American Journal of Botany
Celebrating 100 years 1914–2014

American Journal of Botany 101(10): 1666–1685, 2014.

SPECIAL INVITED PAPER

Phylogenomics of the carrot genus (*Daucus*, Apiaceae)[1]

Carlos Arbizu, Holly Ruess, Douglas Senalik, Philipp W. Simon, and David M. Spooner[2]

U. S. Department of Agriculture, Agricultural Research Service, Vegetable Crops Research Unit; and Department of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison, Wisconsin 53706-1590 USA

- *Premise of the study:* We explored the utility of multiple nuclear orthologs for the taxonomic resolution of wild and cultivated carrot, *Daucus* species.
- *Methods:* We studied the phylogeny of 92 accessions of 13 species and two subspecies of *Daucus* and 15 accessions of related genera (107 accessions total) with DNA sequences of 94 nuclear orthologs. Reiterative analyses examined data of both alleles using ambiguity codes or a single allele with the highest coverage, trimmed vs. untrimmed homopolymers; pure exonic vs. pure intronic data; the use of all 94 markers vs. a reduced subset of markers; and analysis of a concatenated data set vs. a coalescent (species tree) approach.
- *Key results:* Our maximum parsimony and maximum likelihood trees were highly resolved, with 100% bootstrap support for most of the external and many of the internal clades. They resolved multiple accessions of many different species as monophyletic with strong support, but failed to support other species. The single allele analysis gave slightly better topological resolution; trimming homopolymers failed to increase taxonomic resolution; the exonic data had a smaller proportion of parsimony-informative characters. Similar results demonstrating the same dominant topology can be obtained with many fewer markers. A Bayesian concordance analysis provided an overall similar phylogeny, but the coalescent analysis provided drastic changes in topology to all the above.
- *Conclusions:* Our research highlights some difficult species groups in *Daucus* and misidentifications in germplasm collections. It highlights a useful subset of markers and approaches for future studies of dominant topologies in *Daucus*.

Key words: Apiaceae; carrot; *Daucus*; germplasm; next-generation sequencing; phylogenomics.

Until very recently, it was difficult to imagine the availability of genomic information with unprecedented amounts of data for a wide array of organisms (Delsuc et al., 2005; Rokas and Carroll, 2006; McCormack et al., 2013). However, recent developments of next-generation sequencing technologies now make it possible to sequence millions of bases in a single experiment at a relatively low cost (Egan et al., 2012; Soltis et al., 2013), ushering in “phylogenomics” (Delsuc et al., 2005; McCormack et al., 2013), that we use here to refer to the use of genome-scale genetic data for phylogenetic analyses.

Phylogenomics is a new field with as yet many unexplored applications and potential constraints. For example, the reconciliation of well-supported species trees is a primary interest in systematics (Blair and Murphy, 2011), but since phylogenetics is moving away from single-locus to multilocus analyses (Edwards, 2009), debates on gene tree discordance are becoming more common. For many years, the alternative to deal with discordance in multilocus data was concatenation (Rokas et al., 2005; Dunn et al., 2008; Schierwater et al., 2009), with an idea that “incorrect” or “noisy” phylogenetic signal was overcome by huge data sets obtained from concatenation of many loci, leading to strongly supported phylogenetic species trees (Chen and Li, 2001; Rokas et al., 2003; Christelová et al., 2011; Blair et al., 2012; Lang et al., 2013; Salichos and Rokas, 2013). However, even though combining data from multiple genes can result in strongly supported phylogenetic resolution, assuming a single divergent history may undermine interpretation of the phylogeny on a combined gene tree (Kolaczkowski and Thornton, 2004; Lewis et al., 2005; Mossel and Vigoda, 2005).

Biological explanations were proposed for gene tree discordances, such as coalescent stochasticity (Takahata, 1989), the movement of genes among species by hybridization and introgression (Rieseberg et al., 2000), horizontal gene transfer (Doolittle, 1999), gene duplication (Page and Charleston, 1997), and incomplete lineage sorting (Pamilo and Nei, 1988). Baum (2007) proposed a “primary concordance tree” as a valuable summary of the dominant phylogenetic history among a group of organisms. He defined the dominant phylogenetic history as the tree composed of clades with a higher concordance factor than any contradictory clade. We use the term “dominant topology”, as determined by our concatenated data set. Also, this tree should provide a useful estimate of the primary history and the degree of reticulation/divergence at various points in that history. Baum (2007) also indicated that clades on concordance trees can be annotated with their concordance factor (CF), the proportion of the genome for which the clade is true. The CF can be estimated from population histories or from multilocus molecular data sets.

[1] Manuscript received 11 March 2014; revision accepted 17 June 2014.

The authors thank David Baum who advised C.A. in earlier stages of this research during his course “Phylogenetic Analysis of Molecular Data”, Cécile Ané for advice and assistance in running “BUCKy”, and Kathleen Reitsma and staff at the North Central Regional Plant Introduction Station in Ames Iowa for advice and assistance in obtaining the germplasm examined here.

[2] Author for correspondence (e-mail: david.spooner@ars.usda.gov)

doi:10.3732/ajb.1400106

The Apiaceae (Umbelliferae) family contains 455 genera and over 3500 species, and is one of the largest families of seed plants (Pimenov and Leonov, 1993). The genus *Daucus* contains carrot (*Daucus carota* L. subsp. *sativus* Hoffm.), which is the most notable cultivated member of Apiaceae in terms of economic importance and nutrition. Cultivated carrot is grown on an estimated 1.2 million ha annually worldwide (carrots and turnips as aggregated data) (FAO, 2012), with an annual crop value of about $640 M in the United States for fresh and processing carrots (USDA National Agricultural Statistics Service, 2012). It is the single, largest primary source of vitamin A precursors and phytonutrients and is particularly beneficial for human nutrition. The orange carotenoids of carrot, α- and β-carotene, are vitamin A precursors and make carrot the largest single source of provitamin A in the U. S. diet, accounting for about half of dietary intake (Simon et al., 2009). The economic importance of carrot stimulates research into breeding to feed a constantly growing population, to guarantee food security, and to adapt to climate change. Wild *Daucus* species may play an important role in this process, providing genes that can be used for breeding purposes such as pest and disease tolerance or resistance, yield increase, male sterility, nutraceutical, and culinary traits, among others. A better understanding of the species boundaries and phylogenetic relationships of *Daucus* will play a crucial role in future breeding programs.

The taxonomic distinction and phylogenetic relationships among species of genus *Daucus* are not clear, even though there have been studies of its morphology, anatomy and biochemistry (Vivek and Simon, 1999), and phylogeny. Many generic boundaries within the Apiaceae are unnatural as documented by molecular investigations based on DNA sequences from nuclear ribosomal internal transcribed spacers, plastid *rpoC1* intron and *rpl16* intron sequences, plastid *matK*-coding sequences, plastid DNA restriction-site data, and DNA sequences from nuclear orthologs (Plunkett et al., 1996; Downie et al., 2000; Lee and Downie, 2000; Spalik and Downie, 2007; Spooner et al., 2013). Molecular data from these studies place some species from the genera *Agrocharis*, *Athamanta*, *Cryptotaenia*, *Margotia*, *Melanoselinum*, *Monizia*, *Pachyctenium*, *Pseudorlaya*, and *Tornabenea* within a monophyletic *Daucus* clade.

The latest genus-level treatment available using a morpho-anatomical classification is reported by Sáenz Laín (1981) who recognized 20 species divided into five sections: *Daucus* L. (12 species), *Anisactis* DC. (three species), *Platyspermum* DC. (three species), *Chrysodaucus* Thell. (one species), and *Meoides* Lange (one species). Rubatzky et al. (1999) later estimated 25 species of *Daucus*. The genus *Daucus* has a center of endemism in the Mediterranean, with several species occurring in North America, South America, and Australia (Sáenz Lain, 1981). Spalik et al. (2010) provided a biogeographic analysis of *Daucus* with dates for radiations from the Mediterranean region. *Daucus carota* L. subsp. *carota* is the best-known wild species within carrots (Brandenburg, 1981). The cultivated carrot, *D. carota* subsp. *sativus*, was first domesticated from wild populations of *D. carota* subsp. *carota* from Central Asia (Iorizzo et al., 2013).

The taxonomy of *D. carota* L. is particularly problematical. It undergoes widespread hybridization experimentally and spontaneously with commercial varieties and other named subspecies (Krickl, 1961; Saenz de Rivas and Heywood, 1974; McCollum, 1975, 1977; Umiel et al., 1975; Wijnheijmer et al., 1989; St. Pierre et al., 1990; Ellis et al., 1993; Steinborn et al., 1995; Vivek and Simon, 1999; Nothnagel et al., 2000; Hauser and Bjørn, 2001; Hauser, 2002). Coauthor Simon has obtained fertile intercrosses of cultivated carrot and *D. sahariensis* (unpublished data). The haploid chromosome number for *Daucus* ranges from $n = 9$ to $n = 11$. Diploid numbers range from $2n =$ 18, 20, and 22, but two tetraploid species have been reported (Grzebelus et al., 2011). The four species with $2n = 18$ (*D. carota* all subspecies, *D. capillifolius*, *D. sahariensis*, *D. syrticus*) are clearly interrelated based on shared karyotypes (Iovene et al., 2008). Results from our recent morphological studies (Spooner et al., 2014) caused us to question the many wild subspecies and suggest that there may be only two wild subspecies of carrot, *D. carota* subsp. *carota* and subsp. *gummifer*.

The present study comprises 97 accessions for which 94 nuclear orthologous genes were sequenced here, and we later added sequences for 10 accessions with a subset of these 94 nuclear orthologs as described in the methods. The genes are distributed along all nine chromosomes of cultivated carrot (*D. carota* subsp. *sativus*). Orthologs are genes derived from a single ancestral gene in the last common ancestor of the target species (Koonin, 2005). Phylogenetic studies rely on the identification of true orthologs in diverse angiosperms. Nuclear ortholog markers have great potential utility in further studies on comparative genomics and phylogenetics (Fulton et al., 2002; Li et al., 2008; Levin et al., 2009; Rodríguez et al., 2009; Cai et al., 2012). The goals of our study were: (1) to compare the results from maximum parsimony, maximum likelihood, and Bayesian concordance analyses, (2) to examine the effect of concatenated data vs. a coalescent (species tree) analyses, and (3) to evaluate the potential of multiple nuclear orthologs using next-generation technologies to resolve the phylogenetic relationships of *Daucus*.

MATERIALS AND METHODS

Plant species—We examined 92 accessions of 13 *Daucus* species and two subspecies and 15 accessions of 9 species of non-*Daucus* genera (107 accessions in total) collected from around the world (Table 1). We sampled the species diversity as widely as possible, based on the availability of germplasm accessions. This availability left 12 *Daucus* species unsampled: *D. arcanus* García-Martín and Silvestre (Spain), *D. biseriatus* Murb. (Algeria), *D. conchitae* Greuter (Greece), *D. durieua* Lange (Mediterranean), *D. gracilis* Steinh. (Algeria), *D. hochstetteri* A. Braun ex Drude (Eritrea, Ethiopia), *D. jordanicus* Post (Libya, Israel, Jordan), *D. microscias* Bornm. and Gauba (Iran, Iraq), *D. montanus* Humb. and Bonpl. ex Schult. (Central and South America), *D. reboudii* Coss. (Algeria, Tunisia), *D. setifolius* Desf. (Algeria, Morocco, Tunisia, Portugal, Spain), and *D. virgatus* (Poir.) Maire (Algeria, Tunisia). When germplasm was available, we examined more than one accession of the same species. All accessions were obtained from the United States National Plant Germplasm System, with *Daucus* maintained at the North Central Regional Plant Introduction Station in Ames, Iowa. Full details of the collections are available at the Germplasm Resources Information Network (http://www.ars-grin.gov/npgs/acc/acc_queries.html). Vouchers are maintained at the Potato Introduction Station Herbarium (PTIS). Many of the genera mentioned above in the *Daucus* clade were not available as germplasm, which precluded us from obtaining sufficient quantity and quality of DNA for our study. All examined materials are wild taxa except one cultivated accession, *D. carota* subsp. *sativus* (Table 1).

Data set—Figure 1 visually summarizes all procedures described below. A data set was created from the aligned DNA sequences generated by a Roche (Basel, Switzerland) 454 GS FLX+ Platform. Initially, we examined 102 conserved nuclear ortholog markers from 97 accessions. These nuclear orthologs were identified by following a protocol developed by Wu et al. (2006). Expressed sequence tags (ESTs) of *Arabidopsis thaliana* (hereafter, *Arabidopsis*), carrot, sunflower, and lettuce were obtained from different public sources. *Arabidopsis* sequences were obtained from a copy of the TAIR10 assembly at PlantGDB. A set of 41 671 *Arabidopsis* sequences was downloaded from the following website: http://www.plantgdb.org/download/Download/xGDB/AtGDB/ATtranscriptTAIR10. Carrot ESTs were obtained from Additional File 2

of Iorizzo et al. (2011). Only assembled contigs were used; unassembled Sanger reads were excluded, resulting in a set of 58751 sequences.

Sunflower and lettuce sequences were obtained from The Compositae Genome Project website at http://compgenomics.ucdavis.edu/. A set of 31605 *Helianthus annuus* ESTs was downloaded from http://cgpdb.ucdavis.edu/asteraceae_assembly/data_assembly_files/GB_ESTs_Feb_2007.sp.Heli_annu.clean.assembly. In addition, a set of 26720 lettuce ESTs was downloaded from http://cgpdb.ucdavis.edu/asteraceae_assembly/data_assembly_files/GB_ESTs_Feb_2007.sp.Lact_sati.clean.assembly. These sequence sets were each aligned with each other using the program blastn version 2.2.25 (Camacho et al., 2009) with a maximum expected value of 1e-10 and low complexity filtering by DUST. Two sets of three species were aligned in all pairwise combinations to detect reciprocal best matches (RBM). The comparison between *Arabidopsis*, carrot, and sunflower resulted in 4023 RBM, and the *Arabidopsis*, carrot and lettuce comparison resulted in 5180 RBM. Sequence sets were also aligned to themselves, and sequences were designated as single-copy genes when there were no blast alignments to other sequences within the same set. The two RBM sets were then further reduced to contain only sequences which were found to be single-copy genes in all three of the species making up the set. The set containing sunflower yielded 71 sequences, and the set containing lettuce yielded 92 sequences; the two sets combined yielded 128 unique sequences. The carrot sequences passing these steps were used for primer design. For each identified gene, *Arabidopsis* EST, sunflower and/or lettuce EST, carrot EST, and carrot whole genome sequence (WGS) (Iorizzo et al., 2014) were aligned using the program MacClade version 4.08a (Maddison and Maddison, 2005). It was possible to determine the exonic and intronic regions of each gene, and with the use of the WGS of carrot, estimation of intron sizes were obtained. We designed most of the nuclear orthologs to capture sequences of 500–700 bp and more than 60% intron content. Primers were designed selecting regions that were identical in sequence between all species and with a maximum match of the 3′ end of the primer between all sequences (usually at least 5 bp), a melting temperature of around 55°C, and GC content between 40–60%. Primers were checked for melting temperature, hairpins, and self-dimers using the program OligoAnalyzer version 3.1 (Owczarzy et al., 2008).

Using these criteria, we designed 102 marker primer pairs with an expected amplicon size of 427–777 bp based on a draft carrot genomic sequence, realizing that some species in this study could fall outside of this range. Sixty-nine markers contain more than 60% intron content, 23 have 26–60% intron content; 10 were designed to consist entirely of exons. Primers were evaluated for functionality and expected fragment size using the inbred line B493 of *Daucus carota* subsp. *carota*. We performed a clean PCR (minimizing the unused reagents at the end) of genomic DNA of all our accessions with these primers, then evaluated the success of amplification and actual size in a 1.5% agarose gel using standard methods.

Quantification of all amplifications was performed using a Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Paisley, UK). For each of the 97 accessions, equal amounts of product from 102 nuclear ortholog marker amplifications were pooled. Each of these 97 pools was individually purified with magnetic beads to remove the PCR reaction components. For each of the 97 pools, we ligated one of the 12 Roche MIDs (multiplex identifiers) to the pool to barcode the single accession. This reaction was cleaned with magnetic beads again. We then quantified the PCR fragments that were successfully ligated to MIDs using RL (Roche Library, a fluorescent tag that is a Roche proprietary product) attached to the MID using the same machine for Picogreen quantification. Pools for sequencing consisted of 4–6 accession pools, themselves pooled following the Rapid Library Preparation Method Manual (Roche, 2010). Final pools were sent to the University of Wisconsin-Biotechnology Center where libraries were prepared using the em-PCr Method Manual-Lib-A SV (Roche, 2009), and sequenced on a Roche GS FLX+ instrument. We chose the Roche 454 sequencing platform because it provided longer read lengths than available with other technologies (Shendure and Ji, 2008; Egan et al., 2012).

Raw sequence data were parsed by barcode to separate reads from each accession, and vector sequence and barcodes were removed. Reads for each accession were assembled with the program MIRA version 3.4.0 (Chevreux et al., 1999). Average read coverage was determined for each contig/accession combination, i.e., the average number of sequence reads covering each nucleotide of the assembled sequence. Those contigs with average read coverage below 20 were removed. Assembled contigs were matched with the appropriate nuclear ortholog marker using the program MUMmer version 3.0 (Kurtz et al., 2004). For each nuclear ortholog marker, DNA sequences from all accessions were aligned using the program MUSCLE version 3.8.31 (Edgar, 2004), and further manual alignment corrections were performed using MacClade.

Sequence analysis—MIRA assembled one or (more commonly) two alleles. These alleles can differ by one or many single nucleotide polymorphisms (SNPs) or indels. Only *D. glochidiatus* is tetraploid, but it exhibited low allelic variation similar to the diploids. In some cases, more than two alleles were found with our coverage cutoff of 20 for useable data. However, in every case, these low coverage "extra" alleles differed in only minor ways (only 1–5 bp) from the two higher coverage alleles and were discarded from further analysis. Two methods were used to process the information provided by the heterozygous allele state. One method was to construct a single consensus sequence using IUPAC degenerate nucleotide ambiguity codes. A second method was to select the one allele per accession that had the highest average read coverage. DNA sequences from these individual genes of the single allele with the highest coverage are deposited in GenBank (Table 2; Appendix S1; see Supplemental Data with the online version of this article), and the aligned database is deposited in the TreeBase repository (http://purl.org/phylo/treebase/phylows/study/TB2:S15477?x-access-code=52b70011707357994e61de7d36a88e63&format=html, submission ID: 15477). We concatenated the 94 genes (see Results) into a single alignment and analyzed these two data sets (single vs. two alleles) for all 107 species, resulting in an aligned length of 112002 bp for the data set of one allele only, and 116652 bp for the data set where two alleles were merged.

Phylogenetic analyses—We chose the Roche 454 platform to obtain long reads, but according to Margulies et al. (2005), this platform produces unreliable sequence for homopolymers over eight base pairs. We encountered difficult and ambiguous alignments with homopolymers of bases A (adenine) and T (thymine) up to 16 bases long. Long homopolymers were also encountered in the carrot genome by Iorizzo et al. (2011). James Speers and Xiao Liu (personal communication, University of Wisconsin-Madison, Biotechnology Center) suggested that homopolymers over six bases long are unreliable. Hence, in our present study, we shortened homopolymers to a maximum of six using MacClade.

We rooted our trees on *Oenanthe*, based on Downie et al. (2000). We first performed maximum parsimony (MP) analyses of 94 markers and 97 accessions, comparing a data set of a single allele with the highest coverage with homopolymers shortened to a maximum of six, to unmodified homopolymers. After we initiated this work, we obtained 10 accessions important for our analysis from fieldwork in Tunisia and Morocco that were not initially available (Table 1). We performed a MP analysis of each marker separately and identified by visual inspection 10 markers that best approached the topology of the concatenated data sets. Based on this analysis, we performed MP analyses adding these 10 additional accessions (107 accessions total) to the concatenated data set but with DNA sequences of these 10 markers obtained with the dideoxy chain termination technique (Sanger et al., 1977). We next performed a MP analysis of 94 markers and 107 accessions comparing a data set of one allele only chosen by highest coverage, with a data set of a two alleles merged into one using ambiguity codes.

Each study group will have different levels of species divergence depending on the ingroup and outgroup variation and may require different proportions of intronic markers (that are more useful for lower divergence) vs. exonic markers (useful for greater divergence). To explore our choice of markers in our study, we performed a MP analysis of 94 markers and 107 accessions of a data set using the single allele of the pure intronic regions vs. pure exonic regions.

All MP analyses were conducted in PAUP* version 4.0a131 (Phylogenetic Analysis Using Parsimony; Swofford, 2002). Question marks and blank spaces were treated as missing data and gaps, respectively. All characters were treated as unordered and weighted equally (Fitch, 1971). The most parsimonious trees were found using a heuristic search (Farris, 1970) by generating 100000 random-addition sequence replicates and one tree held for each replicate. Branch swapping used tree-bisection reconnection (TBR) retaining all most parsimonious trees. Then, we ran a final heuristic search of the most equally parsimonious trees from this analysis using TBR and MULPARS. Bootstrap values (Felsenstein, 1985) for the clades were estimated using 1000 replicates with simple addition sequence, setting MAXTREES to 1000.

Maximum likelihood (ML) phylogenetic analysis initially was attempted after selecting the best-fit evolutionary models for the individual gene sequence data (Table 2) with model selection computed using the Akaike information criterion (AIC), using jModelTest version 2.1.3 (Darriba et al., 2012). With these models, we attempted to get a ML tree with the program GARLI version 2.0 (Genetic Algorithm for Rapid Likelihood Inference; Zwickl, 2006). However, this was impossible to run with our large data set (111166 bp) due to time limits, estimated to be several years using our Dell PC with 16 GB memory and a 3.4 GHz Intel Core i7-2600 processor. Alternatively, we obtained a ML tree with the program RAxML version 8.0.0 (Randomized Accelerated Maximum Likelihood; Stamatakis, 2014), using GTR+G model and estimating individual

TABLE 1. Accessions examined in this study.

Taxon and 2*n* chromosome number[a]	Tentative new identifications	Accession[b]	Location or source[c]
Ingroups			
Daucus aureus Desf. (22)		PI 295854	Israel. Wadi Rubin (HaMerkaz).
D. aureus		PI 319403	Israel. Mediterranean Region.
D. aureus +#		PI 478858	France. Dijon.
D. broteri Ten. (20) +	*D. guttatus* 1	PI 652233	Iran. Mazandaran: Dhalus Road, Dasht-e Nazir, Kandalus.
D. broteri +#	*D. guttatus* 2	PI 652329	Greece. Peloponnese: 4 km from Skoura, toward Leonidion, Laconia Prefecture.
D. broteri +#	*D. guttatus* 1	PI 652340	Syria. Kassab.
D. guttatus Sibth. and Sm. (20) +	*D. guttatus* 1	PI 652343	Syria. Halwah.
D. broteri +	*D. guttatus* 3	PI 652367	Turkey. Mugla.
D. capillifolius Gilli (18) +		PI 279764	Libya. Near Jefren.
D. capillifolius		Ames 30198	Tunisia. Medenine.
D. capillifolius #		Ames 30202	Tunisia. Medenine.
D. capillifolius +		Ames 30207	Tunisia. Medenine.
D. carota L. subsp. *carota* (18, all subspecies) #		Ames 25017	Germany. Saxony-Anhalt.
D. carota subsp. *carota* +		Ames 26393	Portugal. Castelo Branco.
D. carota subsp. *carota*		Ames 26394	Portugal. Portalegre near Monforte.
D. carota subsp. *carota*		Ames 26401	Portugal. Portalegre near Monforte.
D. carota subsp. *carota*		Ames 26408	Portugal. Beja.
D. carota subsp. *carota*		Ames 27397	Uzbekistan. Between Yalangoch and Sobir Raximova.
D. carota subsp. *carota*		Ames 30250	Tunisia. Nabuel: along Route 28 at junction of road to Takelsa.
D. carota subsp. *carota*		Ames 30251	Tunisia. Nabuel: Route 26, between Takelsa and El Haouaria, 26 km from El Haouaria.
D. carota subsp. *carota* #		Ames 30252	Tunisia. Nabuel: Sidi Daoud, 1 km from Route 27.
D. carota subsp. *carota*		Ames 30253	Tunisia. Nabuel: between El Haouarcae and Dor Allouche.
D. carota subsp. *carota*		Ames 30254	Tunisia. Nabuel: between El Haouarcae and Dor Allouche.
D. carota subsp. *carota* +		Ames 30255	Tunisia. Nabuel: along road between Korba and Beni Khalled.
D. carota subsp. *carota*		Ames 30259	Tunisia. Bizerte: south side of Ischkeul.
D. carota subsp. *carota*		Ames 30260	Tunisia. Bizerte: along Route 51, west of Ghzab.
D. carota subsp. *carota*		Ames 30261	Tunisia. Bizerte: grounds of Direction Regionale Mogods, Khroumerie Sejnane.
D. carota subsp. *carota*		Ames 30262	Tunisia. Beja: road from Route 7, just west of Sejnane to Cap Negro.
D. carota subsp. *carota* *		Ames 31570	Morocco. Larache: approximately 10 kilometers south of Larache, Laouamra Region.
D. carota subsp. *carota* #		PI 274297	Pakistan. Northern areas.
D. carota subsp. *carota*		PI 279759	Spain. Madrid (Botanic Garden).
D. carota subsp. *carota*		PI 279762	Source: Denmark. Copenhagen.
D. carota subsp. *carota*		PI 279775	Source: Hungary. Pest. Botanical Garden.
D. carota subsp. *sativus* #		PI 279777	Source: Egypt. Giza: Orman Botanic Garden.
D. carota subsp. *carota* #		PI 279788	Austria. Vienna.
D. carota subsp. *carota*		PI 279798	Spain. Madrid.
D. carota subsp. *carota*		PI 295862	Spain.
D. carota subsp. *carota*		PI 390887	Israel. Central Israel: From Bet Elazari.
D. carota subsp. *carota*		PI 421301	USA. Kansas: Elk County.
D. carota subsp. *carota*		PI 430525	Afghanistan. Zardek.
D. carota subsp. *carota*		PI 478369	China. Xinjiang: near Chou En Lai Monument Stone River, Sinkiang.
D. carota subsp. *carota*		PI 478873	Italy. Sardinia: St. Elia Beach, 50 m from sea, Cagliari.
D. carota subsp. *carota*		PI 478881	USA. Oregon: roadside between Echo and Pendleton.
D. carota subsp. *carota*		PI 478884	Source: The Netherlands, South Holland: Botanical Garden, Leiden.
D. carota subsp. *carota*		PI 502244	Portugal. Coimbra: Lousa.
D. carota subsp. *carota*		PI 652225	Source: France. Collection site unknown.
D. carota subsp. *carota* #		PI 652226	Greece. N. Khalkidiki: 10 km N of Kassandra on coast road.
D. carota subsp. *carota*		PI 652229	Source: Tunisia.
D. carota subsp. *carota*		PI 652230	Albania. Lushnje.
D. carota subsp. *carota*		PI 652341	Syria. Ash Sheik Hasan.
D. carota subsp. *carota*		PI 652393	Turkey. Konya: 10-15 km to Seydisehir, between Yarpuz and Konya.
D. carota subsp. *gummifer* (Syme) Hook.f.		Ames 7674	Source: Italy. Tuscany: Botanic Garden.
D. carota subsp. *gummifer*		Ames 26381	Portugal. Faro: Near Portunao.
D. carota subsp. *gummifer* +		Ames 26382	Portugal. Faro: Near Sagres.
D. carota subsp. *gummifer*		Ames 26383	Portugal. Faro: Near Aljezur.
D. carota subsp. *gummifer* #		Ames 26384	Portugal. Beja.
D. carota subsp. *gummifer*		Ames 31193	France.
D. carota subsp. *gummifer*		Ames 31198	Unknown.
D. carota subsp. *gummifer*		PI 478883	France. Finistere: maritime turf, Le Conquet.
D. carota subsp. *gummifer* +	*D. guttatus* 1	PI 652387	Turkey. Antalya.
D. carota subsp. *gummifer* +		PI 652411	France. Finistere: Pointe de Rospico, Navez.
D. carota subsp. *carota* +	*D. guttatus* 1	Ames 25898	Turkey. Konya: Konya, toward Beysehir.
D. carota +	*D. guttatus* 1	PI 286611	Source: Lebanon. Faculty of Agricultural Sciences.
D. crinitus Desf. (22) #		Ames 26413	Portugal. Castelo Branco.
D. crinitus		PI 652412	Portugal. Braganca: near Zava.

TABLE 1. Continued.

Taxon and 2*n* chromosome number[a]	Tentative new identifications	Accession[b]	Location or source[c]
D. crinitus		PI 652413	Portugal. Guarda: near Barca de Alva.
D. crinitus		PI 652414	Portugal. Faro: near Bengado.
D. glochidiatus (Labill.) Fisch., C.A.Mey. & Avé-Lall. (44) +#		PI 285038	Source: CSIRO, Australia. Capital Territory.
D. guttatus (20) +	*D. guttatus* 1	PI 279763	Source: Israel. Jerusalem Department of Botany.
D. guttatus +	*D. guttatus* 2	PI 652331	Greece. Peloponnese: village of Loutra Agias Elenis, 17 km south of Korinthos, Korinthia Prefecture.
D. guttatus +	*D. guttatus* 2	PI 652360	Turkey. Mugla: between Soke and Milas.
D. involucratus Sm. (22) +		PI 652332	Greece. Peloponnese: village of Loutra Agias Elenis, 17 km south of Korinthos, Korinthia Prefecture.
D. involucratus +		PI 652350	Turkey. Izmir.
D. involucratus +#		PI 652355	Turkey. Izmir: 5 km north of Kusadasi.
D. littoralis Sibth. & Sm. (20) +		PI 295857	Israel. Beit Alpha.
D. littoralis +#		PI 341902	Israel.
D. littoralis Sm. +#	*D. guttatus* 3	PI 652375	Turkey. Mugla: between Dalaman-Gocik and Fethiye.
D. muricatus L. (20)		Ames 25419	Portugal. Coimbra: Pitanca de Baixo-Condeixa.
D. muricatus +#		Ames 29090	Tunisia. South of Tunis along Hwy. 3 toward Zaghouan.
D. muricatus		PI 295863	Spain. Cordoba. From Villa del Rio (Cordoba).
D. pusillus Michx. (22) +#		PI 349267	Uruguay. Montevideo. Near La Colorado Beach.
D. pusillus		PI 661242	United States. Oregon: near Hunters River Cove, Curry.
D. pusillus		PI 661256	United States. Texas: Bastrop County, along Route 713 (Farm to Market Road), 5 miles south of Rockne.
D. sahariensis Murb. (18)		Ames 29096	Tunisia. Between Tataouine and Bir Lahmer.
D. sahariensis		Ames 29097	Tunisia. Between Tataouine and Remada.
D. sahariensis #		Ames 29098	Tunisia. Between Remada and Chenini.
D. syrticus Murb. (18)		Ames 29107	Tunisia. Near Beni Kdache to the south.
D. syrticus		Ames 29108	Tunisia. Between Medenine and Matmatas.
D. syrticus		Ames 29109	Tunisia. Between Medenine and Matmatas.
D. syrticus +#		Ames 29110	Tunisia. Between Matmatas and El Hamma, near the Gabes airport.
D. tenuisectus Coss. ex Batt. (22) *		Ames 31616	Morocco. Al Haouz: 25.7 km north of center of Ijoukak, 29 km south of Asni, Nfiss River Valley, Imgdal Region.
D. tenuisectus *+		Ames 31617	Morocco. Al Haouz: Along Route 203, 2.3 km south of road going to Oukaimeden from Tahannout (P2028), approximately 12 km north of bridge over river, Nfiss River Valley, Moulay Brahim Region.
Margotia gummifera Lange (22) +		Ames 30292	Tunisia. Jendouba: road to Tabarka, near Tabarka airport.
Pseudorlaya pumila Grande (16) *+		Ames 29088	Tunisia. South of Medenine toward Tataouine, near Bir Lahmer.
Outgroups			
Ammi visnaga (L.) Lam. (20, 22)		Ames 30185	Tunisia. Bizerte: National Park Ischkeul on road to Eco Museum.
Astrodaucus littoralis Drude (20) +		PI 277064	Source: Azerbaijan. Baku Botanical Garden.
Caucalis platycarpos L. (20) +		PI 649446	Germany. Saxony-Anhalt: Mannsdorf.
Oenanthe virgata Poir. (not reported)		Ames 30293	Tunisia. Beja: Route 11, 41 km from Eudiana, 254 km from Beja.
Orlaya daucoides (L.) Greuter (20) +#		PI 649477	Turkey. Aydin: Dilek Peninsula Reserve.
Orlaya daucorlaya Murb. (14) *		PI 649478	Greece. Epirus: 8 km from Aristi, toward Ioannina.
Torilis arvensis (Hudson) Link (24) *∞		Ames 31623	Morocco. Al Haouz: Along Route 203, 2.3 km south of road going to Oukaimeden from Tahannout (P2028), approximately 12 km north of bridge over river, Nfiss River Valley, Moulay Brahim Region.
T. leptophylla (L.) Rchb.f. (12)		Ames 25750	Syria. Salma.
T. leptophylla *∞		Ames 31619	Morocco. Ifrane: 2 km south of N13 on minor road to Ain-Leuh, beginning a few kilometers southeast of Azrou, Tigrigra Region.
T. nodosa (L.) Gaertn. (24) *		Ames 31606	Morocco. Berkane: Montes des Beni Snassen, Fezouane Region.
T. nodosa *		Ames 31607	Morocco. Al Haouz: Moulay Brahim, between Tahannout and Asni, Moulay Brahim Region.
T. nodosa *∞		Ames 31622	Morocco. Al Haouz: Moulay Brahim, between Tahannout and Asni, Moulay Brahim Region.
Turgenia latifolia (L.) Hoffman (24) +		PI 649433	Syria. Ain el Haour.

[a] These names correspond to those in the Germplasm Resources Information Network (GRIN) database, except for the proposed new identifications of the subspecies of *D. carota* listed in Spooner et al. (2014). The 10 accessions designated with an asterisk were added after Roche 454 analyses with Sanger sequencing, and will therefore have more accessions with sequence data in the ninth column of Table 2. The 34 accessions designated with a plus sign were used in the *BEAST analysis, the 21 accessions designated with a pound sign were used in the BUCKy analysis, and the three accessions designated with an infinite sign were not used in our first *BEAST analysis. The 2*n* chromosome numbers are those known for the species, not the individual accessions, and are taken from Grzebelus et al. (2011) and IPCN chromosome reports (http://www.tropicos.org/Project/IPCN).

[b] Plant Introduction (PI) numbers are permanent numbers assigned to germplasm accessions in the National Plant Germplasm System (NPGS). Germplasm centers in the NPGS assign temporary site-specific numbers to newly acquired germplasm (Ames numbers for carrots and other Apiaceae maintained at the North Central Regional Plant Introduction Station in Ames, Iowa, USA) until an accession's passport data and taxonomy is verified, it is determined not to be a duplicate accession, and it has been determined the accession can be successfully maintained. These accessions may or may not be assigned a PI number after the assessment period.

[c] Location refers to where the germplasm was collected in the wild, while source refers to germplasm acquired through another entity such as a market vendor or genebank.

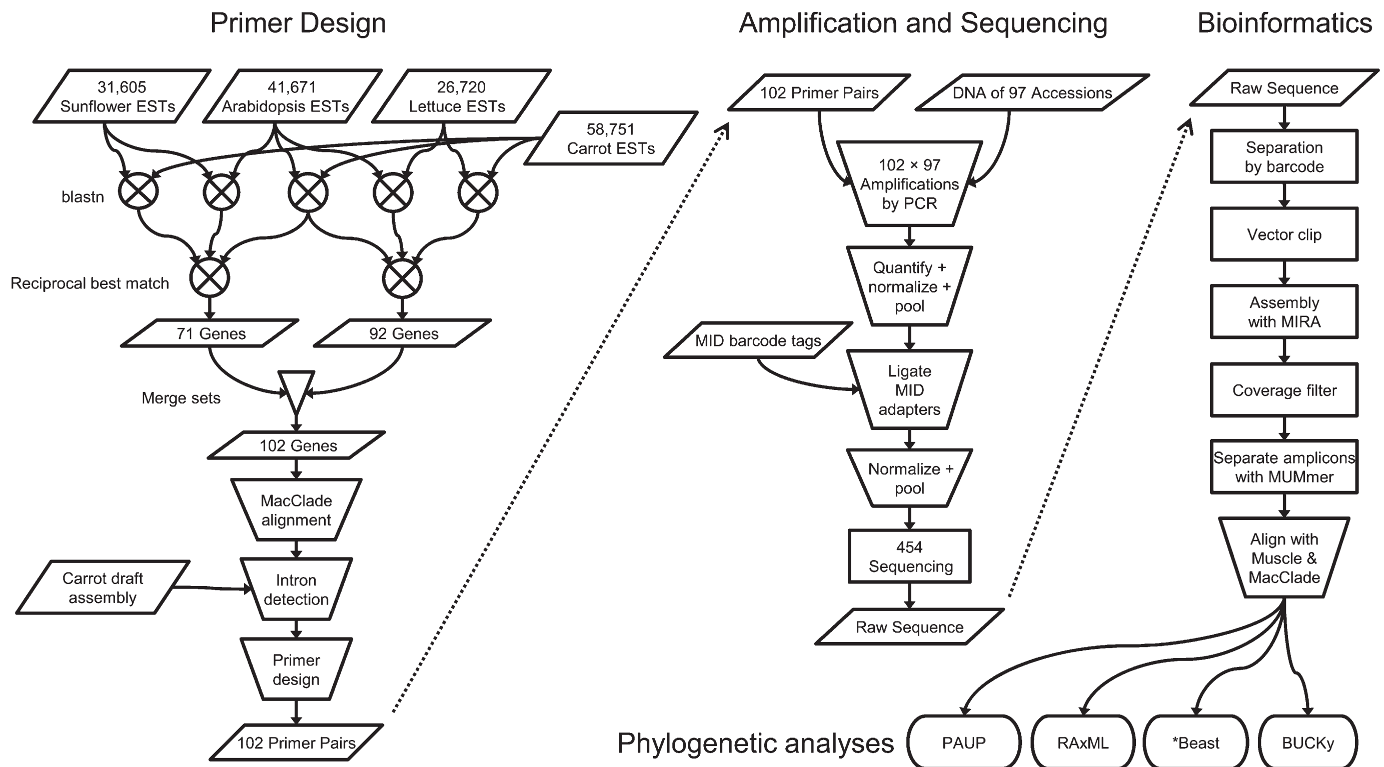

Fig. 1. Flow chart of the laboratory and bioinformatic procedures used in this study.

alpha-shape parameters, GTR rates, and empirical base frequencies for each individual gene. Using the same program, 1000 nonparametric bootstrap inferences were obtained. Both analyses were conducted via the CIPRES (Cyberinfrastructure for Phylogenetic Research; Miller et al., 2010) portal at the San Diego Supercomputer Center (http://www.phylo.org).

We also performed a Bayesian concordance analysis (BCA) (Ané et al., 2007) to obtain the primary concordance tree using the program BUCKy version 1.4.2 (Bayesian Untangling of Concordance Knots; Larget et al., 2010). According to Cécile Ané (personal communication, University of Wisconsin-Madison, Department of Botany), there is a practical limit of 25 accessions for BUCKy. Therefore, we conducted pruned analyses choosing representative accessions (Table 1) from major clades as determined from the maximum parsimony and maximum likelihood analyses to explore gene to gene conflict in our data set. All 94 genes with their corresponding model of nucleotide substitution (Table 2) were analyzed separately in MrBayes version 3.2.2 (Ronquist et al., 2012) using the BEAGLE library (Ayres et al., 2012) with four chains and two searches run simultaneously for 10 million generations sampling every 1000 generations. This analysis was also conducted via the CIPRES. We summarized the MrBayes results for the 94 genes using the program *mbsum* included in BUCKy, removing 1001 trees from each chain as burn-in. We then performed the BCA with four independent runs with four linked chains for all four different levels of discordance: $\alpha = 0.1$, 1, 10, and infinite (a larger value of α corresponds to greater gene tree incongruence); in each run with 1 100 000 generations; 100 000 generations were discarded as the burn-in period. Default settings were used for all other parameters.

We also performed a Bayesian analysis using *BEAST package version 1.8.0 (Bayesian Evolutionary Analysis by Sampling Trees; Drummond et al., 2012) to obtain a species tree estimation using a coalescent approach. An XML format file was generated using BEAUti version 1.8.0. With 94 genes with their corresponding model of evolution (Table 2), one initial analysis used 104 accessions comprising 27 species (Table 1). In addition, one final analysis used only HKY models of evolution (Table 2) and a subset of 37 accessions comprising 22 species (Table 1). All analyses were conducted using the Yule process as a species tree prior (Gernhard, 2008). All Markov chain Monte Carlo (MCMC) chains were run for 1 billion generations sampling every 50 000 generations. We imported the log files of the two runs into the program Tracer version 1.6.0 in *BEAST to analyze the convergence to the stationary distribution and the effective sample size (ESS) of each parameter. The samples of plausible trees from the two runs were individually summarized, and 25% of the trees were discarded as burn-in using the program TreeAnnotator version 1.8.0 in the *BEAST package. The resulting trees were viewed in FigTree version 1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/). The *BEAST analyses were conducted in the same PC used for ML with GARLI, but with the BEAGLE library and NVIDIA GPU GeForce GTX 580.

RESULTS

Sequence data—Eight of the 102 markers had low coverage (less than 20×) as determined from MIRA version 3.4.0 (Chevreux et al., 1999) or had ambiguous alignments and were discarded from further analyses. The remaining 94 markers were distributed on all nine linkage groups of *Daucus carota* (Table 2). Of these 94 marker/97 accession matrices (9118 cells), there were missing data for 558 cells, resulting in 6.1% missing data.

Maximum parsimony (MP) analyses—As explained in the introduction, we conducted reiterative (1) modification of the sequences with and without homopolymers trimmed to a maximum of 6 bp, (2) analyses of single markers one by one, vs. a concatenated data set of 10 or all 94 markers, (3) analyses of a data set of a single allele vs. two alleles merged into one by ambiguity codes, and (4) analyses of intronic vs. exonic regions.

Modifications of topology by trimming the homopolymers to a maximum of six—Our initial data set of 94 markers and 97 accessions with a single allele with highest coverage had an aligned length of 112 002 bp, and the data set with homopolymers trimmed to a maximum of six had an aligned length of

111 166 bp (a reduction in 836 bp). The tree scores (Table 3) and MP topologies of these two analyses (Appendices S2, S3; see online Supplemental Data) are similar, with only minor differences in bootstrap scores, and rearrangements of clade relationships within clade A′, containing the *Daucus* species with $2n = 18$ chromosomes, in contrast to all other species in *Daucus* clades A and B with $2n = 20$, 22, and 44 (*D. glochidiatus*), and 16 (*Pseudorlaya pumila*). Other than clade A′, there is no widespread pattern in chromosome numbers.

Topology of differing numbers of markers—As expected, our data sets with a single allele with highest coverage examined with markers one by one produced MP results with a wide range of topologies. When we compared these individual gene trees to that using all 94 markers (Fig. 2) we noted that some individual gene trees were similar to the concatenated "dominant" topology. Figure 3 shows the MP topology of marker DC10366 that appeared the most similar to the dominant topology, and Fig. 4 shows the MP topology of 10 concatenated markers that, like marker DC10366, approached the dominant topology. These results are useful for those wishing to reconstruct dominant topologies of *Daucus* with additional accessions.

The dominant topology is highly resolved, with 100% bootstrap support for most of the external and many of the internal clades. Notable exceptions are the relationships within the accessions of *D. carota* and *D. capillifolius* in clade A′, and of *D. sahariensis* and of *D. syrticus* also in clade A′, but these two groups are strongly supported as sister clades to each other with 100% bootstrap support. Within the *D. carota* and *D. capillifolius* clade, there are two clusters associated with a geographical component. All accessions with known locality data of *D. capillifolius* and *D. carota* collected in Libya and Tunisia form a weakly supported clade (<70% bootstrap, Fig. 2, highlighted in red). In addition, most accessions of *D. carota* collected in Portugal and Spain form a strongly supported clade (100% bootstrap, Fig. 2, highlighted in blue), but two accessions from Portugal and Spain were not present in this clade (Ames 26401, PI 279798) and one accession from Morocco (Ames 31570) was present in this clade.

The dominant topology grouped different accessions of many different species with strong support, but in addition to the species intermixing in clade A′ as discussed earlier, this topology failed to group *D. broteri* and *D. guttatus* together, placing these two species in three separate well-supported clades (all 100% bootstrap support). We grew these accessions again and resequenced the DNA with the 10 nuclear orthologs mentioned earlier to check for misidentifications. The plants appeared the same as our original vouchers and they grouped the same with these new DNA data. However, the morphological characters distinguishing these species are ambiguous, mirroring our molecular results. Because of uncertainty of the application of these names, we name them here as *D. guttatus* (the earliest name) 1, 2, and 3. *Margotia gummifera* and *Pseudorlaya pumila* were sister to the *D. carota* clade, followed by *D. aureus* and *D. muricatus*, and then the remaining *Daucus* species. *Orlaya* was supported as the closest outgroup to *Daucus*. We labeled the two main clades each with 100% bootstrap support as clade A and clade B.

Maximum parsimony analyses using different scoring of allelic variants—Our MP results comparing a single allele with the highest coverage vs. two alleles merged into one using ambiguity codes differed in the following ways. The tree scores (Table 3) document a longer aligned database for two alleles, 115 882 bp, vs. 111 166 bp for single alleles (4716 bp or 4.2% longer). The consistency index of the resulting two-allele tree is larger (0.641) than of the single-allele tree (0.53). The topology of the two trees (Fig. 2, online Appendix S4) also differed. For example, the two geographic subsets (1) Libya and Tunisia, (2) Portugal and Spain are missing in the two-allele tree. There is a polytomy in clade B of the two-allele tree that is resolved in the single-allele tree, although with only 67% bootstrap support. However, many of the remaining topologies remain the same.

Maximum parsimony analyses examining the pure intronic regions from the pure exonic regions—We designed our analysis of *Daucus* and close outgroups to use a majority of markers with 60% intron content or more, assuming that such regions were needed to give phylogenetic resolution. To broaden the analyses, we designed primers to evaluate 10 purely exonic gene regions to have data potentially useful for the farther outgroups and to explore the phylogenetic utility of these regions for the ingroup. Our pure exonic regions (gleaned from all 94 markers) had 20 478 aligned characters, vs. 90 688 aligned characters for the pure intronic regions. The consistency indices for both trees (online Appendices S5, S6) are nearly identical, but there are many more parsimony-informative characters in the intronic regions as a proportion of the total characters. Specifically, the total database had 18.4% exonic regions and 81.6% intronic regions, and taken as a proportion of these length differences, the introns had 20.6% parsimony-informative characters vs. 13.6% for the exons, about 50.7% larger for the introns. The topologies of the two trees (Appendices S5, S6) reflect these differences in the number and parsimony-informativeness of these two data sets. While the main clades A, A′, and B are the same, there is a reduction in bootstrap support for some of the main clades.

Maximum likelihood analysis—Our initial attempt to obtain a ML tree with 1000 bootstrap replicates using mixed models on the GARLI platform was unsuccessful due to lack of time (we estimated that over 5 yr would be needed based on the run times of our attempt). Hence, we ran the ML analysis with 1000 bootstrap replicates using RAxML with a single model of evolution, but using different alpha-shape parameters, GTR rates, and empirical base frequencies. This ML tree (Fig. 5) has the same overall topology as the MP tree (Fig. 2), including the geographic subsets in clade A, and recovers the same clades A, A′, and B. In addition, there are good bootstrap support values in most components of this tree. Two notable exceptions are (1) *D. aureus* and *D. muricatus* are not on the same clade in ML as they are in MP, but form sister clades. (2) Although *D. guttatus* 1, 2, and 3 have the same sister-group relationships in both analyses, the relationships of these sister-group pairs differ.

Bayesian concordance analysis—Our pruned analysis showed an acceptable result as the standard deviation of concordance factors was less than 0.005. The primary concordance tree (Fig. 6) estimated for 94 genes and 21 accessions with Bayesian analyses showed a similar topology to the MP and ML trees. In addition, there were no significant differences among the concordance factors using the four different prior probabilities on gene tree incongruence (α values). BCA worked well for this pruned analysis. The concordance factors (CF) of these same main clades, despite having comparable taxonomic relationships, are much lower than bootstrap support values in the MP and ML analyses, but they are meant to show different aspects of the topology and are not meant to be

TABLE 2. Characteristics of the 94 nuclear orthologs used in our study.

Markers[a]	Linkage group	Aligned length[b]	Model of evolution[c]	Forward primer (5′–3′)	Reverse primer (5′–3′)	Intron content	GenBank nos.[d]	No. of accessions with sequence data	Exon[e] (nt)	Intron (nt)	No. introns	No. exons	Length of gene (nt)	% of intron
DC10366*	V	1128	TVM+G (GTR+G) [HKY+G]	GTAGTTCTTCACACAGCTTCCTTC	ATCAACTTCGTGCTGCTCTTG	<60% intron	KJ521466–KJ521568	103	2022	2464	6	7	4486	54.9
DC10966*	I	1344	TPM2uf+G (GTR+G) [HKY+G]	TCATTGCACCACCGACATG	CTTCAAGGTCTTTGTGGTGTTCA	>60% intron	KJ520950–KJ521056	107	654	2927	4	5	3581	81.7
DC11423	V	606	HKY+I+G	TCGTTCATGAGGGAAATCGC	GGACCCTCCTGAGTAAACAC	exon only	KJ520089–KJ520179	91	1269	0	0	1	1269	0.0
DC11491	IX	605	HKY+G	TCCTGGTCTCCTGTTGCA	CACAGGTTGAGGAAGTGAAGATTC	exon only	KJ516559–KJ516652	94	1965	1819	3	4	3784	48.1
DC11601	IX	1729	HKY+G	AGATCTCTTCTCGCTCTCC	CAACTGCCTATACTCTTCGCTC	>60% intron	KJ518726–KJ518816	91	1794	18209	16	17	20003	91.0
DC12018	I	1632	GTR+I+G [HKY+I +G]	CTCGAGAGCACAACCACATG	CAACAATGTTCAACCCTCGTTCT	<60% intron	KJ521755–KJ521845	91	855	3474	6	7	4329	80.2
DC13263	I	644	GTR+I+G [HKY+G]	CAAGCTGCAAGACCTTCTCTC	ACCTTCATCAAACACCTTGCAG	exon only	KJ515770–KJ515865	96	1725	0	0	1	1725	0.0
DC1340	VIII	613	TPM3uf+I+G [(HKY+G)]	AGAAGAATCTGAAATGGCAGGTAGA	GGCCTTGTCATTCAAAACAGC	<60% intron	KJ514413–KJ514504	92	378	306	3	4	684	44.7
DC14804	VII	2765	TPM2uf+G (GTR+G) [HKY+G]	GGGAAGTCCTGTACTCATTTTGG	GGAAAGTTTGGGCTAGGTGC	>60% intron	KJ520600–KJ520663	64	1800	5460	17	18	7260	75.2
DC1494	IX	837	TM3uf+G [(HKY+G)]	ACCGTTTCTACAATCTGGAAATGC	TCATTATCCTCGTCATCTGGTT	>60% intron	KJ520761–KJ520852	92	533	2761	4	5	3294	83.8
DC15318	V	1296	TPM3uf+I+G [(HKY+I+G)]	GTGCCTATCGTGGTATTACAGC	GTTTTGGTGATGCCTGTGC	>60% intron	KJ514782–KJ514877	96	1482	6519	15	16	8001	81.5
DC15347*	I	1014	TrN+I+G [(HKY+I+G)]	ATCTCAACCTTTTCTCTGATGGTG	ACCACCCACCTACAGATCC	>60% intron	KJ522039–KJ522142	104	885	2291	6	7	3176	72.1
DC15398	V	668	TrN+G [(HKY+G)]	AGCAGAAGTTGATGATGATGCTG	ATGAGTTGGGCTTCGCATC	<60% intron	KJ520180–KJ520260	81	681	5241	4	5	5922	88.5
DC15442	VI	1267	TVM+I+G (GTR+I+G) [HKY+I+G]	CTGTGGCTTTGACTCTCTTCATC	CCCACAATTCAGACCAAGAAGC	>60% intron	KJ520853–KJ520949	97	1290	1990	5	6	3280	60.7
DC15678	III	645	TPM2uf+I+G (GTR+I+G) [HKY+I+G]	ATTGCAGGAAGTGGGACG	CGAACATGACCATCATCTCCAC	>60% intron	KJ515536–KJ515587	52	455	2542	3	4	2997	84.8
DC15851	III	665	GTR+I+G [HKY+I+G]	GTACCAGAATCAGCTTTCTCCAC	CACTCTCGACTACAAGGTAACATG	exon only	KJ514236–KJ514315	80	321	0	0	1	321	0.0
DC16119	VII	901	HKY+G	GAGATAATCCGACAGAACATCAGC	ATTGCTTCTCCTACGCATTACC	>60% intron	KJ514972–KJ515066	95	1248	4287	7	8	5535	77.5
DC16209	I	734	TPM2uf+G (GTR+G) [HKY+G]	GCCTTCCAGCAACATTCATTC	TGTCATTGATTCTGGTGATGGTG	>60% intron	KJ519530–KJ519624	95	1170	11725	14	15	12895	90.9
DC16308*	VII	2249	TrN+G (GTR+G) [HKY+G]	AGTCGAATGCAACAAGAGTTAGC	GCTGGTCTACGCTTGAATCATC	>60% intron	KJ522336–KJ522436	101	1022	5241	5	6	6263	83.7
DC16340	I	3117	TVM+I (GTR+I) [HKY+I]	CCCATTTGCTTCCGAACAGTAG	GATGAAGAACAACATGCTCATCAGC	>60% intron	KJ521675–KJ521754	80	1119	7091	7	8	8210	86.4
DC16577*	VII	717	HKY+I+G	GAATCGCCATCCCCAATCC	GGTTTCATCTTCTACCAGCAGTTC	<60% intron	KJ521366–KJ521465	100	1527	8239	10	11	9766	84.4
DC16645*	V	1010	TPM2uf+G (GTR+G) [HKY+G]	CAAGCGTATTGTTCCAACTGC	GAATCCATCCATCGGGAGAAC	>60% intron	KJ521057–KJ521163	107	1260	5596	11	12	6856	81.6
DC16778	V	639	TPM2uf+G [(HKY+G)]	GATAACATGTCGGGGATACTGC	GGGGTCCAGCAGATATCTTTG	<60% intron	KJ515679–KJ515769	91	1074	6072	10	11	7146	85.0
DC16928	I	939	HKY+G	CACCCAATACTCCATCGACTG	GATTGGTATCGTACTGGTGTTGG	>60% intron	KJ518344–KJ518438	95	834	4444	7	8	5278	84.2
DC17018	VII	946	TPM2uf+G [(HKY+G)]	ACAGTGAAACCACTTGCACC	AAGCCAGGAAGAGCAAGATG	>60% intron	KJ515258–KJ515353	96	249	680	2	3	929	73.2
DC17278	VI	486	GTR+I+G [HKY+I+G]	CGTAACTCCTTACCCGAGAAC	GCTCCGTAACAACAACGCTAC	exon only	KJ519255–KJ519351	97	1266	0	0	1	1266	0.0
DC17284	I	883	TVM+I (GTR+I) [HKY+I]	GCCATTTCCCCTTAAACGAGTC	GCTCCTGTTCGTGAAGTTTGG	>60% intron	KJ515451–KJ515535	85	2190	2527	9	10	4717	53.6
DC17311	VI	1479	TIM2+G (GTR+G) [HKY+G]	CATTCCAACCAGTACAGAAGAGAT	CACCCAGCATCAAATGAGC	>60% intron	KJ519090–KJ519176	87	948	2751	4	5	3699	74.4
DC17915	I	1702	TVM+G (GTR+G) [HKY+G]	TAAGCAGAATCCGCAAAGTGC	AGGAGGTTTTAGTGAGGGAGG	>60% intron	KJ521270–KJ521365	96	432	1093	3	4	1525	71.7
DC1877	V	1073	HKY+G	GGCCATTTGTTGATCATTGGTG	CGGGAAATAGCTTTATTGTGCC	>60% intron	KJ515163–KJ515257	95	567	5619	7	8	6186	90.8
DC18883	I	1398	TrN+G (GTR+G) [HKY+G]	GCCCTTTGACTGCCTCTTC	GATGCTTCCCTTGACGATGATG	>60% intron	KJ517749–KJ517810	62	519	893	3	4	1412	63.2
DC19270	II	836	HKY+G	GAGAAAAGCCTCCTGGATCAC	GCGGGGTAAAAAGTCTCCTTC	>60% intron	KJ515866–KJ515956	91	1374	3931	13	14	5305	74.1
DC20026	VII	1354	TPM1uf+I [(HKY+I)]	GAGAGTCAAGGAGAAGATGAAGC	CCCTTGCCATTTGTTGCATC	>60% intron	KJ517659–KJ517748	90	369	636	3	4	1005	63.3
DC20383	IV	957	TIM1+I (GTR+I) [HKI+I]	CCCGCTGTTATCAAACACACA	ACTATAAATTCCGAGAGATCCAAGC	>60% intron	KJ516932–KJ517027	96	2148	9189	15	16	11337	81.1
DC20459	II	667	HKY+G	TCCAACAAACAGTCAATGCTCATC	GGAGAAGAAACATGTCTCTACTGC	>60% intron	KJ516836–KJ516931	96	2133	1486	7	8	3619	41.1
DC20759	I	985	TPM2uf+G [(HKY+G)]	GGCCATTTCAGCACTTTGC	AAATTCAGCTTGGAGTTGACCC	<60% intron	KJ519719–KJ519809	91	1533	15841	12	13	17374	91.2
DC2290	VIII	2368	HKY+G	GTTCTTCTTCAGCTCTGGATGG	GCTATTCATGTGTTGTCCCTGAC	>60% intron	KJ519625–KJ519718	94	687	3933	7	8	4620	85.1
DC23389	IV	2426	TPM3uf+I+G [(HKY+I+G)]	TCACAAAACCAATCGCAGGG	CATTGGGTCCAGTGAGTACC	>60% intron	KJ518914–KJ518993	80	1812	6372	16	17	8184	77.9
DC246	VI	1573	TIM2+I (GTR+I) [HKY+I]	CAAACAAGTGACCCTAGAAGTGG	GCTTAATGCCCAGAAAGTGCTC	>60% intron	KJ519810–KJ519905	96	2748	5626	11	12	8374	67.2
DC25791	VI	1637	HKY+G	AGCTTTGGCACAAGAGTCTTG	AGTTGGTTGCCGAGTTTCTATG	>60% intron	KJ516288–KJ516374	87	2451	4486	9	10	6937	64.7
DC2600	II	631	TIM1+G (GTR+G) [HKY+G]	TTGCGGTGGGTGTTTATGG	TTTCCTCCACTTTCGCCTTC	exon only	KJ520261–KJ520339	79	1239	0	0	1	1239	0.0
DC26617	VII	905	HKY+G	ACCACCTAGTCTCCACTGAC	CAGAGGTTACCTGCGAGTAC	>60% intron	KJ520503–KJ520599	97	1104	460	2	3	1564	29.4
DC26889	V	678	TPM3uf+G (GTR+G) [HKY+I]	CCTTAAGACCTTCTGCATTTGC	CTTGTGAATCATTGCCTGCTTC	>60% intron	KJ520434–KJ520502	69	3039	6556	22	23	9595	68.3
DC27381	III	1189	GTR+I+G [HKY+I+G]	GGCACCTATTTCTGTGTTGACA	TGCAGCCTTGTCATCATAGTG	>60% intron	KJ516653–KJ516743	91	456	2639	5	6	3095	85.3
DC2772	II	645	TIM2+I [(HKY+I)]	GGTATTCACAGCGAGGAGG	ATCAATGCGATGGCGTAACTC	>60% intron	KJ520664–KJ520760	97	1656	5640	14	15	7296	77.3
DC28180	V	1145	TPM2uf+G [(HKY+G)]	TGCTGACTTTTCCAAGGAAGTG	GAGAGTTGGCAAGGTGAACTG	>60% intron	KJ518994–KJ519089	96	1254	5829	12	13	7083	82.3
DC282	VI	878	TMP3uf+G [(HKY+G)]	ACATGGTTCAGATGGTGCTG	CACTACCTCTGCTAACCACC	<60% intron	KJ518817–KJ518913	97	978	333	1	2	1311	25.4
DC28973	I	1896	HKY+G	AGAACAGAATGGGCAAGTGG	GCGCATCCCTCCATTATCAG	>60% intron	KJ518249–KJ518343	95	1746	1407	4	5	3153	44.6
DC2965	II	686	HKY+I+G	TATGTGCAGCCTGTAACGATTG	AGACTCTCACTCCTCTTCACC	exon only	KJ517586–KJ517658	73	1461	1229	2	3	2690	45.7
DC30826	V	1118	TIM3+I (GTR+I) [HKY+I]	CCATTGCTAGATGTTGAAACACCC	GACAATGGTGGTCAAATCCCT	>60% intron	KJ514878–KJ514971	94	1008	3321	11	12	4329	76.7
DC31753	IX	676	GTR+I [HKY+ I]	CAGCTATGCTTCATGACTGTACC	CCAAGAGTAACCACATCTGTCTTG	>60% intron	KJ517490–KJ517585	96	1218	3969	8	9	5187	76.5
DC32391	IX	722	TIM1+G (GTR+G) [HKY+G]	AGGCATCTGCTGTGGAAATTG	ACAGCCTATCCATCAATCAATGC	<60% intron	KJ515067–KJ515162	96	576	2512	6	7	3088	81.3
DC3277	VIII	1116	TPM1uf+G (GTR+G) [HKY+G]	GTGCGGAGAATCATGTCGATG	ACAAAGCTGCGGTCCAAAG	>60% intron	KJ514693–KJ514781	89	2529	7766	9	10	10295	75.4
DC32914*	VI	631	TIM3+G (GTR+G) [HKY+G]	TGCTTATTTGCTCCGAGATATGC	CACAAATCGGGTCCTAGTCC	exon only	KJ521164–KJ521269	106	2685	0	0	1	2685	0.0

TABLE 2. Continued.

Markers[a]	Linkage group	Aligned length[b]	Model of evolution[c]	Forward primer (5′–3′)	Reverse primer (5′–3′)	Intron content	GenBank nos.[d]	No. of accessions with sequence data	Exon[e] (nt)	Intron (nt)	No. introns	No. exons	Length of gene (nt)	% of intron
DC3374*	VI	1292	TIM3+G (GTR+G) [HKY+G]	CTTCTTCCATGCGACCCA	CTACAGGGAGCGAAAGAAATATGT	>60% intron	KJ522143–KJ522246	104	2073	10629	20	21	12702	83.7
DC34333	I	709	GTR+G [HKY+G]	GAAATCTGCATATCCCCTAAAGGC	TGCTGCCTCAGAATTGTCATTTG	<60% intron	KJ516744–KJ516835	92	1221	866	5	6	2087	41.5
DC35097*	VI	761	TIM1+I+G (GTR+I+G) [HKY+G+I]	GGTAACTTCTTGTAATCTGCTCCC	CACTATGTAGCACCATCATTCTGG	<60% intron	KJ521569–KJ521674	106	2289	6104	13	14	8393	72.7
DC35833	I	1235	TPM2uf+G [(HKY+G)]	ACTTCCAAGTGTTGTTTGTCCA	GGGACATATGAGGCTGTTAGATAC	>60% intron	KJ520340–KJ520433	94	225	1421	2	3	1646	86.3
DC38098	VI	2070	TIM1+I (GTR+I) [HKY+I]	AACTTCATGAGTTCTTCATCCACC	CATTGGTCAACTGGAAGGGAC	<60% intron	KJ514602–KJ514692	91	1140	2473	4	5	3613	68.4
DC38583	I	2426	TrN+G (GTR+G) [HKY+G]	CCAATCTCCGACATCAAACATGC	GGCCGTTGAGGTTCTTGTTG	>60% intron	KJ519906–KJ519997	92	2151	8715	18	19	10866	80.2
DC3902*	VII	1334	TPM2uf+G [(HKY+G)]	CAAGCCAATTACCAGGAGAGC	CGTCTCTCATAGTTGTCCACAC	>60% intron	KJ521939–KJ522038	100	507	779	1	2	1286	60.6
DC42375	IV	1148	TIM3+I+G [(HKY+I+G)]	TCAAAAATGACCAGCTAGTTGAGG	CTCCCAGTAAAGTTGGTTGCAG	>60% intron	KJ515354–KJ515450	97	1515	2052	7	8	3567	57.5
DC42660	VI	680	TIM1+I+G [(HKY+I+G)]	AGAACTGCTAAAGAAGATGGTTGC	CTGTTGCTCCAATATCCTCTCTTC	exon only	KJ514316–KJ514412	97	1578	0	0	1	1578	0.0
DC43804	IX	2792	HKY+G	CCTGACAAGATAAGGCAGAACATC	TAGAATTCAGAAGAGCATCCTTCG	>60% intron	KJ517124–KJ517211	88	1386	6274	12	13	7660	81.9
DC45860	V	847	TIM2+G (GTR+G) [HKY+G]	GTGAAGCAACTGCCTTCTCTAA	AGGCCCAGGTTTTTATTGAGG	>60% intron	KJ517811–KJ517905	95	1206	4348	8	9	5554	78.3
DC46197	III	1203	TIM2+G (GTR+G) [HKY+G]	CTGGTTTTTACACTGCCGAAAAG	CTTTGTTTCAGGATGATCAGGTGC	>60% intron	KJ519446–KJ519529	84	1020	2325	9	10	3345	69.5
DC46683	VI	1362	TIM2+G (GTR+G) [HKY+G]	TCATCATCACAAAACACCTCACTC	ATAAGCGTATTGGTTCGGCTG	>60% intron	KJ517212–KJ517304	93	3435	12991	25	26	16426	79.1
DC46818	VII	945	TPM2uf+G [(HKY+G)]	ACCAAATAGCTACTGCTTGGATC	CGTAAACTCCTAAGCTGTAGTGC	>60% intron	KJ515588–KJ515678	91	3876	8175	20	21	12051	67.8
DC47239	IV	1241	TPM3uf+I (GTR+I) [HKY+I]	GCAACTTTGATCTATTATGGAGTCCG	GTCAGGCTGACTGCTTTCTC	>60% intron	KJ519177–KJ519254	78	835	924	2	3	1759	52.5
DC48708	IV	1521	GTR+G [HKY+G]	TCGACGAAAGAAATTGGAGTGATG	GTGTCTTTGAGAGTCCTTCCC	>60% intron	KJ514050–KJ514143	94	1347	2886	6	7	4233	68.2
DC49764	I	488	TPM3uf+I [(HKY+I+G)]	CTGCCAGGAACAAGACGAC	GAAAACGGAACCAATAGTCAAGGA	<60% intron	KJ515957–KJ516051	95	1647	4940	11	12	6587	75.0
DC49801	IX	778	TrN+G [(HKY+G)]	GTAGAAGCATTAATAACACGAGGC	GGAAAGAAAAGTTGCAGACATGC	>60% intron	KJ517305–KJ517399	95	2988	14377	17	18	17365	82.8
DC51684	I	1084	TVM+G (GTR+G) [HKY+G]	GATTGGTACCTCAGAGCATGG	CTAATCCAGCCAGTCCACC	>60% intron	KJ516375–KJ516463	89	1089	3591	10	11	4680	76.7
DC54236	III	627	TPM1uf+I [(HKY+I)]	CATCTCATCCGGTTCCTTTGG	CCCTCTGAAGCCCTAAAATCATTG	<60% intron	KJ518536–KJ518630	95	2877	0	0	1	2877	0.0
DC54335	VII	804	TrN+I+G [(HKY+I+G)]	CGGAAAGAGTGTTTTTGTAAGTGC	CTTGTGGTGGAAGAGTTGTACAG	>60% intron	KJ518631–KJ518725	95	2139	15201	18	19	17340	87.7
DC56413	VII	619	TIM2+I +G [(HKY+I+G)]	CTCTCATTACCCATGGCAACTC	AAAGAGTGCAACACAGAGGG	exon only	KJ516052–KJ516126	75	456	0	0	1	456	0.0
DC57194	II	574	TPM3uf+G [(HKY+G)]	CGGGCAATATACATCCTGCATC	GCCTGCATCAAAGGTTGC	<60% intron	KJ513880–KJ513976	97	1212	7137	9	10	8349	85.5
DC57966	II	484	HKY+G	AGTGGAATGTCCGAAAAGGC	GCATAGATAGAATGCCGCTGC	<60% intron	KJ517906–KJ517984	79	867	1687	9	10	2554	66.1
DC58732	II	1712	TPM3uf+I [(HKY+I)]	AGGGACAAGGCATTTATCGC	CATCCTTGTACAGTAGATAGCGTG	>60% intron	KJ517028–KJ517123	96	882	7939	12	13	8821	90.0
DC58941	II	2084	HKY+I+G	GATGTCTCTATCAACACGATCTGC	CTTGGAAAACGACAACATGGTG	>60% intron	KJ521846–KJ521938	93	657	3234	6	7	3891	83.1
DC59353	V	2252	TPM3uf+G [(HKY+G)]	GGTCTGGCATTTGGAACTGG	GGCATTATGTCGTGCAGCAG	>60% intron	KJ516219–KJ516287	69	1083	2246	6	7	3329	67.5
DC68	IV	857	TPM2uf+G [(HKY+G)]	ATTCTGAAGTCATAGCTGCTCAAG	CCAAGATAGTCCAAGCATGAAGC	>60% intron	KJ518439–KJ518535	97	1326	3643	5	6	4969	73.3
DC7428	I	601	HKY+G	GGAAGCCTATCTGTTTCCCAG	GTTGGTTTGAGAGATCTATGGTGG	<60% intron	KJ518056–KJ518151	96	1254	8598	9	10	9852	87.3
DC8031	V	960	HKY+G	ACATTTGTTGCTCTCTTATCCACC	CACCTGCCTCTCTTCCTTC	>60% intron	KJ518152–KJ518248	97	1726	4787	8	9	6513	73.5
DC8465	IX	934	TrN+G (GTR+G) [HKY+G]	AGTGCAAATGGAGATCATACTGC	ATGGCGAGTACCACCAAAC	>60% intron	KJ519352–KJ519445	94	1182	3189	7	8	4371	73.0
DC8584	VIII	1817	HKY+I	CAATCCTTGGTTCCTCAGCAC	CGGGCATGTCTTTGGGTG	>60% intron	KJ517985–KJ518055	71	834	2422	9	10	3256	74.4
DC8796	III	1931	TIM3+G [(HKY+G)]	TTTGATGGCTCGTGGTCC	GCTTTCTTCCTCGTCCTTTCTTTG	>60% intron	KJ514144–KJ514235	92	312	3026	3	4	3338	90.7
DC8872	VI	593	TPM3uf+G (GTR+G) [HKY+G]	AAGCATCACAACCTGCATCTC	CTGATTCTCAGTACTTTGAAGAGCTG	<60% intron	KJ517400–KJ517489	90	882	2536	5	6	3418	74.2
DC8990	I	600	TPM3uf+G [(HKG+G)]	GTTACTCCACAACCTTCTGTACC	GGTTATGCCTCCACTATGGTG	<60% intron	KJ514505–KJ514601	97	1899	6779	12	13	8678	78.1
DC9014	I	1053	TrN+G [(HKY+G)]	CAGGCAACCGAGGTGTAATTC	TGTGGGCTTTACGAAAGGAAAC	<60% intron	KJ522247–KJ522335	89	697	6171	4	5	6868	89.9
DC9186	VII	1546	TIM3+G (GTR+G) [HKY+G]	ATGCTGGGTTGCTTTATTGGG	AGTTGCTTCCTTCAAGTCCG	>60% intron	KJ516127–KJ516218	92	1080	1626	3	4	2706	60.1
DC9206	I	1197	TVM+I (GTR+I) [HKY+G]	ACCAATTCAGTCTGTGGCAAC	TTCAAGAAAGGTTTCAGAGAGGC	>60% intron	KJ519998–KJ520088	91	1279	1083	3	4	2362	45.9
DC9853	VII	1754	TVM+G (GTR+G) [HKY+I+G]	ACTCTTCAAACAGGCAACAGC	CTGATTTGAGAAGTCCAGTGCG	>60% intron	KJ513977–KJ514049	73	2478	2897	7	8	5375	53.9
DC9952	VI	2801	TIM2+G (GTR+G) [HKY+G]	CGGATTGCGTTCAGTAAATTCC	CCTCCTTCGGCTGAATATGTG	>60% intron	KJ516464–KJ516558	95	918	1015	2	3	1933	52.5
Concatenated length		111 166				18.4% exon, 81.6% intron		8557						

[a] An asterisk designates the 10 markers used in the reduced data set.

[b] Refers to the aligned length of the data set with homopolymers trimmed to a maximum of six.

[c] The first model of evolution was obtained in jModelTest. The second model (if present, in parentheses) was the next lower AIC value accepted in MrBayes analysis and *BEAST analysis. The HKY model (if present, in brackets) with the lowest AIC value was used in our *BEAST analysis. Markers in parentheses and brackets were used in MrBayes and *BEAST analyses.

[d] See online Appendix for specific information of accession/marker codes.

[e] Columns 10-15 refer to characteristics of the entire genes, not the smaller portions of the genes sequenced here.

Table 3. Tree scores for maximum parsimony analyses.

	Maximum parsimony tree parameters							
	94 markers, 97 accessions, 1 allele, homopolymers present vs. trimmed		1 or 10 markers, 107 accessions, 1 allele, homopolymers trimmed		94 markers, 107 accessions, 1 allele vs. 2 alleles combined with ambiguity codes, homopolymers trimmed		94 markers, 107 accessions, 1 allele, homopolymers trimmed, intronic vs. exonic regions	
Tree statistics	Homopolymers present	Homopolymers trimmed	1 marker allele (DC10366)	10 markers	1 allele	2 alleles	Intron	Exon
No. characters	112 002	111 166	1128	11 480	111 166	115 882	90 688	20 478
Parsimony-informative characters	21 193	21 011	270	2903	21 502	21 348	18 711	2791
No. parsimonious trees	1	2	1	3	6	16	2	1
Length	92 908	91 361	784	10 122	92 859	74 516	81 160	10 904
Consistency index	0.530	0.530	0.732	0.608	0.530	0.641	0.537	0.525
Retention index	0.729	0.732	0.898	0.830	0.736	0.812	0.739	0.704
Rescaled consistency index	0.387	0.388	0.658	0.505	0.390	0.520	0.397	0.370
Fig. or supplemental Appendix no.	App. S2	App. S3	Fig. 3	Fig. 4	Fig. 2	App. S4	App. S5	App. S6

comparable values. Concordance factors within the subspecies of *Daucus* are low. On the other hand, *D. sahariensis* and *D. syrticus* are grouped in a clade with a high CF (0.69). Clades A and B are supported by concordance factors of 0.348 and 0.396 respectively, which translates to 32.7 and 37.2 genes, respectively (by multiplying the number of genes, 94, by the concordance factors).

Species tree estimation—The analysis using 104 accessions of 27 species and 94 genes with their corresponding model of evolution as indicated in Table 2 column 4 exhibited some ESS values lower than 100 for the posterior, prior, among other parameters. It took 89 d for this analysis to reach completion using the PC mentioned in the Materials and Methods. These low values have been reported by other researchers in the web site for users of *BEAST (https://groups.google.com/d/forum/beast-users). Andrew Rambaut, a coauthor of the *BEAST package, indicates in the website above that the low ESS values are obtained because of a problem when using GTR evolutionary models and Jeffrey's priors (Jeffreys, 1946). He suggested that we also run *BEAST using the HKY models only, because in this case Jeffreys prior provides better statistical properties for estimating the kappa parameter (Drummond et al., 2002). We tried this with a reduced data set of 37 accessions containing all 22 *Daucus* ingroup species and representative outgroups: *Astrodaucus littoralis*, *Caucalis platycarpos*, *Orlaya daucoides*, and *Turgenia latifolia*. ESS values were still lower than 100 for the posterior, prior and other parameters, but higher than the ESS values of the previous analysis.

Figure 7A presents the coalescent analysis using 104 accessions of 27 species. The topologies of this analysis presented significant differences to the MP, ML, and BCA analyses that were largely concordant with each other. The most notable changes in this species tree analysis relative to MP, ML, and BCA are: (1) Clades A and B are no longer coherent; (2) *Astrodaucus littoralis*, *Ammi visnaga*, *Caucalis platycarpos*, *Torilis nodosa*, *Orlaya daucoides*, *O. daucorlaya*, and *Turgenia latifolia* resulted as ingroups to *Daucus.* Figure 7B presents the analysis using a subset of 37 accessions of 22 species. Again, clades A and B are no longer coherent. *Astrodaucus littoralis*, *Caucalis platycarpos*, and *Turgenia latifolia* resulted as ingroups to *Daucus.*

DISCUSSION

Use of next-generation sequencing—The Roche 454 sequencer was released in 2005 (Margulies et al., 2005) and is considered the first commercially available next-generation sequencing platform (Rothberg and Leamon, 2008; Egan et al., 2012). The 454 technology utilizes the pyrosequencing method described by Dressman et al. (2003), providing a mean read length of 700 bp, similar to that obtained by current Sanger capillary technology (Sanger et al., 1977), but at a lower cost per read. Other competing technologies, such as Illumina provide higher coverage but suffer from shorter read lengths (Egan et al., 2012) or have longer reads averaging 8500 bp but with higher error rates, such as Pacific Bioscience (Egan et al., 2012; Koren et al., 2012; Pacific Biosciences, 2013). One weakness of the 454 technology is inaccurate estimation of homopolymer region lengths. By the time our study was in its design stage, the 454 sequencer had already gained a high reputation in the scientific community, shedding light on problems in human genetics, metagenomics, ecology, evolution, and paleobiology (Rothberg and Leamon, 2008). We chose it mainly due to its read lengths, providing individual gene topologies with potential taxonomic resolution.

Phylogenomic analysis—This study is the first phylogenomic analysis of *Daucus,* the economically most important genus in the Apiaceae, using next-generation sequencing technology. *Margotia gummifera* was sister to those *Daucus* with $2n = 18$ chromosomes, concordant with Spooner et al. (2013). Spalik and Downie (2007) and Spalik et al. (2010) provide phylogenetic and chronogramic analyses of *Daucus* based on ribosomal DNA sequence variation that includes more species than we used here, and with many concordant results to our MP and ML concatenated results. Most notably, they support two main clades of *Daucus*, *Daucus* clade I and *Daucus* clade II (here labeled as clade A and clade B). Furthermore, they also report additional non-*Daucus* species (*Pseudorlaya pumila* and *Turgenia latifolia*), which were sampled here, as being within clade A. Our study uses multiple accessions per species and indicates that it was not possible to clearly distinguish the subspecies of *D. carota*. Furthermore, as highlighted below, this

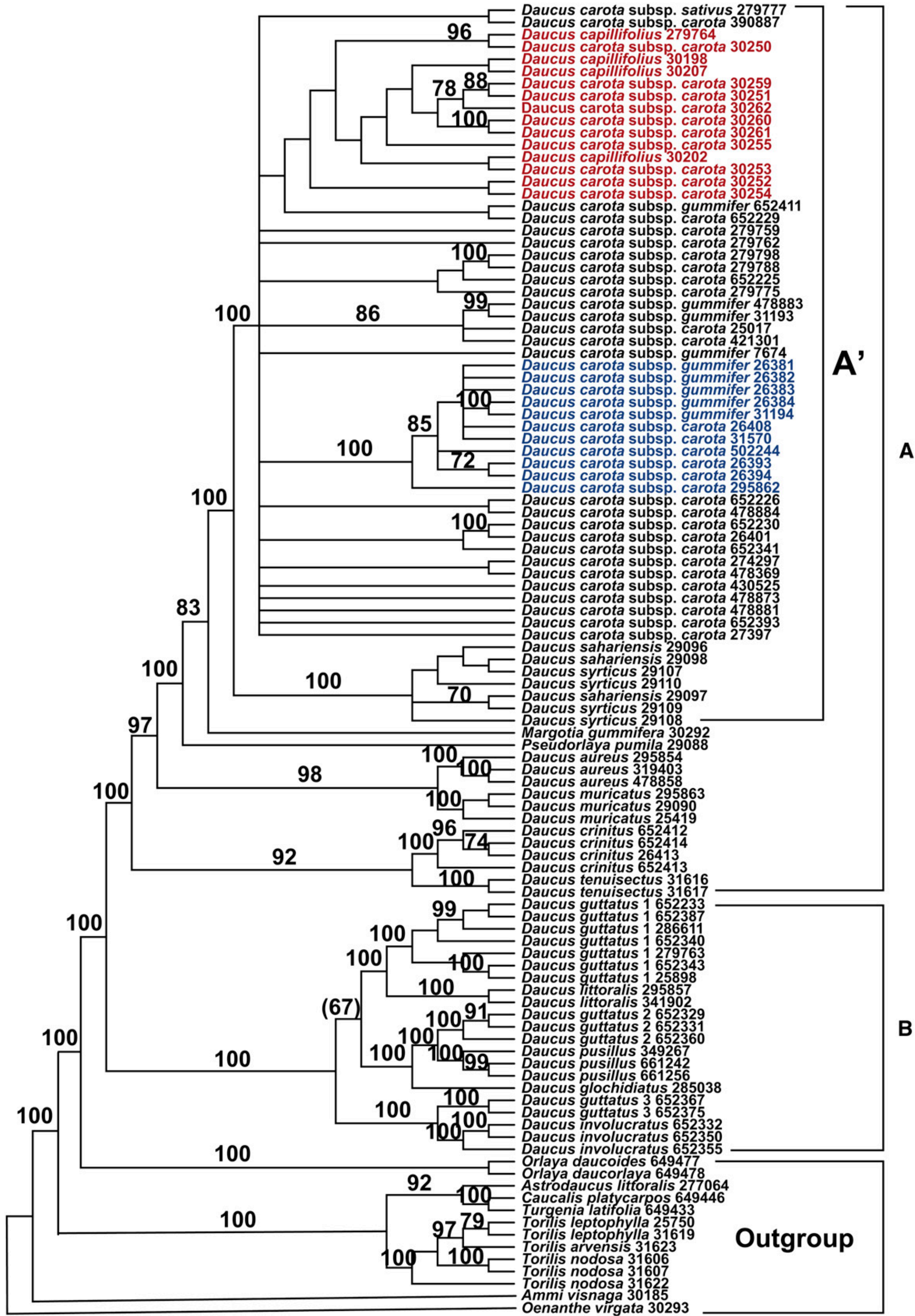

Fig. 2. Phylogeny of *Daucus* from a maximum parsimony analysis using a data set with the allele of highest coverage, homopolymers shortened to a maximum of 6 bp, and based on 94 nuclear orthologs and 107 accessions. All accessions with known locality data of *D. capillifolius* and *D. carota* collected in Libya and Tunisia are highlighted in red; most accessions of *D. carota* collected in Portugal and Spain are highlighted in blue.

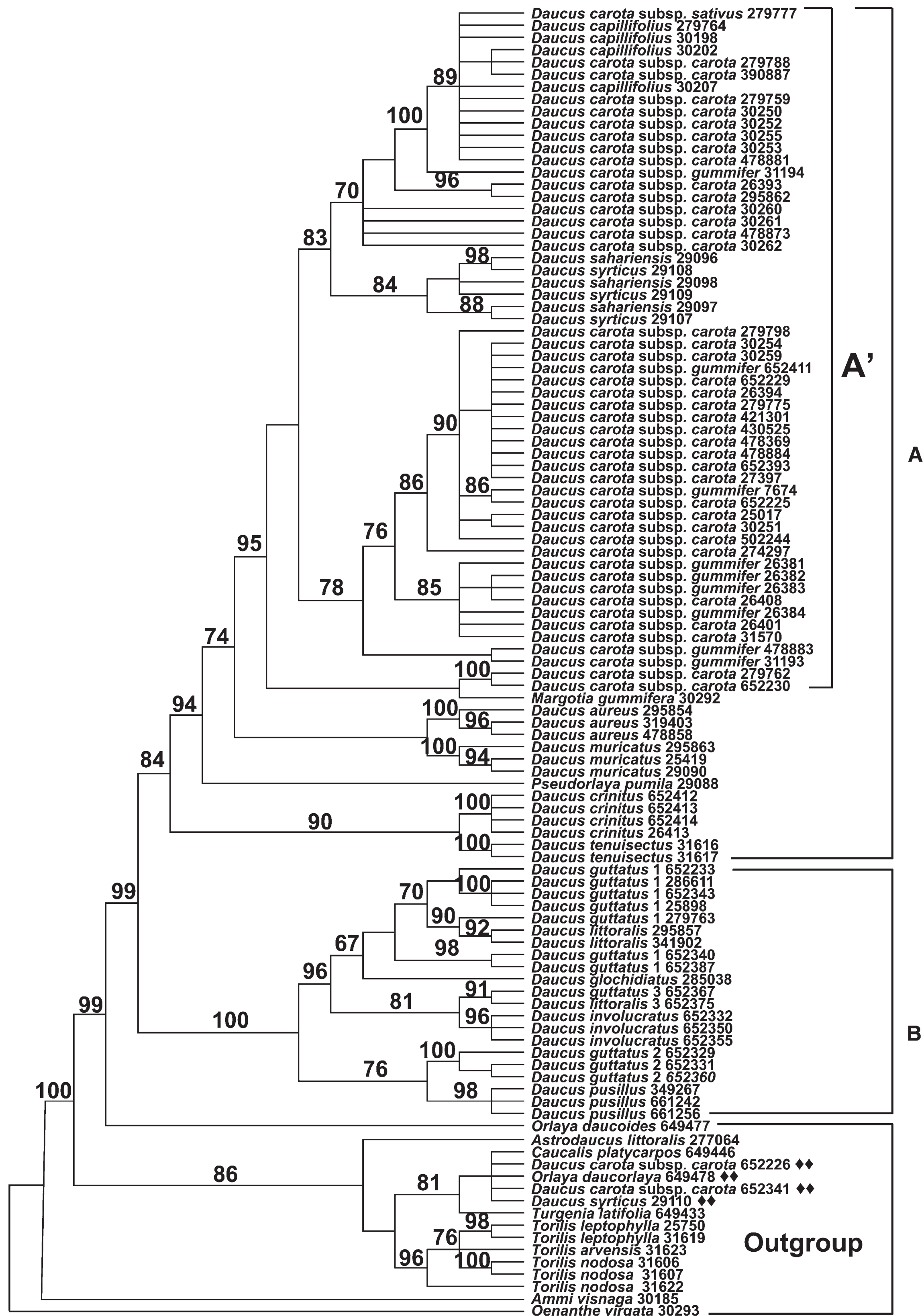

Fig. 3. Phylogeny of *Daucus* from a maximum parsimony analysis using a data set with the allele of highest coverage, homopolymers shortened to a maximum of 6 bp, and based on one nuclear ortholog (marker DC10366) and 107 accessions. The four accessions in the outgroup clade designated by double triangles are misplaced relative to the dominant tree topologies.

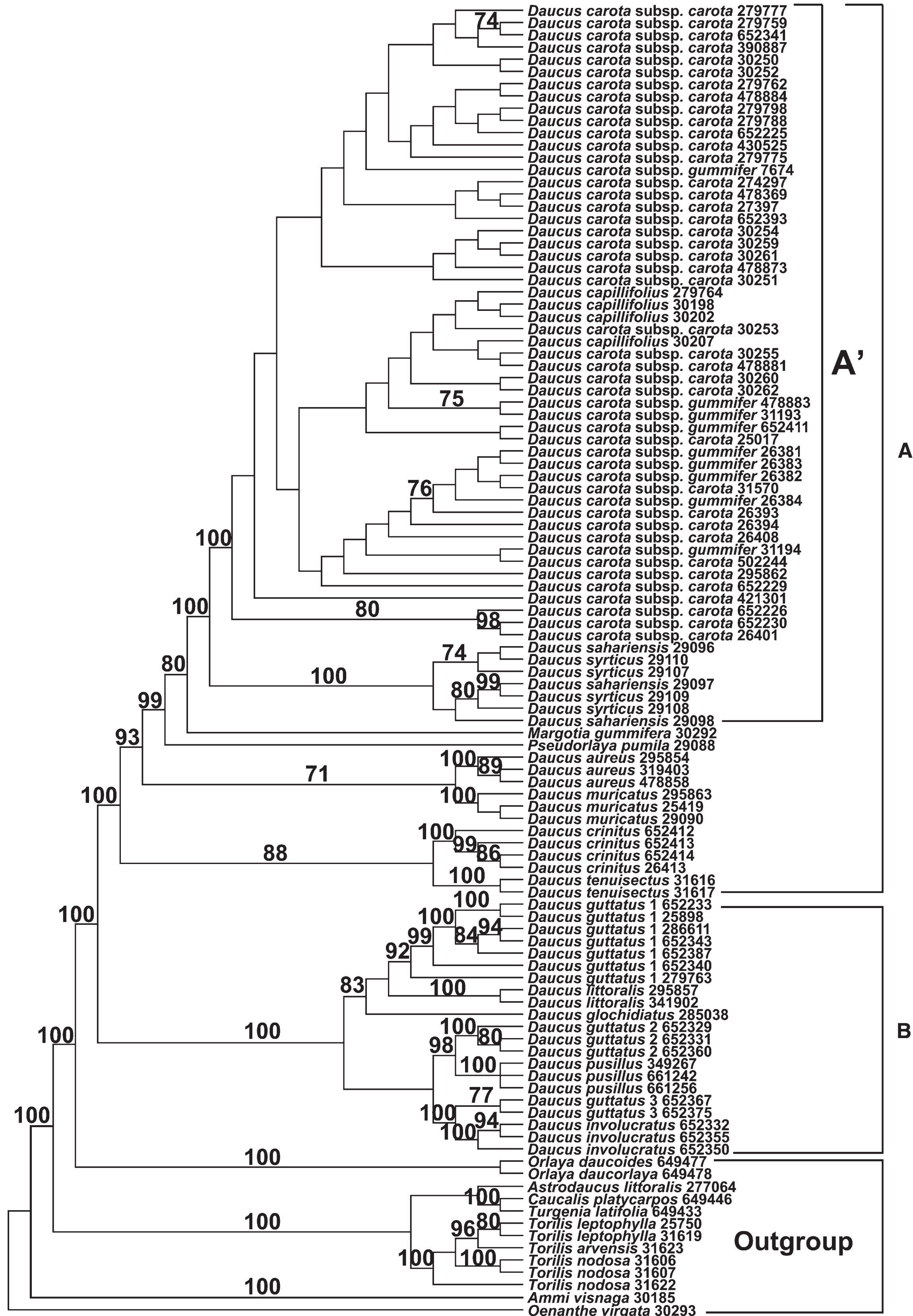

Fig. 4. Phylogeny of *Daucus* from a maximum parsimony analysis using a data set with the allele of highest coverage, homopolymers shortened to a maximum of 6 bp, and based on 10 nuclear orthologs and 107 accessions.

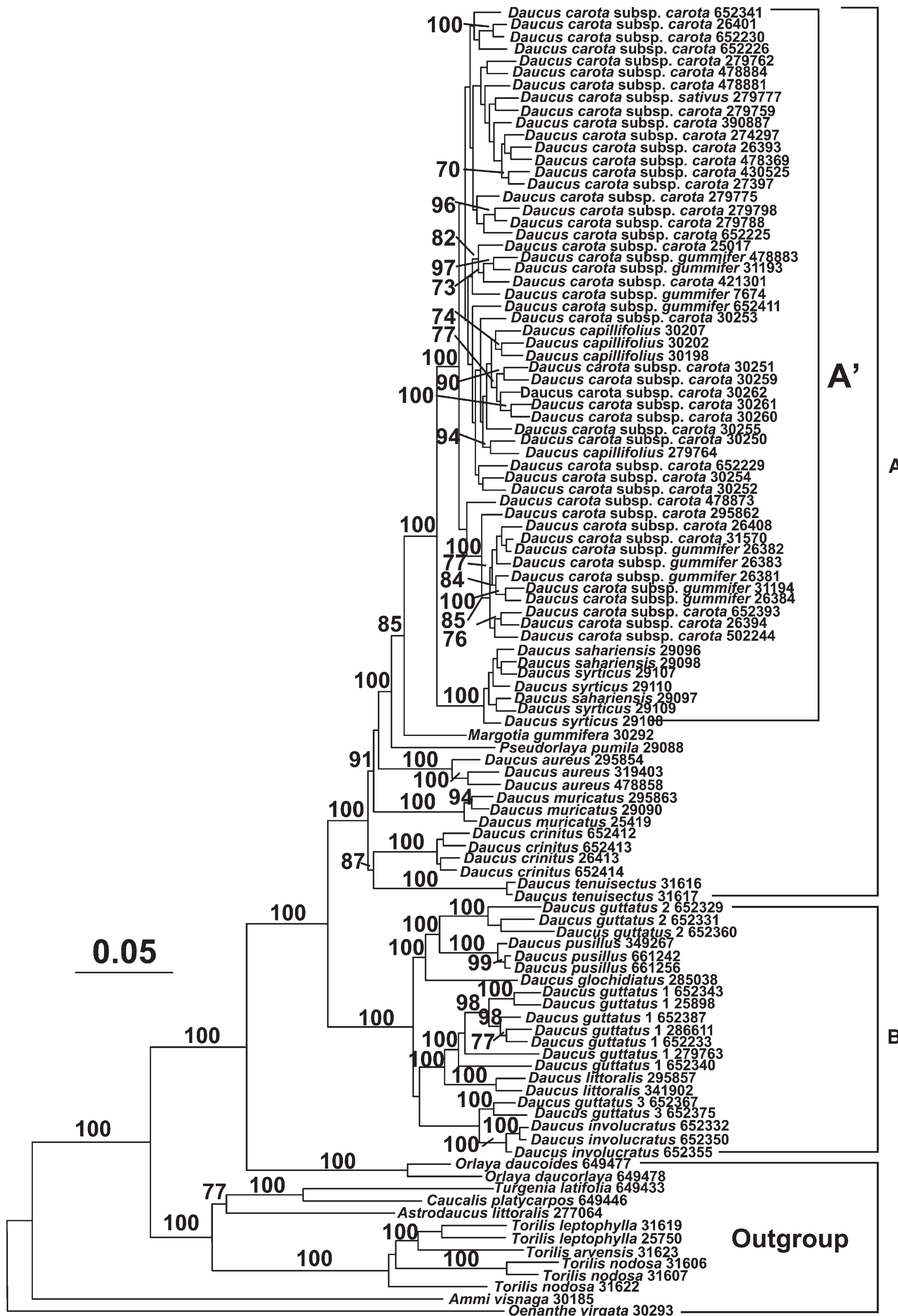

Fig. 5. Phylogeny of *Daucus* from a maximum likelihood analysis using a data set with the allele of highest coverage, homopolymers shortened to a maximum of 6 bp, and based on 94 nuclear orthologs and 107 accessions.

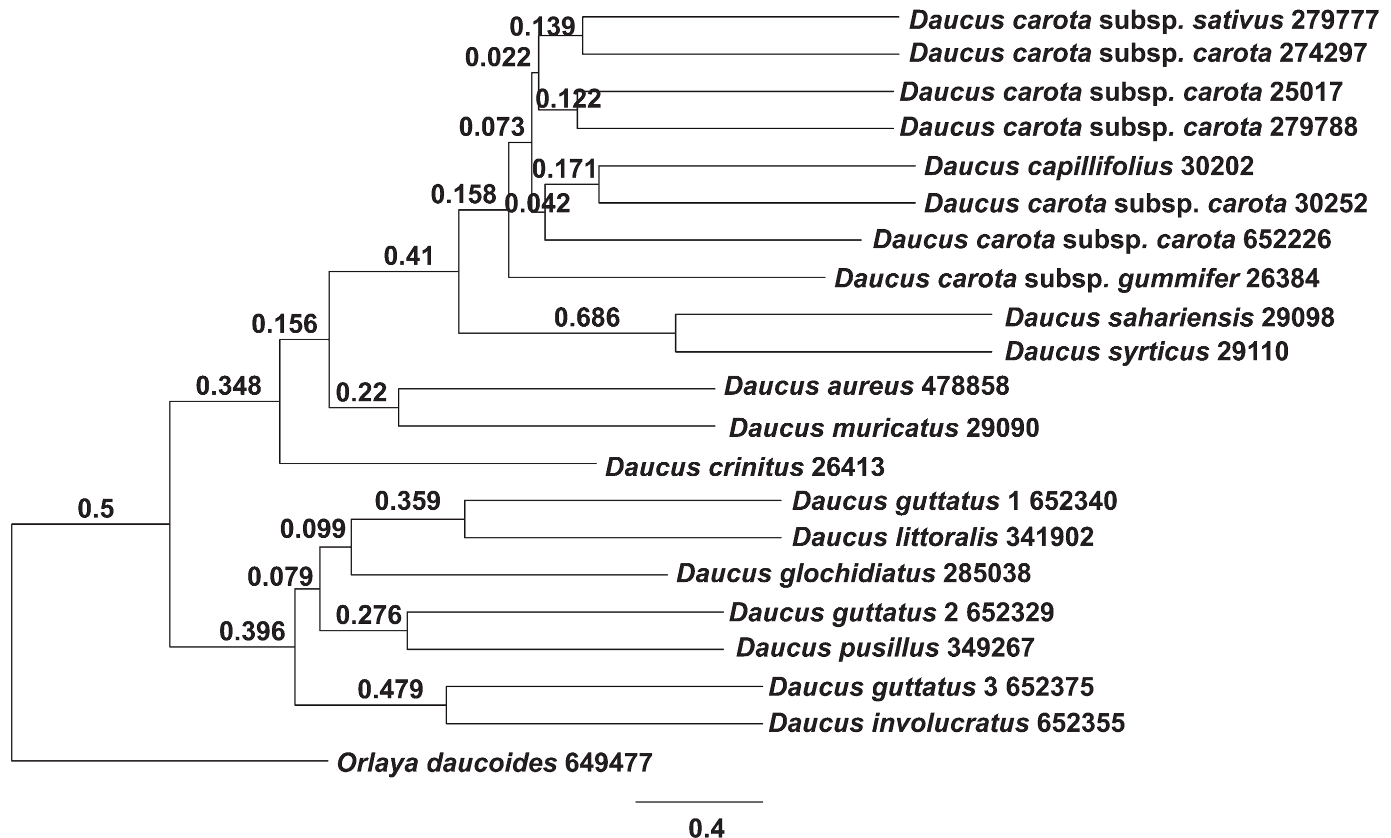

Fig. 6. Primary concordance tree obtained with Bayesian concordance analysis in *Daucus* with 94 nuclear orthologs and 21 accessions. Numbers above the branches are the concordance factors, which do not show significant differences for different α values (0.1, 1, 10, and infinite).

study provides data on well-resolved substructure within *D. broteri* and *D. guttatus* that may indicate separate species status for these accessions. We also provide a better-resolved substructure at the base of clade A and in clade B.

Our results supported two subclades within clade A′ that group wild *Daucus carota* accessions collected in (1) Tunisia and Libya and in (2) Portugal and Spain. This result partially matches that of Iorizzo et al. (2013), using SNP data, who grouped *D. carota* subsp. *carota* and *D. capillifolius* from northern Africa, separate from *D. carota* from Europe. However, our results failed to separate *D. carota* subsp. *carota* from subsp. *gummifer*, a separation that was found by Iorizzo et al. (2013). Clearly, the accessions of *D. carota* (and *D. capillifolius*) are very closely related, and multiple nuclear orthologs are inappropriate markers to examine their relationships. We are further exploring phylogenetic relationships in clade A with SNP data gathered from genotyping by sequencing (GBS) from many additional accessions of *D. carota* from other described subspecies and geographic areas. Our results failing to distinguish *D. capillifolius*, *D. carota* subsp. *carota*, and *D. carota* subsp. *gummifer* could be a result of gene flow, or from multiple origins of these morphotypes, or may be a result of the inappropriateness of nuclear orthologs to separate these closely related taxa. The fact that the taxa in clade A′ all share 18 chromosomes, experimental and field data document ease of gene flow, and they are closely related as documented here suggest that they may be easily incorporated in carrot breeding programs.

Accession numbers 286611, 652387, and 25898 were grouped in clade B. Originally, they were identified as *Daucus carota* (no subspecies designation), *D. carota* subsp. *fontanesii* and *D. carota* subsp. *major*, respectively. These names correspond to those provided by the Germplasm Resources Information Network (GRIN) database. However, in light of our results, and after reevaluation of the morphological information at the Germplasm Resources Information Network, we tentatively labeled these accessions as *Daucus guttatus* 1. Further analyses, including morphological information of all accessions, are needed to determine whether there are more cases of misidentification in the germplasm bank of the USDA.

Our final data were reduced by 836 bp (0.74%) with the shortening of homopolymers to a maximum of 6 bp. In addition, our data set had 6.1% missing data. We did not observe significant differences among the topology of trees with trimmed vs. untrimmed homopolymers (Fig. 2 vs. Appendix S3), showing that our aligned data set was not sensitive to a reduction of such data. We suspect that this trend will be present in other large phylogenomic data sets. No major topological differences in simulated studies using missing data on large alignments of eukaryotes were found by Philippe et al. (2004). To accurately reconstruct the phylogeny of an organism, the number of genes used is considered a more important factor than taxon number (Rokas and Carroll, 2005). It is useful to determine a number of genes that approaches the dominant topology of *Daucus*. We identified gene DC10366 (aligned length: 1128 bp) that produced a tree with high bootstrap values in all major clades and resembled the dominant topology (Fig. 2), and a concatenated data set of 10 genes (11480 bp) that produced even better bootstrap values and topological concordance (Fig. 4).

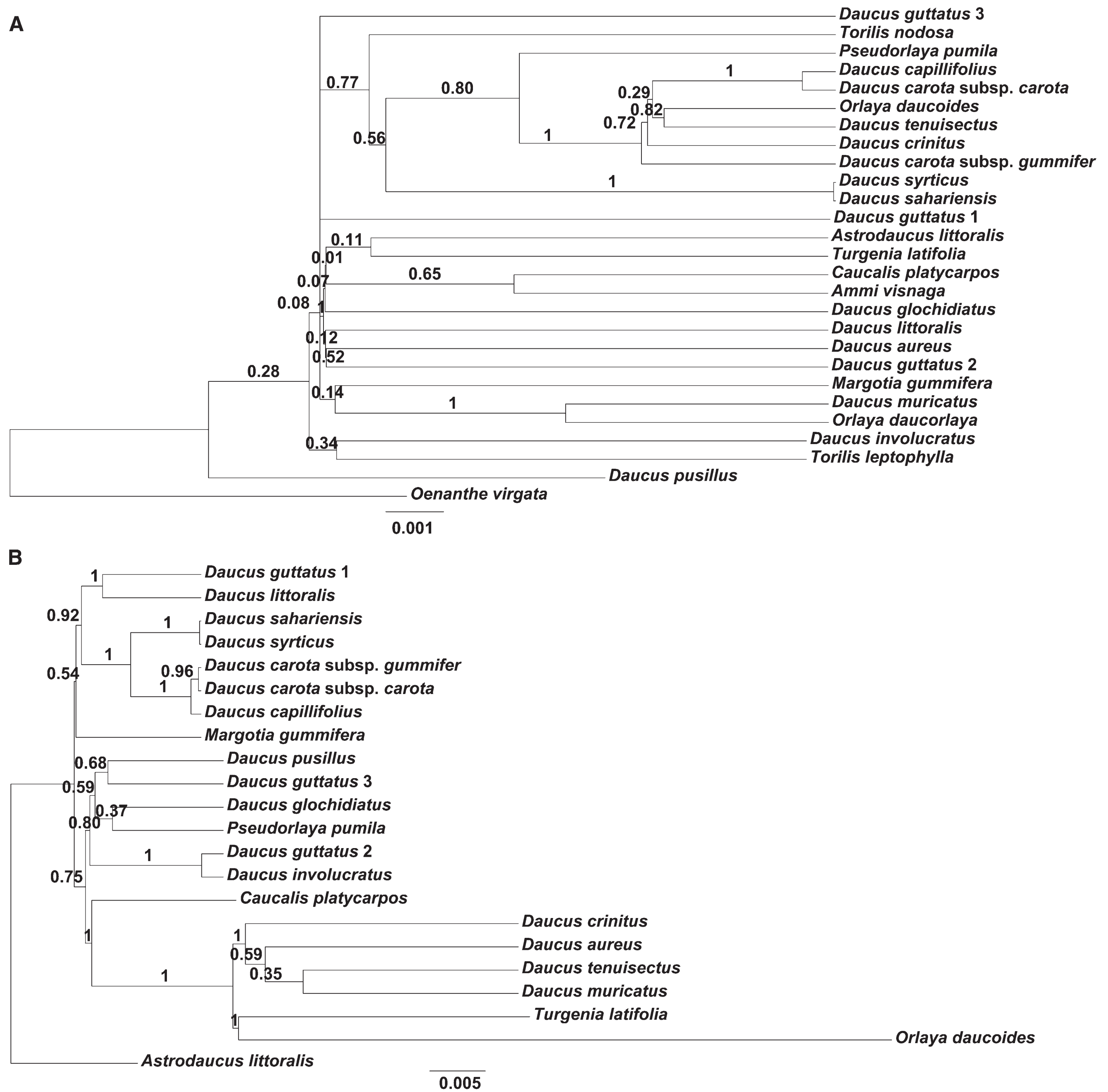

Fig. 7. (A) Species tree based on a coalescent analysis using 104 accessions of 27 species of *Daucus* and outgroups using the models of evolution obtained by jModelTest. (B) Species tree based on a coalescent analysis using the same 37 accessions of 22 species of *Daucus* and outgroups using the HKY models modified to be accepted in *BEAST with the lowest AIC value (Table 2 column 4). Number above the branches of both figures are posterior probabilities.

According to Salichos and Rokas (2013), selecting genes with high average bootstrap support reduces incongruences among many internodes. In addition, concatenation of a set of genes with bootstrap support higher than 60% can produce a species phylogeny similar to that obtained when using all genes together (Salichos and Rokas, 2013).

We decided to use a data set containing a single allele instead of a data set containing two alleles merged into one in our analyses, based on a better-resolved tree topology. The two-allele alignment with ambiguities can be criticized because if the two alleles underwent incomplete lineage sorting, they may not share the same tree; the history of each gene in that alignment may not be tree-like. The one-allele data set does not have this problem (personal communication, Cécile Ané).

We designed our study to include 10 exons to explore their taxonomic utility and to see if they would be more useful to resolve the

outgroups. As expected, the exonic regions had a smaller proportion of parsimony-informative characters as compared with the intronic regions, but the outgroups resolved the same in both data sets.

Concatenation of a large number of genes is not guaranteed to resolve phylogenetic relationships (Blair and Murphy, 2011; Blair et al., 2012). In fact, Salichos and Rokas (2013) stated that the use of bootstrap support values on concatenated analyses of large data sets should be abandoned. The concatenation method is justified when a data set has evolved under the same underlying history, in which differences in the estimated trees are due only to sampling error or model misspecification (Baum, 2007). If this is not the case for our data set, as is very likely, differences among data sets will not be due to sampling error, but to genealogical discordance. Bayesian concordance analysis (BCA) does not assume any single evolutionary history. Our concordance analyses of 94 nuclear orthologs yielded a primary concordance tree, suggesting there are significant discordant histories in *Daucus* genomes. Clades containing the subspecies of *Daucus carota* have very low concordance factors (Fig. 6). The clade containing *D. sahariensis* and *D. syrticus* has the highest concordance factor (0.69), indicating that there are minor discordant histories in *D. sahariensis* and *D. syrticus* genomes relative to *D. carota* subspecies genomes.

As indicated already, our data set had 6.1% of missing data. According to Ané et al. (2007), missing data represents a technical issue in BCA leading to mixing difficulties. However, the standard deviation concordance factor of our analysis was less than 0.005, indicating a good mixing. Discordance between genes was previously reported in different plant species such as potatoes and tomatoes (Rodríguez et al., 2009), rice (Cranston et al., 2009), animals such as salamanders (Williams et al., 2013) and lizards (Leaché, 2009), and plant pathogens such as *Phytophthora* sp. (Blair et al., 2012). The reasons for the discordance in our data set could be explained by a number of causes, from methodological explanations such as alignment bias or undetected paralogy, to biological reasons such as incomplete lineage sorting or hybridization (Wendel and Doyle, 1998). However, Philippe et al. (2011), demon strated that phylogenomics is relatively robust to the possible inclusion of nonorthologous sequences when the genuine phylogenetic signal is abundant. Therefore, the most likely factors that may be causing discordance are recombination, hybridization and introgression (Rieseberg et al., 2000), and incomplete lineage sorting (Pamilo and Nei, 1988).

The results obtained using MP, ML, and BCA are notably different from the *BEAST tree. The multispecies coalescent approach implemented in *BEAST assumes that genealogical discordance is entirely due to incomplete lineage sorting, which is considered one of the most common causes of serious difficulties for phylogenetic inference (Maddison and Knowles, 2006; Baum and Smith, 2013). However, we know that there are other processes that can cause genealogical discordance. As a result, it is better to consider an alternative approach, the BCA. This method integrates over gene tree uncertainty and does not make any particular assumption regarding the reason for discordance (Larget et al., 2010). Furthermore, BCA uses a simple measure of the prior probability of gene-to-gene discordance to convert sequence data from multiple genes into an estimate of the proportion of the genome for which any clade is true, its concordance factor (Baum and Smith, 2013). To date, there is not enough evidence to conclude the cause of genealogical discordance in the *Daucus* genome.

***Taxonomy of Daucus*—**As discussed in the introduction, molecular data place some species from nine non-*Daucus* genera in a *Daucus* clade and suggest the need to redefine the taxonomic boundaries of the genus. Lee et al. (2001) supported some species from three of these genera, *Agrocharis*, *Pachyctenium*, and *Pseudorlaya*, as nested within *Daucus*, based on a cladistic analysis of morphological data; the other six genera have yet to be examined morphologically. However, the congruence of morphological and molecular data provides strong support for a redefinition of *Daucus* to include species from these three genera, and perhaps more in the future.

The three well-supported clades of some accessions previously assigned to *D. broteri*, *D. carota*, and *D. guttatus*, and *D. littoralis* (Table 1, Fig. 2) in the dominant topology provide strong support for their recognition as three separate species. Their recognition as distinct species awaits further molecular and morphological studies of additional accessions. If such studies support distinct species status, however, additional herbarium research of type specimens is needed to assign their proper taxonomic name.

Our present molecular study and the morphological studies of Spooner et al. (2014) show the difficulty of defining subspecies of *D. carota*. In addition, these studies and the SNP analysis of Iorizzo et al. (2013) show *D. capillifolius* to be morphologically distinct, yet nested within *D. carota*. These results and the shared chromosome numbers and ease of crossability (above) suggest that *D. capillifolius* may be better recognized as a subspecies of *D. carota*, but we await our further SNP analyses of additional accessions of *D. capillifolius* and *D. carota* before we consider this taxonomic change.

In summary, relative to our three goals outlined in the introduction, (1) for concatenated data sets, MP and ML analyses of the entire *Daucus* data set of 94 nuclear orthologs produced mostly congruent trees with 100% bootstrap support for most of the external and many of the internal clades. The BCA analysis showed a similar topology to the MP and ML trees, but highlighted the fact that there were often low proportions of genes that supported certain clades. (2) The coalescent analysis is notably different from the MP, ML, and BCA trees. At present, we can only speculate on causes of discordance of our gene trees, but our database is useful for future workers wishing to explore causes of discordance in *Daucus* and other organisms. (3) The use of multiple nuclear orthologs and next-generation technologies highlighted some difficult species groups in *Daucus* and discovered misidentifications in germplasm collections. We identified a useful subset of markers and methodological approaches for future studies of dominant topologies in *Daucus*, potentially saving time and resources.

Since the initiation of our study, the Roche 454 sequencer is being phased out of service and will not be available after 2016. A repeat of our techniques could possibly use an Illumina MiSeq platform, but read lengths currently are at a maximum of 300 bp. Alternatively, the Pacific Bioscience platform could take full advantage of the entire length of the nuclear orthologs we examined (Table 2).

LITERATURE CITED

Ané, C., B. Larget, D. A. Baum, S. D. Smith, and A. Rokas. 2007. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24: 412–426.

Ayres, D., A. Darling, D. Zwickl, P. Beerli, M. T. Holder, P. O. Lewis, J. Huelsenbeck, et al. 2012. BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Systematic Biology* 61: 170–173.

Baum, D. 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* 56: 417–426.

Baum, D. A., and S. D. Smith. 2013. Tree thinking: An introduction to phylogenetic biology. Roberts and Co., Greenwood Village, Colorado, USA.

Blair, C., and R. W. Murphy. 2011. Recent trends in molecular phylogenetic analysis: Where to next? *Journal of Heredity* 102: 130–138.

Blair, J. E., M. D. Coffey, and F. N. Martin. 2012. Species tree estimation for the late blight pathogen, *Phytophthora infestans*, and close relatives. *PLoS ONE* 7: e37003.

Brandenburg, W. A. 1981. Possible relationships between wild and cultivated carrots (*Daucus carota* L.) in the Netherlands. *Die Kulturpflanze* 29: 369–375.

Cai, D., F. Rodríguez, Y. Teng, C. Ané, M. Bonierbale, L. A. Mueller, and D. M. Spooner. 2012. Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section *Petota*). *BMC Evolutionary Biology* 12: 70.

Camacho C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10: 421.

Chen, F. C., and W. H. Li. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *American Journal of Human Genetics* 68: 444–456.

Chevreux, B., T. Wetter, and S. Suhai. 1999. Genome sequence assembly using trace signals and additional sequence information computer science and biology. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)* 99: 45–56.

Christelová, P., M. Valárik, E. H ibová, E. De Langhe, and J. Doležel. 2011. A multi gene sequence-based phylogeny of the Musaceae (banana) family. *BMC Evolutionary Biology* 11: 103.

Cranston, K., B. Hurwitz, D. Ware, L. Stein, and R. A. Wing. 2009. Species trees from highly incongruent gene trees in rice. *Systematic Biology* 58: 489–500.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9: 772.

Delsuc, F., H. Brinkmann, and H. Philippe. 2005. Phylogenomics and the reconstruction of the tree of life. *Nature Reviews. Genetics* 6: 361–375.

Doolittle, W. F. 1999. Lateral genomics. *Trends in Cell Biology* 9: M5–M8.

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.

Dressman, D., H. Yan, G. Traverso, K. W. Kinzler, and B. Vogelstein. 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences, USA* 100: 8817-8822.

Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161: 1307–1320.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.

Dunn, C. W., A. Hejnol, D. Q. Matus, K. Pang, W. E. Browne, S. A. Smith, E. Seaver, et al. 2008. Broad phylogenomics sampling improves resolution of the animal tree of life. *Nature* 452: 745–749.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Edwards, V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 63: 1–19.

Egan, A., J. Schlueter, and D. M. Spooner. 2012. Applications of next-generation sequencing in plant biology. *American Journal of Botany* 99: 175–185.

Ellis, P. R., J. A. Hardman, T. C. Crowther, and P. L. Saw. 1993. Exploitation of the resistance to carrot fly in the wild carrot species *Daucus capillifolius. Annals of Applied Biology* 122: 79–91.

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.

FAO [Food and Agricultural Organization of the United Nations]. 2012. FAOSTAT, Rome, Italy. Website http://www.faostat.fao.org [accessed 4 March 2014].

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. *Plant Cell* 14: 1457–1467.

Gernhard, T. 2008. The conditioned reconstructed process. *Journal of Theoretical Biology* 253: 769–778.

Grzebelus, D., R. Baranski, K. Spalik, C. Allender, and P. W. Simon. 2011. *Daucus*. *In* C. Kole [ed.], Wild crop relatives: Genomic and breeding resources, vol. 5, Vegetables, 129–216. Springer Verlag, Berlin, Germany.

Hauser, T. P. 2002. Frost sensitivity of hybrids between wild and cultivated carrots. *Conservation Genetics* 3: 73–76.

Hauser, T. P., and G. K. Bjørn. 2001. Hybrids between wild and cultivated carrots in Danish carrot fields. *Genetic Resources and Crop Evolution* 48: 499–506.

Iorizzo, M., D. Senalik, S. Ellison, P. Cavagnaro, S. Cheng, P. Zheng, Z. Zheng, et al. 2014. The building of the first Apiaceae genome. Plant and Animal Genome XXII, San Diego, California, USA [online abstract]. Website https://pag.confex.com/pag/xxii/webprogram/Paper11512.html.

Iorizzo, M., D. Senalik, S. Ellison, D. Grzebelus, P. Cavagnaro, C. Allender, J. Brunet, et al. 2013. Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus* L.) (Apiaceae). *American Journal of Botany* 100: 930–938.

Iorizzo, M., D. A. Senalik, D. Grzebelus, M. Bowman, P. F. Cavagnaro, M. Matvienko, H. Ashrafi, et al. 2011. De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genomics* 12: 389.

Iovene, M., E. Grzebelus, D. Carputo, J. Jiang, and P. W. Simon. 2008. Major cytogenetic landmarks and karyotype analysis in *Daucus carota* and other Apiaceae. *American Journal of Botany* 95: 793–804.

Jeffreys, H. 1946. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, A, Mathematical and Physical Sciences* 186: 453–461.

Kolaczkowski, B., and J. W. Thornton. 2004. Performance of maximum parsimony and maximum likelihood phylogenetics when evolution is heterogeneous. *Nature* 431: 980–984.

Koonin, E. V. 2005. Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics* 39: 309–338.

Koren, S., M. C. Schatz, B. P. Walenz, J. Martin, J. T. Howard, G. Ganapathy, Z. Wang, et al. 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature Biotechnology* 30: 693–700.

Krickl, M. 1961. Karotten: zur Frage der Verkreuzung mit der wilden Karotte. *Saatgut-Wirtschaft* 13: 135–136.

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. 2004. Versatile and open software for comparing large genomes. *Genome Biology* 5: R12.

Lang, J. M., A. E. Darling, and J. A. Eisen. 2013. Phylogeny of bacterial and archaeal genomes using conserved genes: Supertrees and supermatrices. *PLoS ONE* 8: e62510.

Larget, B., S. K. Kotha, C. N. Dewey, and C. Ané. 2010. BUCKy: Gene tree/species tree reconciliation with the Bayesian concordance analysis. *Bioinformatics* 26: 2910–2911.

Leaché, A. D. 2009. Species tree discordance traces to phylogeographic clade boundaries in North American fence lizards (*Sceloporus*). *Systematic Biology* 58: 547–559.

LEE, B. Y., AND S. R. DOWNIE. 2000. Phylogenetic analysis of cpDNA restriction sites and *rps*l6 intron sequences reveals relationships among Apiaceae tribes Caucalideae, Scandiceae and related taxa. *Plant Systematics and Evolution* 221: 35–60.

LEE, B. Y., G. A. LEVIN, AND S. R. DOWNIE. 2001. Relationships within the spiny-fruited umbellifers (Scandiceae subtribes Daucinae and Torilidinae) as assessed by phylogenetic analysis of morphological characters. *Systematic Botany* 26: 622–642.

LEVIN, R. A., A. WHELAN, AND J. S. MILLER. 2009. The utility of nuclear conserved ortholog set II (COSII) genomic regions for species-level phylogenetic inference in *Lycium* (Solanaceae). *Molecular Phylogenetics and Evolution* 53: 881–890.

LEWIS, P. O., M. T. HOLDER, AND K. E. HOLSINGER. 2005. Polytomies and Bayesian phylogenetic inference. *Systematic Biology* 54: 241–253.

LI, M., J. WUNDER, G. BISSOLI, E. SCARPONI, S. GAZZANI, E. BARBARO, H. SAEDLER, AND C. VAROTTO. 2008. Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. *Cladistics* 24: 727–745.

MADDISON, D. R., AND W. P. MADDISON. 2005. MacClade 4.08a: Analysis of phylogeny and character evolution. Sinauer, Sunderland, Massachusetts, USA.

MADDISON, W. P., AND L. L. KNOWLES. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* 55: 21–30.

MARGULIES, M., M. EGHOLM, W. ALTMAN, S. ATTIYA, J. BADER, L. BEMBEN, J. BERKA, ET AL. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.

MCCOLLUM, G. D. 1975. Interspecific hybrid *Daucus carota* × *D. capillifolius. Botanical Gazette (Chicago, Ill.)* 136: 201–206.

MCCOLLUM, G. D. 1977. Hybrids of *Daucus gingidium* with cultivated carrots (*D. carota* subsp. *sativus*) and *D. capillifolius. Botanical Gazette* 138: 56–63.

MCCORMACK, J. E., S. HIRD, A. ZELLMER, B. CARSTENS, AND R. BRUMFIELD. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66: 526–538.

MILLER, M. A., W. PFEIFFER, AND T. SCHWARTZ. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE) in New Orleans, 2010, 1–8, USA.

MOSSEL, E., AND E. VIGODA. 2005. Phylogenetic MCMC algorithms are misleading on mixtures of trees. *Science* 309: 2207–2209.

NOTHNAGEL, T., P. STRAKA, AND B. LINKE. 2000. Male sterility in populations of *Daucus* and the development of alloplasmic male-sterile lines of carrot. *Plant Breeding* 119: 145–152.

OWCZARZY, R., A. V. TATAUROV, Y. WU, J. A. MANTHEY, K. A. MCQUISTEN, H. G. ALMABRAZI, K. F. PEDERSEN, ET AL. 2008. IDT SciTools: A suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research* 36: W163–W169.

PACIFIC BIOSCIENCES. 2013. Pacific Biosciences introduces new chemistry with longer read lengths to detect novel features in DNA sequence and advance genome studies of large organisms. Website http://investor.pacificbiosciences.com/releasedetail.cfm?ReleaseID=794692 [accessed 9 March 2014].

PAGE, R. D. M., AND M. A. CHARLESTON. 1997. From gene to organismal phylogeny: Reconciled trees and the gene tree species tree problem. *Molecular Phylogenetics and Evolution* 7: 231–240.

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

PHILIPPE, H., H. BRINKMANN, D. V. LAVROV, D. T. J. LITTLEWOOD, M. MANUEL, G. WORHEIDE, AND D. BAURAIN. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology* 9: e1000602.

PHILIPPE, H., E. A. SNELL, E. BAPTESTE, P. LOPEZ, P. W. H. HOLLAND, AND D. CASANE. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Molecular Biology and Evolution* 21: 1740–1752.

PIMENOV, M. G., AND M. V. LEONOV. 1993. The genera of Umbelliferae: A nomenclator. Royal Botanic Gardens, Kew, UK.

PLUNKETT, G. M., D. E. SOLTIS, AND P. S. SOLTIS. 1996. Evolutionary patterns in Apiaceae: Inferences based on *mat*K sequence data. *Systematic Botany* 21: 477–495.

RIESEBERG, L. H., S. J. BAIRD, AND K. A. GARDNER. 2000. Hybridization, introgression, and linkage evolution. *Plant Molecular Biology* 42: 205–224.

ROCHE. 2009. em-PCR method manual—Lib-A SV. Website http://www.high-throughput-sequencing.com/manuals_roche/emPCR_liba_sv_Oct2009.pdf [accessed 18 July 2014].

ROCHE. 2010. Rapid library preparation method manual. Website http://dna.uga.edu/wp-content/uploads/2013/12/GS-FLX-Titanium-RapidLibrary-RL-Preparation-Method-Manual-Roche.pdf [accessed 18 July 2014].

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.

ROKAS, A., AND S. B. CARROLL. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* 22: 1337–1344.

ROKAS, A., AND S. B. CARROLL. 2006. Bushes in the tree of life. *PLoS Biology* 4: e352.

ROKAS, A., D. KRUGER, AND S. CARROLL. 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science* 310: 1933–1938.

ROKAS, A., B. L. WILLIAMS, N. KING, AND S. B. CARROLL. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804.

RONQUIST, F., M. TESLENKO, P. VAN DER MARK, D. L. AYRES, A. DARLING, S. HÖHNA, B. LARGET, ET AL. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539–542.

ROTHBERG, J. M., AND J. H. LEAMON. 2008. The development and impact of 454 sequencing. *Nature Biotechnology* 26: 1117–1124.

RUBATZKY, V. E., C. F. QUIROS, AND P. W. SIMON. 1999. Carrots and related vegetable Umbelliferae. CABI, New York, New York, USA.

SAENZ DE RIVAS, C., AND V. H. HEYWOOD. 1974. Preliminary study on the *Daucus* species of peninsular Spain). *Anales del Jardin Botánico A. J. Canavilles* 31: 97–118.

SÁENZ LAÍN, C. 1981. Research on *Daucus* L. (Umbelliferae). *Anales del Jardin Botánico de Madrid* 37: 481–534.

SALICHOS, L., AND A. ROKAS. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497: 327–333.

SANGER, F., S. NICKLEN, AND A. R. COULSON. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* 74: 5463–5467.

SCHIERWATER, B., M. EITEL, W. JAKOB, H. J. OSIGUS, H. HADRYS, S. L. DELLAPORTA, S. O. KOLOKOTRONIS, AND R. DESALLE. 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biology* 7: e1000020.

SHENDURE, J., AND H. JI. 2008. Next-generation DNA sequencing. *Nature Biotechnology* 26: 1135–1145.

SIMON, P., L. POLLAK, B. CLEVIDENCE, J. HOLDEN, AND D. HAYTOWITZ. 2009. Plant breeding for human nutrition. *Plant Breeding Reviews* 31: 325–392.

SOLTIS, D. E., M. A. GITZENDANNER, G. STULL, M. CHESTER, A. CHANDERBALI, S. CHAMALA, I. JORDON-THADEN, ET AL. 2013. The potential of genomics in plant systematics. *Taxon* 62: 886–898.

SPALIK, K., AND S. R. DOWNIE. 2007. Intercontinental disjunctions in *Cryptotaenia* (Apiaceae, Oenantheae): An appraisal using molecular data. *Journal of Biogeography* 34: 2039–2054.

SPALIK, K., M. PIWCZYŃSKI, C. A. DANDERSON, R. KURZYNA-MŁYNIK, T. S. BONE, AND S. R. DOWNIE. 2010. Amphitropic amphiantarctic disjunctions in Apiaceae subfamily Apioideae. *Journal of Biogeography* 37: 1977–1994.

SPOONER, D. M., P. ROJAS, M. BONIERBALE, L. A. MUELLER, M. SRIVASTAV, D. SENALIK, AND P. SIMON. 2013. Molecular phylogeny of *Daucus. Systematic Botany* 38: 850–857.

SPOONER, D. M., M. P. WIDRLECHNER, K. R. REITSMA, D. E. PALMQUIST, S. ROUZ, Z. GHRABI-GAMMAR, M. NEFFATI, ET AL. 2014. Reassessment of practical subspecies identifications of the USDA *Daucus carota* germplasm collection: Morphological data. *Crop Science* 54: 706–718.

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.

STAMATAKIS, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

STEINBORN, R., B. LINKE, T. NOTHNAGEL, AND T. BÖRNER. 1995. Inheritance of chloroplast and mitochondrial DNA in alloplasmic forms of the genus *Daucus. Theoretical and Applied Genetics* 91: 632–638.

SWOFFORD, D. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0a131. Sinauer, Sunderland, Massachusetts.

TAKAHATA, N. 1989. Gene genealogy in three related populations: Consistency probability between gene and population trees. *Genetics* 122: 957–966.

UMIEL, N., R. JACOBSON, AND D. GLOBERSON. 1975. Pollination of the cultivated carrot (*Daucus carota* L.) by the wild carrot (*D. carota* var. *maximus*), and its implications on commercial seed production. *Hassadeh* 56: 478–480 [In Hebrew, English summary].

USDA [U. S. DEPARTMENT OF AGRICULTURE], NATIONAL AGRICULTURAL STATISTICS SERVICE. 2012. Data and statistics. Website http://www.nass.usda.gov/Data_and_Statistics/index.asp [accessed 4 March 2014].

VIVEK, B. S., AND P. SIMON. 1999. Phylogeny and relationships in *Daucus* based on restriction fragment length polymorphisms (RFLPs) of the chloroplast and mitochondrial genomes. *Euphytica* 105: 183–189.

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.

WIJNHEIJMER, E. H. M., W. A. BRANDENBURG, AND S. J. TER BORG. 1989. Interactions between wild and cultivated carrots (*Daucus carota* L.). *Euphytica* 40: 147–154.

WILLIAMS, J. S., J. H. NIEDZWIECKI, AND D. W. WEISROCK. 2013. Species tree reconstruction of a poorly resolved clade of salamanders (Ambystomatidae) using multiple nuclear loci. *Molecular Phylogenetics and Evolution* 68: 671–682.

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.

ZWICKL, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, University of Texas at Austin, Austin, Texas, USA.