within *Daucus*, as was *Margottia gummifera*, a new finding. All accessions of *D. capillifolius*, *D. carota*, and *D. sahariensis* formed a $2n = 18$ clade with all other species within the *Daucus* clade with chromosome numbers of = 20, 22, and 44 (*D. glochidiatus*). Sister to the *D. carota* clade was a clade containing *Margottia gummifera* and *Pseudorlaya pumila*, sister to these species was *D. crinitus*, sister to all the above was *D. muricatus*, and sister to all of the above was a clade containing the remaining *Daucus* species. Bayesian analyses of individual regions analyzed separately and averaged over multiple trees with *BEAST software (a coalescent approach), however, provided a phylogeny at variance with the concatenated approach, most notably in firmly imbedding *Turgenia latifolia* within *Daucus*.

Keywords—Carrot, taxonomy, Umbelliferae.

The Umbelliferae Juss. comprise some 300–455 genera and 3,000–3,750 species (Constance 1971; Pimenov and Leonov 1993). The family is cosmopolitan and most diverse in the northern hemisphere. Carrot is by far its economically most important member, but the family also contains the vegetables, flavorings, or garnishes angelica, anise, caraway, celeriac, celery, chervil, coriander (cilantro), cumin, dill, fennel, lovage, parsley, and parsnip. Most members of Umbelliferae are easily identifiable to family by the distinctive characters such as herbs with hollow or pith-filled stems, pinnately divided leaves with sheathing bases, small unspecialized flowers in compound umbels, and specialized fruits (Heywood 1971, 1986a).

The ease of identifying members of the Umbelliferae, however, is strongly contrasted with vague circumscription of genera within the family. Traditional classifications within the Umbelliferae are presented by Drude (1897–1898), Heywood (1971, 1993), and Pimenov and Leonov (1993). Recent molecular investigations using DNA sequences from nuclear ribosomal ITS, plastid *rpoC1* intron and *rpl16* intron sequences, plastid *matK* coding sequences and plastid DNA restriction site data (Downie and Katz-Downie 1996; Plunkett et al. 1996; Lee and Downie 1999, 2000; Downie et al. 2000; 2001; Kuo-Fang et al. 2005; Spalik and Downie 2007) do not support many genera within the Umbelliferae as monophyletic. Indeed, fully 11 of 16 genera with more than one species examined per genus in these molecular studies have not been supported as monophyletic (*Aciphylla, Arracacia, Conioselinum, Daucus, Ferula, Heracleum, Ligusticum, Pastinaca, Peucedanum, Pleurospermum,* and *Seseli*). Similarly, studies investigating *Daucus* have clearly documented the paraphyly of the genus as currently circumscribed as it contains the following species in other genera (followed by their distribution): *Agrocharis*, three species (northeastern, tropical, southern Africa) *Athamanta dellacellae* (Libya), *Cryptotainia elegans* (Canary Islands), *Melanoselenium decipiens* (Madeira), *Monizia edulis* (Madeira), *Pachyctenium mirabile* (Libya), *Pseudorlaya minuscula* (Portugal, Spain), *Pseudorlaya pumila* (Morocco, Israel, France, Spain, Greece, Albania, Bulgaria, Crete, Sicily), *Tornabenea annua* (Cape Verde Islands), *Tornabenea tenuissima* (Cape Verde Islands).

Daucus is most common in the Mediterranean region, with some species in other continents and in the southern hemisphere. The latest comprehensive taxonomic monograph of *Daucus* was by Sáenz Laín (1981), but this is largely an intuitive classification lacking a modern phylogeny, specimen citations, distribution maps, detailed descriptions, and use of living specimens for analysis; and it was produced from an examination of a limited number of herbarium specimens. Sáenz Laín (1981) recognized 20 species divided into five sections: *Anisactis, Chrysodaucus, Daucus, Meoides,* and *Platyspermum*. The sole cultivated species, *Daucus carota* L., exists in both cultivated and wild forms. More than 60 species have been proposed for variants within the "*D. carota* complex" alone, for which there are no or only poorly developed barriers to interbreeding among either the wild forms or their domesticates (Small 1978). Sáenz Laín (1981) recognized four subspecies within *D. carota*, Heywood (1986b) 11 subspecies, and Pujadas Salvà (2002) ten subspecies for the Iberian Peninsula alone. The subspecies are diverse phenotypically (Small 1978). Currently, germplasm curators and taxonomists are forced to rely on local floras for identifying *Daucus* germplasm and herbarium specimens.

Current molecular phylogenetic studies of *Daucus* are conducted with few gene regions. As additional gene regions are being discovered for phylogenetic estimation, we are beginning to appreciate that different genes often provide different reconstructions of phylogeny, and that different unlinked loci can have conflicting genealogical histories as a result of the stochastic process of coalescence of alleles due to random genetic drift and incomplete lineage sorting (Wendel and Doyle 1998). Different methods have been used to sort out this problem; one to analyze data from many genes together (a concatenated approach), or to reconstruct a phylogeny from individual gene trees analyzed separately and generating a single tree by averaging over multiple trees (a coalescent approach).

The present study builds upon prior molecular studies of *Daucus* by the use of eight nuclear orthologs and a single plastid region. Establishing orthology in divergent species is difficult, and in plants, identifying true orthologs is complicated by paleopolyploidy and extensive gene duplication (Ku et al. 2000; Fulton et al. 2002; Blanc and Wolfe 2004; O'Brien et al. 2005). Fulton et al. (2002) screened a large tomato expressed sequence tag (EST) database against the *Arabidopsis* genomic sequence and reported the identification of 1,025 genes (referred to as a conserved ortholog set, or COS markers) that are single or low copy in both genomes and have potential utility in examination of phylogeny in dicotyledons. Wu et al. (2006) improved on the discovery of more COS with better determination of orthology across wider phylogenetic distances and identified 2,869 single-copy orthologs. COS genes are shared by most, if not all, euasterid II plant species and *Arabidopsis thaliana* (Cruciferae). The euasterid II clade represents the largest clade of flowering plants, encompassing > 75,000 or one-quarter of the estimated 300,000 flowering plant species and includes many important and diverse crop species, including tomato, potato, and carrot. Alignments of the orthologous sequences across multiple species enabled the design of universal PCR primers which can be used to amplify the corresponding orthologs from a broad range of taxa, including those currently lacking any sequence data. We use the simple term nuclear orthologs because all low copy nuclear orthologous genes are similar, requiring about the same level of care concerning technical issues (e.g. PCR recombination) and are subjected to the same set of lineage-specific and hence variable evolutionary properties (variation in rates, degree of gene conversion, gene amplification or loss).

The purpose of our study is to explore the use of COS DNA sequences (and a single plastid sequence for comparison) in a representative subset of *Daucus* species, potential generic ingroups, and close outgroups.

Materials and Methods

Species—We studied 22 accessions of *Daucus* (see Appendix 1 for species authors) and seven accessions of related ingroup and close outgroup genera (*Astrodaucus littoralis*, *Margotia gummifera*, *Orlaya daucoides*, *Pseudorlaya pumila*, *Torilis leptophylla*, *Turgenia latifolia*), depending on availability of living collections from genebanks, and based upon topologies from prior studies (e.g. Spalik and Downie 2007). Unlike prior studies using plastid or nuclear ITS regions, the single-copy nuclear regions studied here required access to fresh material from living collections. We chose *Margottia gummifera* because of its overall similarity to *Daucus* as noted on our recent collections in Tunisia. Vouchers are deposited in PTIS.

Selection of COS Sequences and Sequence Generation—We used eight COS and one plastid primer (plastid (*psbA/trnH*)). Five COS primers (X1, P28n, D28n, P26n, 171Hn) were screened from 57 COS sequences previously screened from 450 putative COS markers orthologous in the euasterid II clade (Wu et al. 2006), and providing single amplicons in a carrot genetic mapping population (Santos and Simon 2002). We then blasted these 57 sequences against carrot ESTs (Iorizzo et al. 2011) to design refined primers for carrot. Additionally, we filtered *Arabidopsis* ESTs (http://www.arabidopsis.org/) and carrot ESTs (Iorizzo et al. 2011) for unique sequences using blastn within each group at $e \leq 1E^{-10}$. We then conducted reciprocal best match searches between these filtered *Arabidopsis* and carrot datasets, finding 284 best matches. Then we performed multiple sequence alignments of these 284 single matches against genomic database of *Arabidopsis* as a reference to find in-silico the size of introns with over 60% on length in a range of 500–1,200 base pairs (bp) on the expressed sequences of *Daucus*. Ten of these matches were selected to design refined additional primers specific for the carrot group. Subsequently, three primers (CA2, CA6, and CA7) were selected based on single PCR products of 500–1,200 bp in 1.5% agarose gels, and good quality for sequencing. Primer sequences and PCR temperatures are found in Table 1. The remaining PCR conditions for COS follow Rodríguez et al. (2009) and for *psbA/trnH* follow Tate and Simpson (2003).

Data Analyses—Sequences were edited using the Staden package 4.10 (Staden 1996) and aligned in CLUSTALX 2.0.6 (Thompson et al. 1997), with further manual alignments by MacClade 4.08 OS X (Maddison and Maddison 2005). Individual sequences were deposited in GenBank (Appendix 1) and the alignments are available in TreeBASE (Study Accession S13876). Maximum parsimony (MP) and maximum likelihood (ML) analyses used a concatenated data approach and *BEAST analyzed each plastid and COS locus separately (a coalescent approach). Molecular substitution models were evaluated with Modeltest (Posada and Crandall 1998) to select the preferred model among those that could be used in a concatenated dataset for ML or each locus separately for *BEAST.

The MP analyses were conducted in PAUP* version 4.0b8 (Swofford 2002). Gaps were treated as missing data. All characters were treated as unordered and weighted equally (Fitch 1971). The most parsimonious trees were found by heuristic searches under Fitch criteria (Farris 1970) and equal weight for all characters by generating 100,000 replicates and one tree held for each replicate, a random order entry and tree-bisection reconnection (TBR) as the branch swapping method, retaining all most parsimonious trees. Support values for individual clades were estimated with bootstrap analyses (Felsenstein 1985) using 1,000 replicates and the same search criteria as above.

Maximum likelihood (ML) analysis was conducted using RAxML (version 7.3.1), released by Stamatakis in May 2012 (Miller et al. 2010). The analysis was performed using one single model with joint branch length optimization, followed by 1,000 non-parametric bootstrap inferences. The likelihood settings from best-fit model (GTR + I + G) were selected by AIC in Modeltest version 3.7.

For *BEAST analyses, the dataset was imported into BEAUTi (*BEAST 1.6.1 package) to generate the XML format file for *BEAST (Heled and Drummond 2010). Empirical base frequencies were used and the Yule speciation process was selected as a prior on the species phylogeny. All MCMC chains were run for 100 million generations with subsampling every 10,000 generations and three independent runs, and once again for one billion runs with subsampling every 50,000 generations. The log files of the three runs of 100 million runs or the single run of one billion runs were then imported into Tracer 1.5 to get a combined tracer file and to check convergence to the stationary distribution and the effective sample size (ESS) of each parameter. The sample files from the three independent runs were combined after discarding their first 25% as burn-in. They were summarized with a greedy consensus in TreeAnnotator 1.6.1 in the

Table 1. The eight COS oligonucleotide primers and single plastid marker used in this study.

Original eight COS and one plastid (*trnH-psbA*) marker	Redesigned COS markers (see text)	Forward Primer (5′ – 3′)	Reverse Primer (5′ – 3′)	T°a
U237757	X1	ATCGGCTGCTGATGTTTATGATCG	ACAACATCCCTCCATAGAGTTTCAAG	55
C2At5g35970	P28n	ACCATCCTGCTGGATCCCAATGTTGCA	TTCCATATGGCATTCTTGTATCAAGC	60
C2At4g02680	D28n	GCAGCCATGCGAAGTTTGCAGTTGGCT	TCTGATCTCTTTTCATAAGCAGATGC	60
C2At5g17560	P26n	CTGTCAGTGTTAAAGATGCTTATGGCG	TCAAAAGTTGAAGCGATCACATCAATGCT	60
C2At3g63200	171Hn	AAGTGTGCCGTGCCACGTCAGC	TCAACACTCCTCCCACTCCGCC	65
AT1g65270	CA2	GCCAGTATCCTCGACAGTAT	CTCTTTAAGACTGTGTGAGAGGT	55
AT3g19630	CA6	CCAAGTGATGACAACTTCTCCAT	GCTGATGAGCATGTTGTTCTTCAT	55
AT4g14210	CA7	GGTCACAAGCCCATATTGTT	CGTTAATTCCTAGCTCTCCAA	55
trnH-psbA	–	CGCGCATGGTGGATTCACAATCC	GTTATGCATGAACGTAATGCTC	55

*BEAST package. The resulting estimated trees were viewed in FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/) in the *BEAST package.

Results

***Model Selection, Phylogenetic Trees*—**The best fit model of evolution for the concatenated and individual datasets is listed in Table 2. The topology of the MP (Fig. 1; Suppl. Figure 1 presents the nine individual MP gene trees) and ML (Suppl. Figure 2) trees are similar. They both resolved the same three main clades A, A′, and B, all three with 100% bootstrap support, placed *Orlaya daucoides* as sister to these clades (100% bootstrap support), and placed *Astrodaucus littoralis*, *Caucalis platycarpos*, *Torilis leptophylla*, and *Turgenia latifolia* as sister to all other examined taxa. Clade A′ contains all taxa possessing $n = 18$ chromosomes, while the remaining taxa in clade A and in clade B possess $n = 16$, 20, 22, and 44 chromosomes. All branches within clade B were resolved at 99–100% bootstrap support with the topology identical between MP and ML analyses. Within clade A′ the following infraspecific taxa of *D. carota* grouped in both trees: (*D. carota* subsp. *carota* and subsp. *fontanesii*), (*D. carota* subsp. *commutatus* and subsp. *gummifera*), (*D. carota* subsp. *halophilus*, subsp. *maritimus*, and subsp. *maximus*). However, the topology of the remaining accessions within clade A′ varies between MP and ML. Clade A′ contains *D. capillifolius* and *D. sahariensis* in addition to all the infraspecific taxa of *D. carota*. The taxa in clade A outside of clade A′ (*D. crinitus* and *D. muricatus*, as well as two non-*Daucus* species *Margottia gummifera* and *Pseudorlaya pumila*) form a grade. The topology of the remaining outgroup non-*Daucus* species at the base of the tree resolve *Astrodaucus littoralis* as sister, but the remaining three species have minor topological differences.

The topology of the *BEAST tree (Fig. 2), however, has significant topological differences. Most notable are: 1) Clades A, A′, and B of the MP and ML analyses are no longer coherent, and *D. sahariensis* is outside of most species from MP and ML analyses, 2) *Turgenia latifolia* resolves as a well-supported (100% likelihood score) ingroup to *Daucus*, 3) *Orlaya daucoides*, rather than *Astrodaucus littoralis*, is sister to all other taxa. To test whether the plastid sequences or missing data from *D. sahariensis* or *Turgenia latifolia* were responsible for these discordances we ran *BEAST three additional times: 1) without the plastid data (Suppl. Figure 3), 2) without *D. sahariensis* (Suppl. Figure 4), 3) without *D. sahariensis* plus *Turgenia latifolia* (Suppl. Figure 5). In all three of these additional analyses, the *BEAST trees were not notably different.

Discussion

***Maximum Parsimony and Maximum Likelihood Results*—**A significant result from our analysis was the placement of two species from other genera within *Daucus* (*Margottia gummifera*, *Pseudorlaya pumila*), and, as discussed below, a third species in *BEAST analysis, *Turgenia latifolia*. As discussed in the introduction, the placement of non-*Daucus* species within a *Daucus* clade has already been established by the use of plastid or ribosomal DNA characters. New to our result was *Margottia gummifera* within this clade. In addition, the basic topological groupings we found here concur with prior studies. For example, Spalik and Downie (2007) highlighted two main groups of *Daucus* that they termed "Daucus I" and "Daucus II," which also group together in our studies.

Two factors are clear from these generic ingroup results: 1) Populations of these species all occur in the Mediterranean region or Macronesia (Introduction), to which we add *Margottia gummifera* distributed in southwestern Europe and northern Africa. 2) While all of these generic ingroups occur in the Umbelliferae subfamily Apoideae Drude as outlined by Pimenov and Leonov (1993), they are concentrated in only two of the 12 tribes of this subfamily: tribe Caucalideae (*Agrocharis* Hochst., *Astrodaucus* Drude, *Daucus* L. *Pseudorlaya* (Murb.) Murb., *Turgenia* Hoffm.) and tribe Laserpitieae Benth (*Margottia* Boiss., *Melanoselenium* Hoffm., *Monizia* Lowe, *Tornabenea* Webb). The other species occur in the tribe Apieae (*Athamanta* L., *Cryptotaenia* DC.), and in an undetermined group of species ("incertae sedis," *Pachyctenium* Pampan.). Future studies in *Daucus* should concentrate on examining additional species from tribes Caucalidae and Laserpitieae, with full realization that not all species from these genera may group with the *Daucus* clade as was shown dramatically by Spalik and Downie (2007) with the genus *Cryptotaenia*.

Another significant result was the close relationship of the infraspecific taxa of *Daucus carota* and the two other species in clade A′ (*D. capillifolius*, *D. sahariensis*), all with $2n = 18$, in contrast to the other species with different chromosome numbers that are generally divergent as assessed by high bootstrap values. This molecular similarity agreed with other results that failed to distinguish taxa within *D. carota* using isozymes (St. Pierre et al. 1990; St. Pierre and Bayer 1991), AFLPs and ISSRs (Bradeen et al. 2002). These results were clearly reflected in the difficulty to settle on a clearly defined infraspecific classification of *D. carota* discussed in the introduction and they suggest a relatively recent divergence of populations of *D. carota*, emphasizing the need to reconsider many of the intraspecific taxa of this species.

Table 2. Aligned length, model of sequence evolution, and coverage of the one plastid and eight conserved orthologous nuclear sequences used in the present phylogenetic analysis of 29 accessions of *Daucus* and related genera.

Markers	Aligned length	Model of evolution	Sequence coverage
X1	897	HKY + G	Complete
P28n	519	HKY + I	No *Daucus sahariensis*, *Pseudorlaya pumila*
D28n	764	GTR + G	Complete
P26n	1,244	HKY + G	No *Daucus pusillus*, *D. sahariensis*
171Hn	976	GTR + G	Complete
CA2	1,026	HKY + G	No *Daucus guttatus*, *D. littoralis*, *D. sahariensis*, *Turgenia latifolia*
CA6	613	HKY + I	No *Daucus capillifolius*, *D. car*. var. *atrorubens*
CA7	1,134	GTR + G	No *Astrodaucus littoralis*, *Daucus crinitus*, *D. glochidiatus*, *D. sahariensis*, *Turgenia latifolia*
psbA/trnH	339	HKY + G	complete
Concatenated length	7,212	GTR + G	246 of the 261 (94%) species/sequences present

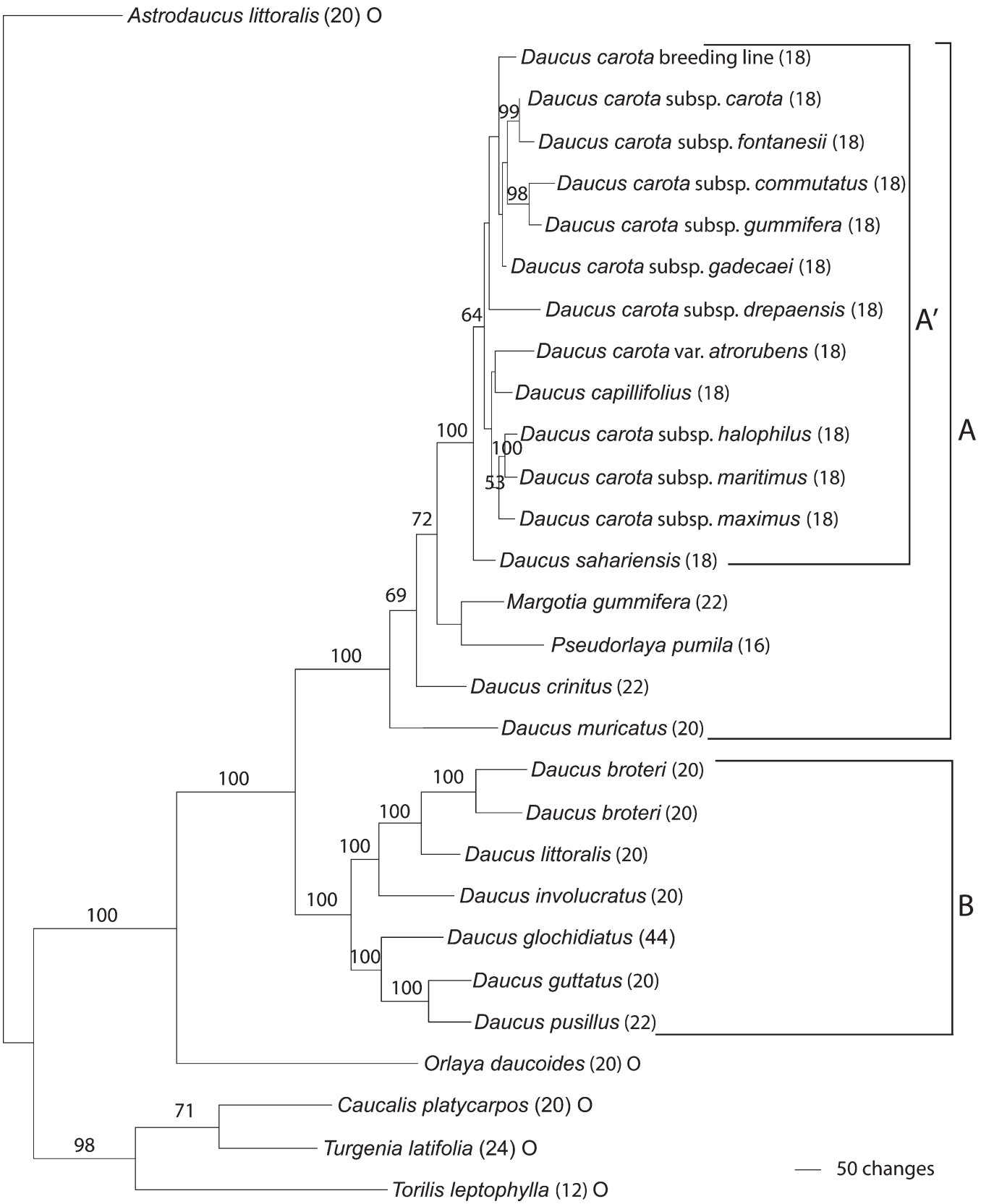

FIG. 1. Strict consensus tree of *Daucus* and closely related genera from maximum parsimony analyses. Numbers at nodes are parsimony bootstrap values, with clades A, A′, and B as discussed in the text. Tree length = 5,347, consistency index = 0.748, retention index = 0.709; rescaled consistency index = 0.5300.

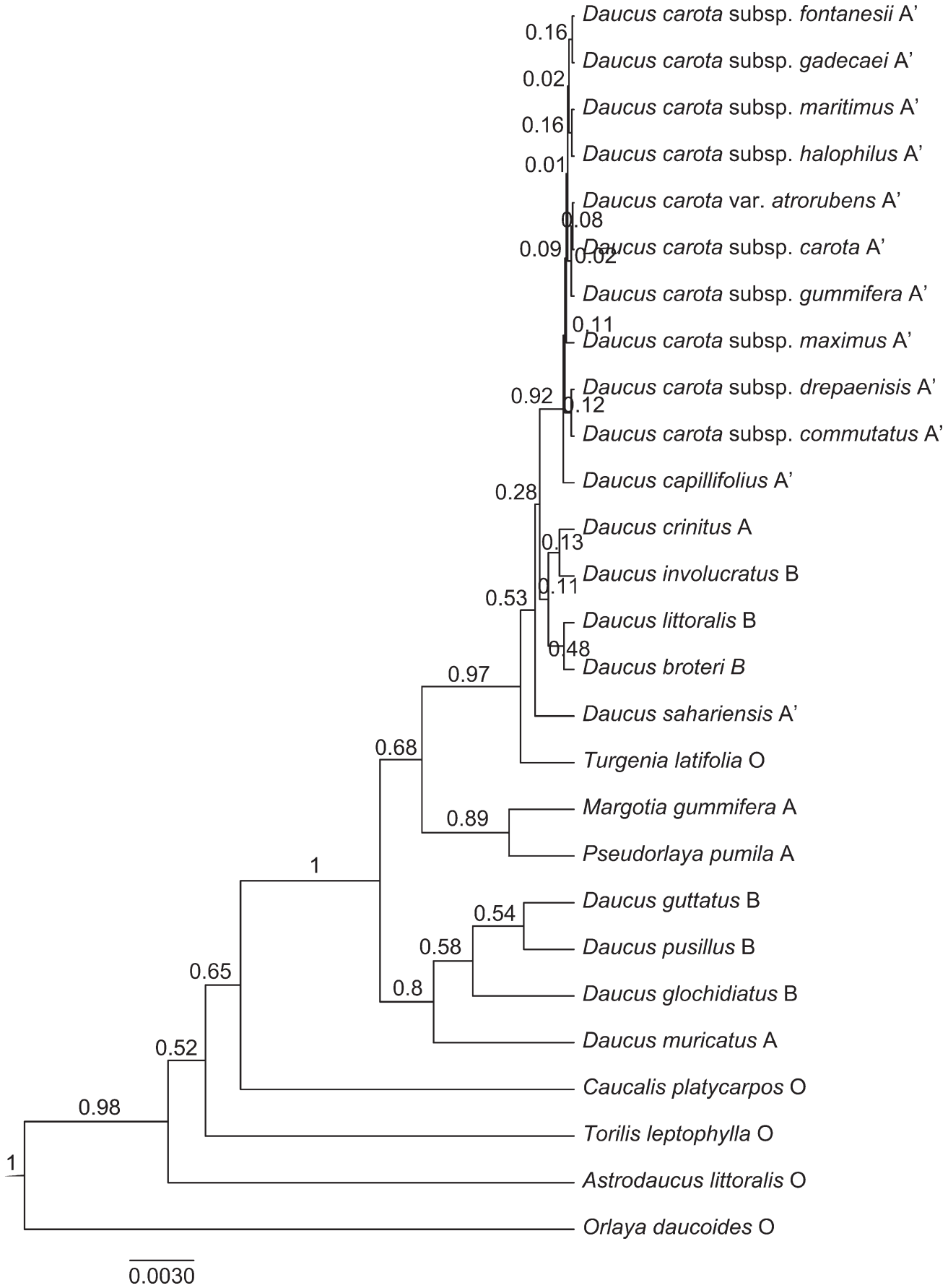

FIG. 2. Bayesian estimate of *Daucus* and closely related genera using one plastid and eight nuclear conserved orthologous sequences with the coalescent-based program *BEAST. Clade posterior probabilities above 0.50 are indicated.

Discrepancy of Concatenated and Coalescent Analyses—Some results from *BEAST were discordant from MP and ML, most notably in the well-supported placement of *Turgenia latifolia* in the ingroup. *BEAST analyzes individual loci in a Bayesian framework to test evolutionary hypotheses without conditioning on a single locus while it analyzes these individual loci to average over tree space. *BEAST analysis assumes that individual loci experience incomplete lineage sorting. Incomplete lineage sorting occurs when an ancestral species undergoes several speciation events in a short period of time, as occurs with similar and recently evolved taxa (Galtier and Daubin 2008) possibly present in morphologically similar taxa. If, for a given gene, the ancestral polymorphism is not fully resolved into two monophyletic lineages when the second speciation occurs, then with some probability the gene tree will be different from the species tree. At present, there is no way to know which species phylogeny is correct, or if ancient hybridization is responsible for this discordance. We are exploring this with additional markers and accessions.

Significance of This Research and Future Plans—With an annual value of $627 million, carrot is among the top ten vegetable crops in the U. S. A. and worldwide (USDA, National Agricultural Statistics Service, http://www.nass.usda.gov/) and the single largest primary source of vitamin A precursors and phytonutrients in the U. S. A. diet. Carrots have a significant nutritional impact on U. S. A. consumers, and can play an important role in ameliorating cancer, heart disease, and obesity.

Diverse germplasm accessions have been identified with effective resistance to carrot pests and with improved consumer quality traits. Carrot is by far the most intensively bred umbelliferous crop. Carrot breeders need better assessments of the diversity of carrot genetic resources, currently numbering approximately 1,500 collections in the U. S. A. carrot genebank at the North Central Regional Plant Introduction Station in Ames, Iowa. These assessments serve as one way for breeders to efficiently tap into this diversity and breed for improved cultivars useful for U. S. A. growers, and to meet growing challenges affecting market qualities demanded by consumers. In carrots, nematodes (*Meloidogyne incognita* and *M. javanica*) and *Alternaria* leaf blight and other foliar diseases are major pests in virtually all production areas of the world, but especially non-desert regions. These are particularly critical challenges to growers of fresh market and processing carrots while flavor, appearance, and nutrition are key market qualities important to consumers.

Carrot is particularly useful for human nutrition. Orange carotenoids of carrot, α- and β-carotene, are the vitamin A precursors and make carrot the largest single source of provitamin A in the U. S. A. diet, accounting for about half of our intake (Simon et al. 2009). Heirloom and foreign carrot cultivar germplasm colors including purple, yellow, white, and red have been bred for U. S. A. production. These novel colors are not only striking in appearance, but the purple anthocyanins, yellow lutein, and red lycopene are important natural "phytonutrient" compounds that may serve an important role in promoting health, potentially reducing the risk of atherosclerosis, cancer, inflammation, and improving antioxidant status (Hung et al. 2004). Nutritious and flavorful vegetables and fruits are important foods to counteract the national trend toward obesity (Epstein et al. 2001).

A major goal of this research, as in all taxonomic research focused on crop plants, was to use phylogenetic data, in concert with crossing and phenotypic data, as a guide to improve cultivars. Our (MP and ML) data clearly demonstrated all of the subspecies of *D. carota*, *D. capillifolius*, and *D. sahariensis* to be closely related. These observations, in concert with chromosome number data, suggested that they are primary initial targets for crossing programs. *Daucus capillifolius* is fully intercrossable with *D. carota* (McCollum 1975), as well as *D. sahariensis* (unpublished data by P. Simon). Similarly, the close relationship of generic ingroups shown here and in prior studies suggests that they have potential for gene transfer to *Daucus*, but with increased difficulty because of different chromosome numbers. The present study shows the utility of conserved nuclear orthologs to present well-resolved phylogenies in the *Daucus* clade. Our continuing studies with additional nuclear markers, ramped up considerably with additional accessions using next generation "targeted" sequencing approaches, greatly expands the scope and reduces costs. In concert with field and herbarium collections our long-term goal is to produce a taxonomic monograph of *Daucus*.

Acknowledgments. We thank Cécile Ané, Brian Just, and Patricia Lu-Irving for advice on *BEAST analyses; Holly Ruess for technical assistance; and David Brenner and Kathleen Reitsma at the North Central Regional Plant Introduction Station in Ames, Iowa for providing germplasm. This work was supported by USDA National Research Initiative grant number 2008-35300-18669 to DS, MB, and LM; and by NSF DEB (grant number 0949121) to DS, MB, and LM.

Literature Cited

Blanc, G. and K. H. Wolfe. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell* 16: 1667–1678.

Bradeen, J. M., I. C. Bach, M. Briard, V. le Clerc, D. Grzebelus, D. Senalik, and P. W. Simon. 2002. Molecular diversity analysis of cultivated carrot (*Daucus carota* L.) and wild *Daucus* populations reveals a genetically nonstructured composition. *Journal of the American Horticultural Society* 127: 383–391.

Constance, L. 1971. *History of the classification of Umbelliferae (Apiaceae)*. Pp. 1–8 in *The biology and chemistry of the Umbelliferae*, ed. V. H. Heywood. New York: Academic Press.

Downie, S. R. and D. S. Katz-Downie. 1996. A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 83: 234–251.

Downie, S. R., D. S. Katz-Downie, and M. F. Watson. 2000. A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl16* and *rpoc1* intron sequences: towards a suprageneric classification of subfamily Apioideae. *American Journal of Botany* 87: 273–292.

Downie, S. R., G. M. Plunkett, M. F. Watson, K. Spalik, D. S. Katz-Downie, C. M. Valiejo-Roman, E. I. Terentieva, A. V. Troitsky, B.-Y. Lee, J. Lahham, and A. El-Oqlah. 2001. Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. *Edinburgh Journal of Botany* 58: 301–330.

Drude, C. G. O. 1897–1898. Umbelliferae. Pp. 63–96 [Apr 1997]; 97–144 [Dec 1897]; 144–192 [Mar 1898]; 193–272 [Jul 1898] in *Die natürlichen Pflanzenfamilien*, eds. A. Engler and K. Prantl. Leipzig, Germany: Wilheim Engelmann.

Epstein, L. H., C. C. Gordy, H. A. Raynor, M. Beddome, C. K. Kilanowski, and R. Paluch. 2001. Increasing fruit and vegetable intake and decreasing fat and sugar intake in families at risk for childhood obesity. *Obesity (Silver Spring, Md.)* 9: 171–178.

Farris, J. S. 1970. Methods for computing Wagner trees. *Systematic Zoology* 19: 83–92.

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

Fitch, W. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology* 20: 406–416.

Fulton, T. M., R. Van der Hoeven, N. T. Eannetta, and S. D. Tanksley. 2002. Identification, analysis, and utilization of conserved ortholog

set markers for comparative genomics in higher plants. *The Plant Cell* 14: 1457–1467.
Galtier, N. and V. Daubin. 2008. Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363: 4023–4029.
Heled, J. and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27: 570–580.
Heywood, V. H. 1971. Systematic survey of Old World Umbelliferae. in *The biology and chemistry of the Umbelliferae* ed. V. Heywood. *Supplement 1 to the Botanical Journal of the Linnean Society* 64: 31–41.
Heywood, V. H. 1986a. The Umbelliferae - an impossible family? *Symbolae Botanicae Upsalienses* 26: 73–80.
Heywood, V. H. 1986b. Flora Europaea notulae systematicae ad floram Europeam spectantes. *Feddes Repertorium* 79: 1–68.
Heywood, V. H. 1993. *Flowering plants of the world*. New York: Oxford University Press.
Hung, H. C., M. C. Huang, J. M. Lee, D. C. Wu, H. K. Hsu, and M. T. Wu. 2004. Association between diet and esophageal cancer in Taiwan. *Journal of Gastroenterology and Hepatology* 19: 632–637.
Iorizzo, M., D. A. Senalik, D. Grzebelus, M. Bowman, P. F. Cavagnaro, M. Matvienko, H. Ashrafi, A. Van Deynze, and P. W. Simon. 2011. De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genomics* 12: 389.
Ku, H. M., T. Vision, J. Liu, and S. D. Tanksley. 2000. Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proceedings of the National Academy of Sciences USA* 97: 9121–9126.
Kuo-Fang, C., C.-I. Peng, S. R. Downie, K. Spalik, and B. A. Schaal. 2005. Molecular systematics of the trans-Pacific alpine genus *Oreomyrrhis* (Apiaceae): phylogenetic affinities and biogeographic implications. *American Journal of Botany* 92: 2054–2071.
Lee, B. Y. and S. R. Downie. 1999. A molecular phylogeny of Apiaceae tribe Caucalideae and related taxa: inferences based on ITS sequence data. *Systematic Botany* 24: 461–479.
Lee, B. Y. and S. R. Downie. 2000. Phylogenetic analysis of cpDNA restriction sites and *rpsl6* intron sequences reveals relationships among Apiaceae tribes Caucalideae, Scandiceae and related taxa. *Plant Systematics and Evolution* 221: 35–60.
Maddison, D. R. and W. P. Maddison. 2005. MacClade 4.08: Analysis of phylogeny and character evolution. Sunderland: Sinauer.
McCollum, G. D. 1975. Interspecific hybrid *Daucus carota* × *D. capillifolius*. *Botanical Gazette (Chicago, Ill.)* 136: 201–206.
Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 1–8 in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans: San Diego Supercomputer Center, La Jolla, California, U. S. A.
O'Brien, K. P., M. Remm, and E. L. Sonnhammer. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Research* 33: D476–D480.
Pimenov, M. G. and M. V. Leonov. 1993. *The genera of Umbelliferae: a nomenclator*. Kew: Royal Botanic Gardens.
Plunkett, G. M., D. E. Soltis, and P. S. Soltis. 1996. Evolutionary patterns in Apiaceae: inferences based on *matK* sequence data. *Systematic Botany* 21: 477–495.
Posada, D. and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
Pujadas Salvà, A. J. 2002. El complejo de *Daucus carota* L (Apiaceae) en la Flora Ibérica. *Anales Jardín Botánica de Madrid* 59: 368–375.
Rodríguez, F., F. Wu, C. Ané, S. Tanksley, and D. M. Spooner. 2009. Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? *BMC Evolutionary Biology* 9: 191.
Sáenz Laín, C. 1981. Research on *Daucus* L. (Umbelliferae). *Anales del Jardin Botanico de Madrid* 37: 497–498.
Santos, C. A. F. and P. W. Simon. 2002. QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots. *Molecular Genetics and Genomics* 268: 122–129.
Simon, P., L. Pollak, B. Clevidence, J. Holden, and D. Haytowitz. 2009. Plant breeding for human nutrition. *Plant Breeding Reviews* 31: 325–392.
Small, E. 1978. A numerical taxonomic analysis of the *Daucus carota* complex. *Canadian Journal of Botany* 56: 248–276.
Spalik, K. and S. R. Downie. 2007. Intercontinental disjunctions in *Cryptotaenia* (Apiaceae, Oenantheae): an appraisal using molecular data. *Journal of Biogeography* 34: 2039–2054.
Staden, R. 1996. The Staden sequence analysis package. *Molecular Biotechnology* 5: 233–241.
St. Pierre, M. D., R. J. Bayer, and I. M. Weiss. 1990. An isozyme-based assessment of the genetic variability within the *Daucus carota* complex (Apiaceae: Caucalideae). *Canadian Journal of Botany* 68: 2449–2457.
St. Pierre, M. D. and R. J. Bayer. 1991. The impact of domestication on the genetic variability in the orange carrot, cultivated *Daucus carota* ssp. *sativus* and the genetic homogeneity of various cultivars. *Theoretical and Applied Genetics* 82: 249–253.
Swofford, D. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4.0b3a PPC. Sunderland: Sinauer Associates.
Tate, J. A. and B. B. Simpson. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* 28: 723–737.
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.
Wendel, J. F. and J. J. Doyle. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. Pp. 265–296. in *Molecular systematics of plants II: DNA sequencing*, eds. D. E. Soltis, P. S. Soltis, and J. J. Doyle. Boston: Chapman and Hall.
Wu, F., L. A. Mueller, D. Crouzillat, V. Pétiard, and S. D. Tanksley. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: A test case in the euasterid plant clade. *Genetics* 174: 1407–1420.

Appendix 1. List of DNA vouchers used in this study. "PI" and "Ames" prefixes refer to U. S. germplasm collections with data and germplasm availability listed in the U. S. Germplasm Resources Information Network (GRIN); http://www.ars-grin.gov/npgs/searchgrin.html. Vouchers of all collections are listed in italics and are deposited at PTIS. Accession information is listed as follows: species name in alphabetical order, geographic origin, voucher information, and GenBank accession numbers for the DNA sequence data in the following order: nuclear 171Hn, CA2, CA6, CA7, D28n, P26n, P28n, X1; and plastid (trnH-psbA). The code NS indicates no sequence for that species/COS combination (Table 2), or in the case of Astrodaucus littoralis/171Hn the partial sequence was under 200 bp and was not accepted by GenBank, but is available in the aligned TreeBASE file.

Astrodaucus littoralis (M. Bieb.) Drude. AZERBAIJAN. *PI 277064*, NS, KC526246, KC526271, NS, KC526322, KC526351, KC526378, NS, KC526433. ***Caucalis platycarpos*** L. GERMANY. *PI 649446*, KC526222, KC526251, KC526275, KC526301, KC526327, KC526356, KC526383, KC526409, KC526438. ***Daucus broteri*** Ten. accession 1. IRAN. *PI 652233*, KC526219, KC526248, KC526273, KC526299, KC526324, KC526353, KC526380, KC526406, KC526435. ***Daucus broteri*** Ten. accession 2. TURKEY. *PI 652387*, KC526226, KC526255, KC526278, KC526305, KC526331, KC526360, KC526387, KC526413, KC526442. ***Daucus capillifolius*** Gilli. LIBYA. *PI 279764*, KC526225, KC526254, NS, KC526304, KC526330, KC526359, KC526386, KC526412, KC526441. ***Daucus carota*** var. ***atrorubens*** Alef. EGYPT. *PI 279777*, KC526220, KC526249, NS, KC526300, KC526325, KC526354, KC526381, KC526407, KC526436. ***Daucus carota*** L. subsp. ***carota***. SPAIN. *PI 279759*, KC526223, KC526252, KC526276, KC526302, KC526328, KC526357, KC526384, KC526410, KC526439. ***Daucus carota*** L. breeding line, *Simon 2566B*, KC526218, KC526247, KC526272, KC526298, KC526323, KC526352, KC526379, KC526405, KC526434. ***Daucus carota*** subsp. ***commutatus*** Thell. ITALY. *Ames 7674*, KC526224, KC526253, KC526277, KC526303, KC526329, KC526358, KC526385, KC526411, KC526440. ***Daucus carota*** subsp. ***drepanensis*** (Arcang.) Heywood. DENMARK. *PI 279762*, KC526227, KC526256, KC526279, KC526306, KC526332, KC526361, KC526388, KC526414, KC526443. ***Daucus carota*** subsp. ***fontanesii*** Thell. PORTUGAL. *Ames 26383*, KC526233, KC526261, KC526285, KC526311, KC526338, KC526367, KC526394, KC526420, KC526449. ***Daucus carota*** subsp. ***gadecaei*** (Rouy and E. G. Camus) Heywood. FRANCE. *Ames 31193*, KC526228, KC526257, KC526280, KC526307, KC526333, KC526362, KC526389, KC526415, KC526444. ***Daucus carota*** subsp. ***gummifer*** (Syme) Hook. f. HUNGARY. *PI 279775*, KC526231, KC526259, KC526283, KC526309, KC526336, KC526365, KC526392, KC526418, KC526447. ***Daucus carota*** subsp. ***halophilus*** (Brot.) A. Pujadas. cult from botanical garden in France, source unknown, *Ames 31194*, KC526232, KC526260, KC526284, KC526310, KC526337, KC526366, KC526393, KC526419, KC526448. ***Daucus carota*** subsp. ***maritimus*** (Lam.) Batt. PORTUGAL. *Ames 26393*, KC526236, KC526263, KC526288, KC526314, KC526341, KC526370, KC526397, KC526423, KC526452. ***Daucus carota*** subsp. ***maximus*** (Desf.) Ball. SPAIN. *PI 295862*, KC526239, KC526266, KC526291, KC526317, KC526344, KC526373, KC526400, KC526426, KC526455. ***Daucus crinitus*** Desf.

PORTUGAL. *Ames 26417*, KC526221, KC526250, KC526274, NS, KC526326, KC526355, KC526382, KC526408, KC526437. ***Daucus glochidiatus*** (Labill.) Fisch., C. A. Mey. and Avé-Lall. AUSTRALIA *HRI (Warwich UK genebank) 8251*, KC526229, KC526258, KC526281, NS, KC526334, KC526363, KC526390, KC526416, KC526445. ***Daucus guttatus*** Sibth. and Sm. GREECE. *Ames 25601*, KC526230, NS, KC526282, KC526308, KC526335, KC526364, KC526391, KC526417, KC526446. ***Daucus involucratus*** Sibth. and Sm. TURKEY. *Ames 25813*, KC526234, KC526262, KC526286, KC526312, KC526339, KC526368, KC526395, KC526421, KC526450. ***Daucus littoralis*** Sibth. and Sm. ISRAEL. *PI 295857*, KC526235, NS, KC526287, KC526313, KC526340, KC526369, KC526396, KC526422, KC526451. ***Daucus muricatus*** L. PORTUGAL *Ames 25419*, KC526238, KC526265, KC526290, KC526316, KC526343, KC526372, KC526399, KC526425, KC526454. ***Daucus pusillus*** Michx. ARGENTINA. *ECMC [E. Camadro and M. Cauhépé]* Pus3, KC526242, KC526269, KC526294, KC526320, KC526347, NS, KC526402, KC526429, KC526458. ***Daucus sahariensis*** Murb. TUNISIA. *Ames 30283*, KC526243, NS, KC526295, NS, KC526348, NS, NS, KC526430, KC526459. ***Margotia gummifera*** (Desf.) Lange. TUNISIA. *Ames 30292*, KC526237, KC526264, KC526289, KC526315, KC526342, KC526371, KC526398, KC526424, KC526453. ***Orlaya daucoides*** (L.) Greuter. TURKEY. *PI 649477*, KC526240, KC526267, KC526292, KC526318, KC526345, KC526374, KC526401, KC526427, KC526456. ***Pseudorlaya pumila*** (L.) Grande. TUNISIA. *PI 662301*, KC526241, KC526268, KC526293, KC526319, KC526346, KC526375, NS, KC526428, KC526457. ***Torilis leptophylla*** (L.) Rchb. f. SYRIA. *Ames 25750*, KC526245, KC526270, KC526297, KC526321, KC526350, KC526377, KC526404, KC526432, KC526461. ***Turgenia latifolia*** (L.) Hoffm. SYRIA. *PI 649433*, KC526244, NS, KC526296, NS, KC526349, KC526376, KC526403, KC526431, KC526460.