American Journal of Botany
Celebrating 100 years 1914–2014

American Journal of Botany 101(1): 86–91. 2014.

Diversification times among *Brassica* (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence[1]

Tatiana Arias[2,5], Mark A. Beilstein[3], Michelle Tang[2], Michael R. McKain[4], and J. Chris Pires[2]

[2]Division of Biological Sciences, University of Missouri, 311 Bond Life Sciences Center, 1201 Rollins Street, Columbia, Missouri 65211 USA; [3]School of Plant Sciences, The University of Arizona, P.O. Box 210036, Forbes Building, Room 303, Tucson, Arizona 85721 USA; and [4]Department of Biology, University of Missouri-St. Louis, One University Blvd., St. Louis, Missouri 63121 USA

- *Premise of the study:* Cruciferous vegetables, many of which are in the genus *Brassica* (Brassicaceae), are prized for their nutritive value and have been cultivated for thousands of years. There are numerous wild northwestern Mediterranean species in the tribe Brassiceae, and it is therefore assumed this center of diversity is also the region of origin. Within the tribe, the Nigra and Oleracea clades contain the three diploid *Brassica* crops, *B. oleracea*, *B. rapa*, and *B. nigra.* These three species hybridized in the past to form the tetraploid crop species *B. juncea*, *B. carinata*, and *B. napus*. Collectively, these crop Brassicas have been thought to be closely related because they can still hybridize.
- *Methods:* Using a combination of molecular phylogenetics, diversification analysis, and historical biogeography, we evaluated the relationships and origins of four nested clades: the tribe Brassiceae, the Nigra-Oleracea clade, the core Oleracea (includes *B. oleracea* + *B. rapa* and their respective wild relatives), and *Brassica oleracea* and relatives.
- *Key results:* We found evidence that the tribe originated around the intersection forming between the Arabian Peninsula and Saharan Africa approximately 24 million years ago (Mya). Our data also suggest that the maternal genomes of the three diploid crop Brassicas are not closely related and that the Nigra-Oleracea clade diverged 20 Mya. Finally, our analyses indicate that the core Oleracea lineage giving rise to *B. oleracea* + *B. rapa* originated ≈3 Mya in the northeastern Mediterranean, from where ancestors of *B. oleracea* spread through Europe and *B. rapa* to Asia.
- *Conclusions:* These results challenge previous hypotheses about the biogeographic origins of the tribe Brassiceae and the crop *Brassica* species and appear to be correlated with major geological and climatic events in the Mediterranean basin.

Key words: *Brassica*; Brassicaceae; crops; Mediterranean; Miocene.

Cruciferous vegetables (e.g., Chinese cabbage, broccoli, and mustard) are a major food crop contributing to the diet of millions of people and are of significant importance for agricultural economies worldwide. They have been independently domesticated for consumption, industrial products, and medicine in Europe, the Middle East, and Asia (Hedge, 1976; Gómez-Campo, 1980; Al-Shehbaz, 1985; Gómez-Campo and Prakash, 1999). Wild and domesticated *Brassica* species (here after tribe Brassiceae) exhibit not only spectacular morphological variation (Snogerup, 1980; Hodgkin, 1995; Hall et al., 2011) and physiological adaptations (pest resistance, heavy metal, drought, and salinity tolerance among others) (Fahey et al., 2001), but they possess unparalleled levels of genomic diversity (Lukens et al., 2004; Lysak et al., 2005; Schranz et al., 2006; Lysak et al., 2007). Genomes of all Brassiceae share an ancient hexaploidization event that occurred before the group radiated (Lukens et al., 2004; Lysak et al., 2005; Parkin et al., 2005) ≈28.3 million years ago (Mya) to 15.6 Mya (Beilstein et al., 2010). Additionally, some tribe members underwent multiple hybridization and whole genome duplication events that likely facilitated species diversification within the group and the origin of the *Brassica* crops (Lysak et al., 2005; Parkin et al., 2005).

The geographic origins and domestication of many crops, including cereals (Ge et al., 1999; Salamini et al., 2002; Buckler and Stevens, 2006), legumes (Carter et al., 2004), nightshade vegetables (Jenkins, 1948; Hawkes, 1988), and gourds (Smith, 1997; Sanjur et al., 2002; Schaefer et al., 2009), have been resolved using a combination of dated molecular phylogenies and historical biogeography. However, dated biogeographical reconstructions of *Brassica* crops and wild relatives have remained beyond reach due primarily to the lack of a robust molecular phylogeny. The Mediterranean region and adjacent areas are particularly problematic in regard to phylogenetic, and thus biogeographic inference. For example, both diploid and tetraploid *Brassica* crops have evolved since the intersection of the three continents that form the region, and human agricultural civilizations have been present for thousands of years in the area. Both of these factors are expected to potentially confound phylogenetic signal due to the formation of new hybridization and introgression zones as the region formed and as domesticated

[1]Manuscript received 25 August 2013; revision accepted 6 November 2013.

The authors thank P. McSteen, G. Conant, J. Birchler, M. Liscum, C. Henriquez, J. C. Hall, M. Lundberg, A. Cardenas, and two anonymous reviewers for valuable comments on the manuscript. The authors acknowledge the following research grants and granting institutions: National Science Foundation (DEB 1146603, DEB 1209137).

[5]Author for correspondence (e-mail: tatiana.tatianaarias@gmail.com); present address: Division of Ecology & Biodiversity, School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, China; phone: (852) 98-16-7273

doi:10.3732/ajb.1300312

crops spread through agricultural regions. Hence, the recent publication of a fully resolved chloroplast phylogeny (Arias and Pires, 2012) marks a major advance in our understanding of relationships among *Brassica* species and provides the necessary initial framework to test evolutionary hypotheses in the group.

Molecular phylogenies have identified eight named clades within the tribe Brassiceae: Vella, Zilla, Henophyton, Crambe, Cakile, Savignya, Nigra, and Oleracea (Arias and Pires, 2012). Vella and Zilla are early-diverging groups sister to all other tribe members; Henophyton is sister to core Brassiceae containing Crambe, Cakile, Savignya, and two species-rich clades identified as Nigra and Oleraceae (Warwick and Black, 1991). The Nigra and Oleracea clades contain the three diploid and three tetraploid *Brassica* crops collectively termed the triangle of U species (U, 1935). The Nigra lineage includes *Brassica nigra* ($n = 8$; black mustard), found in North Africa and extending to the Middle East. The Oleracea lineage includes the diploid sister species *B. oleracea* ($n = 9$; cabbage, broccoli, Brussels sprouts, cauliflower, kale) common across Europe, and *B. rapa* ($n = 10$; pak-choi, Chinese cabbage, turnip) occurring naturally from Eastern Europe to Asia. Diploid triangle of U species are thought to have diverged from a recent common ancestor due to their ability to interbreed. Hybridization among diploid *Brassica* crops also suggests little genetic divergence and incomplete reproductive isolation (Coyne and Orr, 1989; Foltz, 1997) among these species. Crosses among the diploids generate three stable tetraploid *Brassica* species: *B. juncea* ($n = 18$, Indian mustard) distributed along the contact zone between *B. nigra* and *B. rapa* in the Middle East; *B. napus* ($n = 19$; rapeseed and rutabega) naturally found in northern Europe, the contact zone between *B. rapa* and *B. oleracea*; and *B. carinata* ($n = 17$; Ethiopian mustard) native to Ethiopia and Kenya. Previous studies of diversification dates based on nuclear and chloroplast markers estimate that *Brassica rapa* + *B. oleracea* diverged from *B. nigra* 7.9 Mya (Lysak et al., 2005), while the genomes of *B. rapa* and *B. oleracea* diverged from each other 3.7 Mya (Inaba and Nishio, 2002) or 1.37–0.12 Mya (Cheung et al., 2009).

Here we re-examine geographical origins and the timing of diversification of four nested clades: the tribe Brassiceae, the Nigra-Oleracea clade, the core Oleracea (includes *B. oleracea* + *B. rapa* and their respective wild relatives), and *Brassica oleracea* and relatives. We added nine outgroups to a recently published, well-resolved, four-locus plastid phylogeny of Brassiceae (Arias and Pires, 2012) and used advances in our understanding of divergence dates in Brassicales (Beilstein et al., 2010). We have included representatives of most genera within the Brassiceae crown group (≈47 genera and 227 species) (Al-Shehbaz, 2012), covering their entire natural range from Central Asia to the western Mediterranean (Hedge, 1976; Gómez-Campo, 1980; Al-Shehbaz, 1985; Gómez-Campo and Prakash, 1999; Thompson, 2005). Using historical biogeographic reconstructions, we evaluate three primary hypotheses about the origin of the tribe. In the first scenario, ancestors of the tribe arose in the northwestern Mediterranean, the current center of tribal diversity (Gómez-Campo, 1980). The second hypothesis is an origin in the Irano-Turanian region (including Turkey and Iran), which corresponds with the center of diversity and proposed region of origin for the family Brassicaceae (Hedge, 1976; Al-Shehbaz, 1985; Franzke et el., 2009). A third hypothesis is that the ancestors of the tribe originated in the Saharo-Sindian region (Sahara, Egypt, and Arabia), similar to ancestors of several plant and animals lineages that later diversified in the Mediterranean (Zardoya and Doadrio, 1999; Sanmartín, 2003; Lo Presti and Oberprieler, 2009; Mico et al., 2009). These alternatives have implications for the pattern of colonization and expansion of the tribe throughout the Mediterranean basin. Similar biogeographic and dating analyses were used to clarify the biogeographical origin of the Nigra-Oleracea clade and the *Brassica* crops within the tribe. Our results challenge currently accepted ideas about the origins and diversification of the tribe Brassiceae and the crop brassicas.

MATERIALS AND METHODS

***New sequences and alignments*—**Four phylogenetically informative and variable chloroplast markers (*atpI-atpH, psbD-trnT, rpl32-trnL, ycf6-psbM*) were previously used across members of tribe Brassiceae to infer trees using maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) (Arias and Pires, 2012). Here sequences for nine Brassicaceae outgroup species were added to the original phylogenetic analysis (Arias and Pires, 2012) to validate the diversification analysis with recently published studies (Beilstein et al., 2010). To obtain the aforementioned chloroplast intergenic regions for the nine Brassicaceae outgroup species, we took several approaches: (1) for *Caulanthus amplexicaulis*, *Conringia planisiliqua*, *Isatis canescens*, and *Streptanthus drepanoides* regular Sanger sequencing was used. (2) For *Arabidopsis thaliana* and *Capsella bursa-pastoris*, regions were extracted from the chloroplast genomes available in GenBank (NC_000932 and NC_009270). (3) For *Schrenkiella parvula* (Schrenk), raw transcriptome data were obtained from GenBank (SRX047632), and chloroplast regions were recovered as described below. (4) For *Thlaspi arvense* and *Eutrema salsugineum*, regions were obtained from total genomic DNA that were then sequenced on an Illumina GAIIx Genome Analyzer (see Steele et al., 2012 for sequencing details). Raw reads from transcriptomes (*Schrenkiella parvulum*) and genomic DNA (*Thlaspi arvense* and *Eutrema salsugineum*) were used to recover chloroplast regions in (3) and (4). Reads were examined using the program FastQC v. 0.10.1 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) before and after cleaning data. Reads with Phred quality scores lower than 20 were filtered using ea-tools v. 1.1.2 from Google Code (Aronesty, 2011). FASTX-Toolkit v. 0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/) was used to trim the beginning and end of sequences where there were biases in bases content. The filtered and clean sequences were assembled against the chloroplast genome of *Brassica rapa* obtained from GenBank (NC_015139) using the program Geneious v. 5.5.8 (Drummond et al., 2011; mean coverage of 1000×). Subsequently, the four chloroplast regions used in the original phylogeny (Arias and Pires, 2012) were extracted, concatenated, and aligned with the original data set using the program MUSCLE v. 3.8.31 (Edgar, 2004). A hundred nucleotides from a hypervariable region of *rpl32-trnL* were excluded (see Arias and Pires, 2012). The final alignment was deposited in the database TreeBase (http://treebase.org, 14830) and comprised 3250 nucleotide positions (*atpI-atpH*: 1–514, *psbD-trnT*: 515–1544, *rpl32-trnL*: 1545–2720, *ycf6-psbM*: 2721–3250), representing 101 accessions from 91 species.

***Diversification times*—**Bayesian approaches in the program BEAST v. 1.7.5 (Drummond and Rambaut, 2007) were used to estimate divergence times, topology, and branch lengths, for the combined four-marker chloroplast alignment. Phylogenies were calibrated using a fossil and secondary calibration points. The fossil *Thlaspi primaevum* (Wing, 1987), dated at 30.8–29.2 Mya (Beilstein et al., 2010), was used as a calibration point for the split between *Alliaria petiolata* and *Thlaspi arvense* (Fig. 1; position 1). The secondary calibration point of 43.2 Mya (95% highest posterior density [HPD]: 50.7–36.6 Mya) was used toward the root of the tree at the split between LINI from LINII (Fig. 1; position 2), leaving *Capsella bursa-pastoris* and *Arabidopsis thaliana* as outgroups (Beilstein et al., 2010).

The BEAST analyses accounted for rate variation using an uncorrelated relaxed clock drawn from a lognormal distribution that allows different rates to be optimized independently on each branch of the tree (Drummond and Rambaut, 2007). The Yule process (Yule, 1924) described the likelihood of speciation for this analysis, and branching rates were determined under a GTR model of nucleotide evolution with four categories (Drummond and Rambaut, 2007). In the analysis of combined data (four chloroplast markers and 3250 nucleotide positions), the substitution model was unlinked from the clock model, allowing each data partition (*atpI-atpH*: 1–514, *psbD-trnT*: 515–1544, *rpl32-trnL*: 1545–2720,

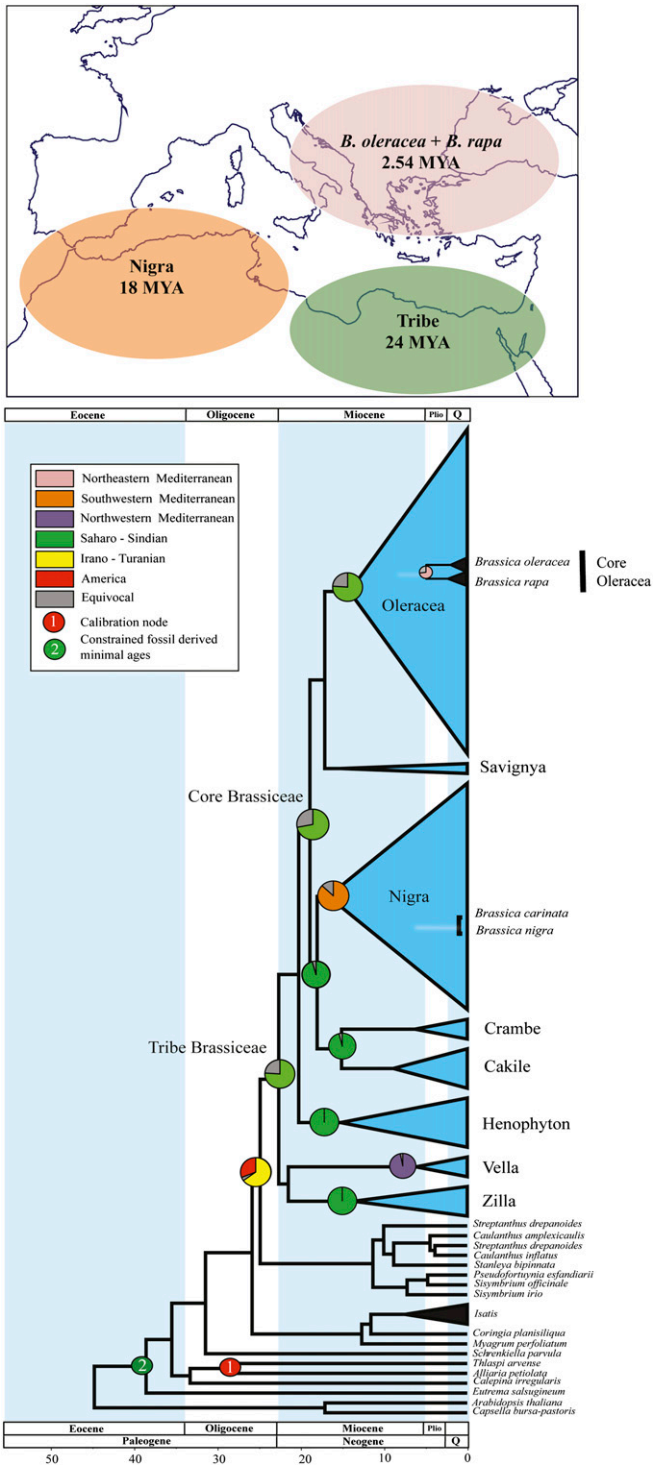

Fig. 1. Brassiceae chronogram inferred using the program BEAST. Main clades are represented by wedges proportional to species diversity. Colors in the map match geographical distributions in the figure key to symbols. Pie charts represent percentage of trees with uniquely best state from a tree distribution of 10 000 trees. Plio, Pliocene; Q, Quaternary.

ycf6-psbM: 2721–3250) to evolve independently across the same tree. Different types of priors were implemented for the two calibration points used to date divergences. A normal distribution was used for the secondary calibration point defining the age of the split between LINI and LINII (43.2; 95% HPD: 50.7–36.6 Mya; standard deviation: 9.97) because secondary calibration dates require the use of soft minimum and maximum bounds (Ho and Phillips, 2009). In contrast, the prior on the age of the split between *T. arvense* and *A. petiolata* was described by a uniform distribution with the lower bound set to ≈29.2 Ma and the upper bound at ≈30.8 Ma (early Oligocene).

Four independent runs in BEAST were conducted for 5×10^7 generations sampling every 1000 generations for the combined four-maker chloroplast analysis. Convergence statistics for each run were analyzed in the program Tracer v. 1.5 (Rambaut and Drummond, 2003), using the resulting 45 000 post-burn-in generations. Chain convergence was obtained when all statistics had effective sample size (ESS) of >200 in the combined file for the four runs in Tracer and also by visually examining the data distribution for each run and the combined file. The program LogCombiner v. 1.7.5 (Drummond and Rambaut, 2007) was used to combine log and trees files. TreeAnnotator v. 1.7.5 (Drummond and Rambaut, 2007) was used to produce maximum clade credibility trees from the postburn-in trees and to determine the 95% posterior density of ages for all nodes in the tree.

***Historical biogeography*—**Analysis of ancestral distribution was conducted using a parsimony approach in Mesquite v. 2.75 (Maddison and Maddison, 2007). BEAST v. 1.7.5 (Drummond and Rambaut, 2007) was implemented to infer topologies with branch lengths for the combined alignment of four chloroplast regions (see section above for details about the analysis performed to recover these phylogenies). After a 45 000 postburn-in, 10 000 trees were used. Ancestral state reconstructions were calculated from counting the number of trees with the uniquely best state. Results were summarized in a majority rule consensus tree resulting from the BEAST analysis.

To maximize congruence with other biogeographic studies in the region, the operational areas from Mico et al. (2009) were used, except the Mediterranean region was further subdivided resulting in these areas: (A) Euro Siberian, (B) northwestern Mediterranean, (C) northeastern Mediterranean, (D) southwestern Mediterranean, (E) Irano-Turanian, (F) Saharo-Sindian, (G) Atlantic, (H) America and (I) East Africa/Red Sea. Widespread species were coded as present in multiple regions, and only the native distributions were taken into account. Distributional data were compiled from Warwick (http://www.brassica.info) and topographic maps were taken from the National Geophysical Data Center (www.ngdc.noaa.gov).

RESULTS

***Diversification times*—**New sequences generated in this study were deposited in GenBank (Appendix S1, see Supplemental Data with the online version of this article). Using Bayesian approaches, we found the tribe Brassiceae originated around 24.02 Mya [95% HPD: 31.55–18.46] (Fig. 1; online Appendix S2). The eight named major clades (Arias and Pires, 2012) and the majority of species in the Brassiceae diversified during the Miocene (≈20 to 10 Mya) (Fig. 1; Appendix S2). Appendices S2 and S3 provide mean divergence times for nodes within the Brassiceae and credibility intervals (see Supplemental Data with the online version of this article).

Our results indicate that the deep coalescence of the two main lineages in the tribe Nigra and Oleracea (core Brassiceae) (sensu Arias and Pires, 2012) occurred 20.04 Mya [95% HPD: 27.38–15.45]. Diversification within the Nigra lineage including *Brassica nigra* and its wild relatives is dated to 17.72 Ma [95% HPD: 27.69–13.29]. The Oleracea crown group diversified beginning approximately 15 Mya [95% HPD: 19.88–11.15] (Fig. 1; Appendix S2).

As reported previously (Arias and Pires, 2012), molecular phylogenies for the tribe also identified a subclade within Oleracea know as core Oleracea containing *Brassica oleracea*, *B. rapa* and their respective wild relatives (Fig. 1; Appendix S2). This clade originated 6.47 Mya [95% HPD: 9.25–4.17] late in the Miocene.

Finally, *B. oleracea* varieties and their wild relatives diversified 1.44 Mya [95% HPD: 2.45–0.66] (Appendix S2).

***Historical biogeography*—**Using a combination of parsimony and Bayesian approaches, we found that the tribe Brassiceae emerged from the forming Saharo-Sindian region (77% of 1×10^4 trees), which today is a broad desert belt stretching from northwest Africa to Sind (India) (Blondel et al., 2010) (Fig. 1). Our results indicate that the biogeographic origin of the two main lineages in the tribe, Nigra and Oleracea (core Brassiceae) (sensu Arias and Pires, 2012), occurred in the Saharo-Sindian region [6% of 1×10^4 trees]) (Fig. 1). Online Appendix S4 provides ancestral ranges of distribution with pie charts representing the marginal probabilities of ancestral areas reconstructed at each node within the Brassiceae.

The origin of the Nigra lineage excluding *Coincya longirostra* but including *Brassica nigra* and its wild relatives was centered in the southwestern Mediterranean region (85% of 1×10^4 trees) (Fig. 1; Appendix S4). The Oleracea crown group originated in the Saharo-Sindian region (73% of 1×10^4 trees) and later colonized the European Mediterranean through the east (Fig. 1; Appendix S4).

The core Oleracea (*B. oleracea* and *B. rapa* varieties and their wild relatives) diversified in the northeastern Mediterranean (72% of 1×10^4 trees) and later spread through the rest of Europe (Fig. 1; Appendix S4). Finally, *B. oleracea* varieties and their wild relatives diversified in the northeastern Mediterranean (72% of 1×10^4 trees) and later spread through the rest of Europe (Fig. 1; Appendix S4).

DISCUSSION

***The tribe Brassiceae evolved in the forming intersection of the Irano-Turanian and Saharo-Sindian regions late in the Oligocene*—**Our analyses indicate that the center of origin for the tribe Brassiceae is not its center of diversity, rejecting the Mediterranean origin hypothesis (Gómez-Campo, 1980). We found that the tribe Brassiceae emerged around 24.02 Mya from the Saharo-Sindian region, today a desert belt stretching from northwest Africa to India. Our results are congruent with biogeographic scenarios laid out for other organisms with high species diversity in the Mediterranean (Zardoya and Doadrio, 1999; Sanmartín, 2003; Lo Presti and Oberprieler, 2009; Mico et al., 2009) and reveal the oldest estimates of divergence among the *Brassica* crops and wild relatives. Recent phylogenetic analyses recovered Henophyton (North African clade) as sister to all other members of core Brassiceae (Arias and Pires, 2012). The phylogenetic position of Henophyton influences historical biogeographic inference of the ancestral distribution for Brassiceae. The Irano-Turanian region has been proposed as the center of diversity and origin of Brassicaceae (Hedge, 1976; Al-Shehbaz, 1985; Coyne and Orr, 1989; Koch et al., 2012). This hypothesis would agree with the paleographic scenario suggested here for the tribe ancestors: they would have inhabited the intersection between the Irano-Turanian and Saharo-Sindian regions late in the Oligocene and then expanded and colonized North Africa while the Mediterranean was still forming.

Major geological and climatic events in the Mediterranean basin appear to be correlated with the origin and diversification of the tribe Brassiceae. The eight major clades (Arias and Pires, 2012) diversified during the Miocene when the Anatolian plate collided with Africa and Arabia and the Mediterranean was cut off from the Indian Ocean, forming a landbridge between continents (Rogl, 1999). Our results agree with recently dated phylogenies for the Brassicaceae (Beilstein et al., 2010), likely because we have used the same methods. Our dates differ from the diversification times of Lysak et al. (2005), who suggested the tribal origin at 14.6–7.9 Mya, estimated using mutation rate of synonymous sites and equal rates of evolution across the tree.

Crambe and Vella diversified ≈7 Mya [(95% HPD: 10.6–3.5, 10.5–3.4 respectively), coincident with the emergence of core Oleracea (95% HPD: 9.3–4.2) (Bocquet et al., 1978, Crespo et al., 2000) (Fig. 1). All three groups overlap the Messian salinity crisis (MSC), which is dated 5.96–5.33 Mya. The MSC occurred when the Mediterranean Sea became enclosed by the collision of Africa and Eurasia, causing a drawdown of the sea level due to widespread salt precipitation, eventually causing the largest rivers to become nearly desiccated (Thompson, 2005; Blondel et al., 2010; Garcia-Castellanos and Villaseñor, 2011). This event could have promoted lineage diversification either via adaptive radiation or through geographically induced allopatric events, as has been suggested for other plant groups (Antonelli and San Martin, 2011). With an increase in aridity and transcontinental connections appearing, new migration routes may have been introduce, allowing groups to colonize the previously submerged steppes and mountain chains thus providing the impetus for diversification (Crespo et al., 2000; Thompson, 2005).

Species diversification in the Mediterranean has usually been attributed to geographical barriers and climatic change at the end of the Tertiary and during the Quaternary Period. Geographical isolation in glacial refugia has been proposed for creating the context that would allow allopatric speciation (Thompson, 2005, p. 43). However, in the Brassiceae most of the species within the Oleracea and Nigra clade appeared during the Plioceone–Pleistocene boundary (≈2.6 Mya), suggesting that although they were eventually affected by periodic glaciations, much of the diversity in these clades was already established.

***Crop brassicas diverged 20 Mya and then hybridized in zones of secondary contact*—**Given that *Brassica nigra* in the Nigra lineage readily hybridizes with *B. rapa* and *B. oleracea* in the Oleracea lineage, it has been thought that the diploid crop brassicas are closely related. The average genetic distances between *B. rapa* and *B. oleracea* have been observed to be lower than between *B. rapa* and *B. nigra* (Song et al., 1995). The probability of hybrid formation should decrease as reproductive barriers (both prezygotic and postzygotic) become stronger between parents with increased genomic divergence (Paun et al., 2009). Additionally, time since divergence is often positively correlated with the degree of reproductive isolation between species (Coyne and Orr, 1989; Foltz, 1997). However, this does not appear to be the case for the diploid triangle of U species; after 20 Myr of separation, *B. rapa* and *B. oleracea* can still hybridize with *B. nigra*. We found that the Nigra and Oleracea lineages diversified 20.04 Mya (27.38–15.45) in the Saharo-Sindian region; the Nigra lineage excluding *Coincya longirostra* but including *Brassica nigra* and its wild relatives diversified 17.72 Mya (27.69–13.29) in the southwestern Mediterranean region; and the Oleracea crown group diversified beginning approximately 15 Mya (19.88–11.15) in the Saharo-Sindian region and later colonized the European Mediterranean through the east 10.65 Mya (Fig. 1). Collectively, these findings demonstrate that natural hybrid formation is possible for distantly related species in the tribe. Perhaps this is not surprising given that hybrids from crosses between parental species whose genomes have been separated for long periods have also been reported elsewhere, including up to 79 Mya in angiosperms (Tripp et al., 2013), 70 Mya in ferns (Wagner et al., 1992; Sessa

et al., 2012), 55–60 Mya for birds and frogs (Prager and Wilson, 1975; Price and Bouvier, 2002), and 8 Mya for mammals (Prager and Wilson, 1975).

Diversification within the Nigra lineage excluding *Coincya longirostra* but including *Brassica nigra* and its wild relatives is dated to 17.72 Mya (95% HPD: 27.69–13.29) and was centered in the southwestern Mediterranean region (85% of 1×10^4 trees). The Oleracea crown group diversified beginning approximately 15 Mya (95% HPD: 19.88–11.15) in the Saharo-Sindian region (73% of 10^4 trees) and later colonized the European Mediterranean through the east 10.65 Mya (Fig. 1). Thus, our analyses indicate that not only did the lineages giving rise to the crop diploids *B. nigra* and *B. oleracea* + *B. rapa* diverge from each other 20 Mya, but also in distinctly different biogeographical regions of the Mediterranean.

Molecular phylogenies for the tribe also identified a subclade within Oleracea know as core Oleracea containing *B. oleracea*, *B. rapa*, and their respective wild relatives (Fig. 1). This clade originated in the northeastern Mediterranean (72% of 1×10^4 trees) 6.47 Mya (95% HPD: 9.25–4.17) late in the Miocene and possibly during the Messinian Salinity Crisis. Moreover, *B. oleracea* varieties and their wild relatives diversified in the northeastern Mediterranean (72% of 1×10^4 trees) 1.44 Mya (95% HPD: 2.45–0.66) and later spread through the rest of Europe. The origin of *B. oleracea* and crop varieties has been hotly debated for the past 20 years. Authors have suggested the species originated in the North Atlantic region of Europe (Song et al., 1990), others proposed a Mediterranean origin (Schiemann, 1932; Schulz, 1936; Maggioni et al., 2010), while the last have proposed multiple origins for the crops (Snogerup, 1980). Our results support Mediterranean origin for ancestors of *B. oleracea* and *B. rapa.* Later, *B. rapa* ancestors dispersed throughout the Irano-Turanian region reaching central Asia ≈2 Mya during the Pleistocene. Data on diversification times and the biogeographic analyses presented here are a step toward understanding the origin of *Brassica* crops varieties; however, further population sampling will be required to assess the origin of crop diversity and allotetraploid crops.

In summary, we have provided data that bear on both the timing and geographic origin of the tribe Brassiceae, as well as analyses indicating a much earlier divergence of crop Brassicas than previously proposed. Our results provide insight into geologic events associated with the diversification in the tribe as well as the split of the *Brassica oleracea* + *B. rapa* clade from the *B. nigra* clade. Ancestors of these important crop species then diversified throughout the Mediterranean before contacting each other again in secondary hybridization zones. This secondary contact is proposed to have occurred nearly 20 Mya after the initial divergence of these lineages. To our knowledge, this is one of the few examples of deep coalescence hybridization within a group of crops. Our results challenge currently accepted ideas about the origins and diversification of the tribe Brassiceae and the crop brassicas and set the stage for future studies on the domestication of crops such as broccoli, cauliflower, kale, kohlrabi, Brussels sprouts (*B. oleracea*), turnip, rape, pak-choi, Chinese cabbage (*B. rapa*), and many other oil seeds and vegetables.

LITERATURE CITED

AL-SHEHBAZ, I. A. 1985. The genera of Brassiceae (Cruciferae: Brassicaceae) in the southeastern United States. *Journal of the Arnold Arboretum* 66: 279–351.

AL-SHEHBAZ, I. A. 2012. A generic and tribal synopsis of the Brassicaceae (Cruciferae). *Taxon* 61: 931–954.

ANTONELLI, A., AND I. SAN MARTIN. 2011. Mass extinction, gradual cooling, or rapid radiation? Reconstructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. *Systematic Biology* 60: 596–615.

ARIAS, T., AND J. C. PIRES. 2012. A fully resolved chloroplast phylogeny of the *Brassica* crops and wild relatives (Brassicaceae: Brassiceae): Novel clades and potential taxonomic implications. *Taxon* 61: 980–988.

ARONESTY, E. 2011. Ea-utils: "Command-line tools for processing biological sequencing data",. Available at http://code.google.com/p/ea-utils [accessed September 2012].

BEILSTEIN, M. A., N. S. NAGALINGUM, M. D. CLEMENTS, S. R. MANCHESTER, AND S. MATHEWS. 2010. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana. Proceedings of the National Academy of Sciences, USA* 107: 18724–18728.

BLONDEL, J., J. ARONSON, J.-Y. BODIOU, AND G. BOEUF. 2010. The Mediterranean region: biological diversity in space and time, vol.4, 1–401. Oxford University Press, Oxford, UK.

BOCQUET, G., B. WIDLER, AND H. KIEFER. 1978. The Messinian model—A new outlook for the floristics and systematics of the Mediterranean area. *Candollea* 33: 269–287.

BUCKLER, E. S. IV, AND N. M. STEVENS. 2006. Maize origins, domestication, and selection. *In* T. J. Motley, N. Zerega, and H. B. Cross [eds.], Darwin's harvest: New approaches to the origins, evolution, and conservation of crops, 67–90. Columbia University Press, New York, New York, USA.

CARTER, T. JR., T. HYMOWITZ, AND R. NELSON. 2004. Biogeography, local adaptation, Vavilov, and genetic diversity in soybean. *In* D. Werner [ed.], Biological resources and migration, 47–57. Springer-Verlag, Berlin, Germany.

CHEUNG, F., C. TOWN, M. TRICK, N. DROU, I. BANCROFT, Y. P. LIM, J. Y. PARK, ET AL. 2009. Comparative analysis between homoeologous genome segments of *Brassica napus* and its progenitor species reveals extensive sequence-level divergence. *Plant Cell* 21: 1912–1928.

COYNE, J. A., AND H. A. ORR. 1989. Patterns of speciation in *Drosophila. Evolution* 43: 362–381.

CRESPO, M. B., M. D. LLEDÓ, M. F. FAY, AND M. W. CHASE. 2000. Subtribe Vellinae (Brassiceae, Brassicaceae): A combined analysis of ITS nrDNA sequences and morphological data. *Annals of Botany* 86: 53–62.

DRUMMOND, A. J., AND A. RAMBAUT. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

DRUMMOND, A. J., B. ASHTON, S. BUXTON, M. CHEUNG, A. COOPER, C. DURAN, M. FIELD, ET AL. 2011. Geneious v5.4. Available from http://www.geneious.com/ [accessed September 2012].

EDGAR, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 591–597.

FAHEY, J. W., A. T. ZALCMANN, P. TALALAY. 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* 56: 5–51.

FOLTZ, D. W. 1997. Hybridization frequency is negatively correlated with divergence time of mitochondrial DNA haplotypes in a sea star (*Leptasterias* spp.) species complex. *Evolution* 51: 283–288.

FRANZKE, A., D. GERMAN, I. A. AL-SHEHBAZ, AND K. MUMMENHOFF. 2009. *Arabidopsis* family ties: Molecular phylogeny and age estimates in Brassicaceae. *Taxon* 58: 425–427.

GARCIA-CASTELLANOS, D., AND A. VILLASEÑOR. 2011. Messinian salinity crisis regulated by competing tectonics and erosion at the Gibraltar arc. *Nature* 480: 359–364.

GE, S., T. SANG, B. R. LU, AND D. Y. HONG. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proceedings of the National Academy of Sciences, USA* 96: 14400–14405.

GÓMEZ-CAMPO, C. 1980. Morphology and morpho-taxonomy of the tribe Brassiceae. *In* S. Tsunoda and K. Hinata [eds.], Brassica crops and the wild allies, biology and breeding, 3–31. Japan Scientific Societies Press, Tokyo, Japan.

GÓMEZ-CAMPO, C., AND S. PRAKASH. 1999. 2. Origin and domestication. *Developments in Plant Genetics and Breeding* 4: 33–58.

HALL, J. C., T. E. TISDALE, K. DONOHUE, A. WHEELER, M. A. AL-YAHYA, AND E. M. KRAMER. 2011. Convergent evolution of a complex fruit structure in the tribe Brassiceae (Brassicaceae). *American Journal of Botany* 98: 1989–2003.

HAWKES, J. G. 1988. The evolution of cultivated potatoes and their tuber-bearing wild relatives. *Genetic Resources and Crop Evolution* 36: 189–208.

HEDGE, I. C. 1976. A systematic and geographical survey of the Old World Cruciferae. *In* J. G. Vaughn, A. J. MacLeod, and B. M. G. Jones [eds.], The biology and chemistry of the Cruciferae, 1–45. London Academic Press, London, UK.

HO, S. Y. W., AND M. PHILLIPS. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* 58: 367–380.

HODGKIN, T. 1995. Cabbages, kales, etc. *In* J. Small and N. W. Simmonds [eds.], Evolution of crop plants, 76–82. Longman Group, Harlow, UK.

INABA, R., AND T. NISHIO. 2002. Phylogenetic analysis of Brassiceae based on the nucleotide sequences of the S-locus related gene, SLR1. *Theoretical and Applied Genetics* 105: 1159–1165.

JENKINS, J. 1948. The origin of the cultivated tomato. *Economic Botany* 2: 379–392.

KOCH, M. A., M. KIEFER, D. A. GERMAN, I. A. AL-SHEHBAZ, A. FRANZKE, K. MUMMENHOFF, AND R. SCHMICKL. 2012. BrassiBase: Tools and biological resources to study characters and traits in the Brassicaceae—version 1.1. *Taxon* 61: 1001–1009.

LO PRESTI, R. M., AND C. OBERPRIELER. 2009. Evolutionary history, biogeography and eco-climatological differentiation of the genus *Anthemis* L. (Compositae, Anthemideae) in the circum-Mediterranean area. *Journal of Biogeography* 36: 1313–1332.

LUKENS, L. N., P. A. QUIJADA, J. UDALL, J. C. PIRES, M. E. SCHRANZ, AND T. C. OSBORN. 2004. Genome redundancy and plasticity within ancient and recent Brassica crop species. *Biological Journal of the Linnean Society* 82: 665–674. doi:

LYSAK, M. A., K. CHEUNG, M. KITSCHKE, AND P. BUREŠ. 2007. Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. *Plant Physiology* 145: 402–410.

LYSAK, M. A., M. A. KOCH, A. PECINKA, AND I. SCHUBERT. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Research* 15: 516–525.

MADDISON, W. P., AND D. R. MADDISON. 2007. Mesquite: A modular system for evolutionary analysis, version 2.0. Available at http://mesquiteproject.org [accessed February 2013].

MAGGIONI, L., R. VON BOTHMER, G. POULSEN, AND F. BRANCA. 2010. Origin and domestication of cole crops (*Brassica oleracea* L.): Linguistic and literary considerations. *Economic Botany* 64: 109–123.

MICO, E., I. SANMARTÍN, AND E. GALANTE. 2009. Mediterranean diversification of the grass-feeding Anisopliina beetles (Scarabaeidae, Rutelinae, Anomalini) as inferred by bootstrap-averaged dispersal-vicariance analysis. *Journal of Biogeography* 36: 546–560.

PARKIN, I. A. P., S. M. GULDEN, A. G. SHARPE, L. LUKENS, M. TRICK, T. C. OSBORN, AND D. J. LYDIATE. 2005. Segmental structure of the *Brassica napus* genome based on comparative analysis with Arabidopsis thaliana. *Genetics* 171: 765–781.

PAUN, O., F. FOREST, M. F. FAY, AND M. W. CHASE. 2009. Hybrid speciation in angiosperms: Parental divergence drives ploidy. *New Phytologist* 182: 507–518.

PRAGER, E. M., AND A. C. WILSON. 1975. Slow evolutionary loss of the potential for interspecific hybridization in birds: A manifestation of slow regulatory evolution. *Proceedings of the National Academy of Sciences, USA* 72: 200–204.

PRICE, T. D., AND M. M. BOUVIER. 2002. The evolution of F1 postzygotic incompatibilities in birds. *Evolution* 56: 2083–2089.

RAMBAUT, A., AND A. J. DRUMMOND. 2003. Tracer v. 1. 4. Available at http://beast.bio.ed.ac.uk/Tracer [accessed February 2013].

ROGL, F. 1999. Mediterranean and paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography. *Geological Carpathian* 50: 339–349.

SALAMINI, F., H. OZKAN, A. BRANDOLINI, R. SCHAFER-PREGL, AND W. MARTIN. 2002. Genetics and geography of wild cereal domestication in the Near East. *Nature Reviews Genetics* 3: 429–441.

SANJUR, O. I., D. R. PIPERNO, T. C. ANDRES, AND L. WESSEL-BEAVER. 2002. Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. *Proceedings of the National Academy of Sciences, USA* 99: 535–540.

SANMARTÍN, I. 2003. Dispersal vs. vicariance in the Mediterranean: Historical biogeography of the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *Journal of Biogeography* 30: 1883–1897.

SCHAEFER, H., C. HEIBL, AND S. S. RENNER. 2009. Gourds afloat: A dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proceedings of the National Academy of Sciences, USA* 276: 843–851.

SCHIEMANN, E. 1932. Entstehung der Kulturpflazen. *Handbuch Verebwis.* 15.

SCHRANZ, M. E., M. A. LYSAK, AND T. MITCHELL-OLDS. 2006. The ABC's of comparative genomics in the Brassicaceae: Building blocks of crucifer genomes. *Trends in Plant Science* 11: 535–542.

SCHULZ, O. E. 1936. Cruciferae. *In* A. Engler and K. Prantl [eds.], Die naturelichen Pflanzenfamilien, 227–558. Verlag von Wilhelm Engelmann, Leipzig, Germany.

SESSA, E. B., E. A. ZIMMER, AND T. J. GIVNISH. 2012. Phylogeny, divergence times, and historical biogeography of New World *Dryopteris* (Dryopteridaceae). *American Journal of Botany* 99: 730–750.

SMITH, B. D. 1997. The initial domestication of *Cucurbita pepo* in the Americas 1×10^4 years ago. *Science* 276: 932–934.

SNOGERUP, S. 1980. The wild forms of the *Brassica oleracea* group (2*n*=18) and their possible relations to the cultivated ones. *In* S. Tsunoda, K. Hinata, and C. Gómez-Campo [eds.], *Brassica* crops and wild allies, biology and breeding, 121–132. Japan Scientific Societies Press, Tokyo, Japan.

SONG, K., P. LU, K. TANG, AND T. C. OSBORN. 1995. Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences, USA* 92: 7719–7723.

SONG, K., T. C. OSBORN, AND P. H. WILLIAMS. 1990. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs): 3. Genome relationships in *Brassica* and related genera and the origin of *B. oleracea* and B. *rapa* (syn. *campestris*). *Theoretical and Applied Genetics* 79: 497–506.

STEELE, P. R., K. L. HERTWECK, D. MAYFIELD, M. R. MCKAIN, J. LEEBENS-MACK, AND J. C. PIRES. 2012. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99: 330–348.

THOMPSON, J. D. 2005. Plant evolution in the Mediterranean, 2nd ed. Oxford University Press, Oxford, UK.

TRIPP, E. A., S. FATIMAH, I. DARBYSHIRE, AND L. A. MCDADE. 2013. Origin of African *Physacanthus* (Acanthaceae) via wide hybridization. *PLoS ONE* 8: e55677.

U, N. 1935 Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* 7: 389–452.

WAGNER, W. H. JR., F. S. WAGNER, A. A. REZNICEK, AND C. R. WERTH. 1992. ×*Dryostichum singulare* (Dryopteridaceae), a new fern nothogenus from Ontario. *Canadian Journal of Botany* 70: 245–253.

WARWICK, S. I., AND L. D. BLACK. 1991. Molecular systematics of *Brassica* and allied genera (subtribe Brassicinae, Brassiceae)—Chloroplast genome and cytodeme congruence. *Theoretical and Applied Genetics* 82: 81–92.

WING, S. L. 1987. Eocene and Oligocene floras and vegetation of the Rocky Mountains. *Annals of the Missouri Botanical Garden* 74: 748–784.

YULE, G. U. 1924. A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 213: 21–87.

ZARDOYA, R., AND I. DOADRIO. 1999. Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *Journal of Molecular Evolution* 49: 227–237.