Phylogeny and Biogeography of the Eastern North American Rose Gentians (*Sabatia*, Gentianaceae)

Author(s): Katherine G. Mathews, Michelle S. Ruigrok, and Guilhem Mansion
Source: Systematic Botany, 40(3):811-825.
Published By: The American Society of Plant Taxonomists
URL: http://www.bioone.org/doi/full/10.1600/036364415X689276

Systematic Botany (2015), 40(3): pp. 811–825

DOI 10.1600/036364415X689276
Date of publication September 22, 2015

Phylogeny and Biogeography of the Eastern North American Rose Gentians (*Sabatia*, Gentianaceae)

Katherine G. Mathews,[1,4] Michelle S. Ruigrok,[2] and Guilhem Mansion[3]

[1]Biology Department, Western Carolina University, Cullowhee, North Carolina, 28723, U. S. A.
[2]Highlands Biological Station, 265 North Sixth Street, Highlands, North Carolina 28741, U. S. A.
[3]Botanischer Garten und Botanisches Museum, Freie Universität Berlin, 14195 Berlin, Germany.
[4]Author for correspondence (kmathews@email.wcu.edu)

Communicating Editor: James Smith

Abstract—*Sabatia* (Gentianaceae) contains ca. 20 species, distributed mainly on the U. S. A. Gulf and Atlantic Coastal Plains. Our aims were to determine 1) phylogenetic relationships among *Sabatia* species, 2) the time and place of *Sabatia*'s origin and main areas of diversification, 3) relationships among sympatric species, and 4) how morphological and karyological characters evolved. We sequenced five noncoding cpDNA regions and nrITS for 30 accessions of *Sabatia*, *Gyrandra,* and *Eustoma*. Parsimony, likelihood, and Bayesian phylogenetic analyses were performed. Bayesian dating was done on a reduced-taxon, combined molecular dataset. The maximum clade credibility chronogram was used for ancestral area reconstruction and character optimization. Correlations between distributional, environmental and phylogenetic matrices were tested with spatial analyses. Phylogenetic analyses reveal that a *Sabatia* + *Gyrandra* clade diverged in the late Middle Miocene, with *Sabatia* subsequently splitting into western and eastern Gulf Coast clades during the early Late Miocene. Further diversification took place in the Late Miocene-Pliocene, with more recent range expansion. Pliocene glacial/interglacial periods could have triggered range contraction/expansion, associated with chromosomal changes. Closely related species of *Sabatia* tend to share both distributions and habitat types. Character optimization showed potential synapomorphies for a polymerous clade and a white-flowered clade.

Keywords—Aneuploidy, U. S. A. Southeast Coastal Plain, floral polymery, molecular dating, S-DIVA.

Molecular phylogenetic methods have helped elucidate biogeographic patterns long recognized in modern-day organismal distributions across the southeastern U. S. A. and have clarified historical causes for concurrent patterns (Avise 2000). Studies have shown that there are multiple patterns of disjunct distributions throughout the southeastern U. S. A., explainable by different dispersal and/or vicariant scenarios across different timespans (Degner et al. 2010; Ellison et al. 2012; Soltis et al. 2006). The distribution of the flowering plant genus *Sabatia* (Gentianaceae), a diverse group of flowering plants primarily found growing across the U. S. A. Southeast Coastal Plain, provides the opportunity to further study biogeography of this region in relation to adjacent areas of North America. Using a phylogenetic context, we can discern concurrent patterns of disjunctions between *Sabatia* and other taxa, which may have common historical explanations, and we can propose hypotheses for causes of diversification and character evolution in this group within an historical framework.

Sabatia is a genus of 21 species (Jim Pringle, unpublished ms.) of herbs whose native range encompasses eastern North America north to Newfoundland, west to central Texas, and south to central Coahuila and Veracruz, Mexico, and the West Indies. Pringle's revision (in prep.) of Wilbur's 1955 monograph of *Sabatia* recognizes two species described since then and resurrects two formerly described species, bringing the number of recognized species from 17 to 21. Most species (14) in the genus occur and have highly overlapping distributions on the Atlantic and/or Gulf Coastal Plains of the southeastern U. S. A., where they grow in fresh to brackish wetland habitats (Wilbur 1955; Wood and Weaver 1982; Fig. 1). Six species are found in upland habitats, including *S. angularis*, a widespread species spanning mid-Atlantic, southeastern and Gulf Coastal Plain, Appalachian Highlands, Ouachitas, and Ozark Plateau physiographic provinces; *S. capitata* of the southern Appalachian foothills; and five other species predominantly west of the Mississippi River: *S. campestris* and *S. arkansana* mainly of the Ozark Plateau, *S. formosa*, mainly of central Texas, *S. arenicola* of the western Gulf Coastal Plain from Louisiana to northern Mexico, and *S. tuberculata*, endemic to Cuatro Cienegas, Mexico. *Sabatia stellaris*, predominantly eastern Coastal Plain in distribution, has a disjunct population in northern Mexico. Despite broad sympatry and high artificial crossability (Perry 1971), *Sabatia* species are morphologically well-distinguished for the most part, and there is scarce evidence for natural hybridization.

Sabatia was first described by Adanson (1763) and named for Italian botanist, Liberato Sabbati (Perry 1971; Wilbur 1955). Adanson's misspelling of the name ("Sabatia" instead of "Sabbatia") was accepted by Pursh (1814), who further added all North American species at the time included in "*Chironia*" into *Sabatia*. Since then, the genus has been treated taxonomically by Grisebach (1839, 1845), Gray (1878), Blake (1915), Wilbur (1955), and Perry (1971). Previous phylogenetic studies of subtribe Chironiinae found *Gyrandra*, a New World segregate of *Centaurium*, to be the sister genus to *Sabatia*, and *Eustoma* to be sister to *Gyrandra* + *Sabatia* (Mansion 2004; Mansion and Struwe 2004; Mansion and Zeltner 2004).

Sabatia displays interesting morphological and karyological character evolution in keeping with other taxa of Chironiinae. The majority of *Sabatia* species have pentamerous flowers, but five species are consistently polymerous (*S. bartramii*, *S. capitata*, *S. dodecandra*, *S. foliosa*, *S. gentianoides*, and *S. kennedyana*), with 9–12 sepals, petals and stamens (Wilbur 1955; Perry 1971). One species, *S. calycina*, is occasionally polymerous with up to seven sepals, petals and stamens. Floral polymery appears elsewhere in subtribe Chironiinae (*Blackstonia*) and in family Gentianaceae (*Anthocleista*, *Potalia*).

Petal color varies from pink to white (often with a yellow or greenish eye); *Sabatia brevifolia*, *S. difformis*, *S. macrophylla*, and *S. quadrangula* have white flowers, while the remainder of the species have pink flowers or are polymorphic. Most species are protandrous and have been shown to be primarily outcrossing (Perry 1971). Two species, *S. calycina* and *S. arenicola*, are autogamous (Perry 1971). Finally, species vary among annual, biennial and perennial longevity.

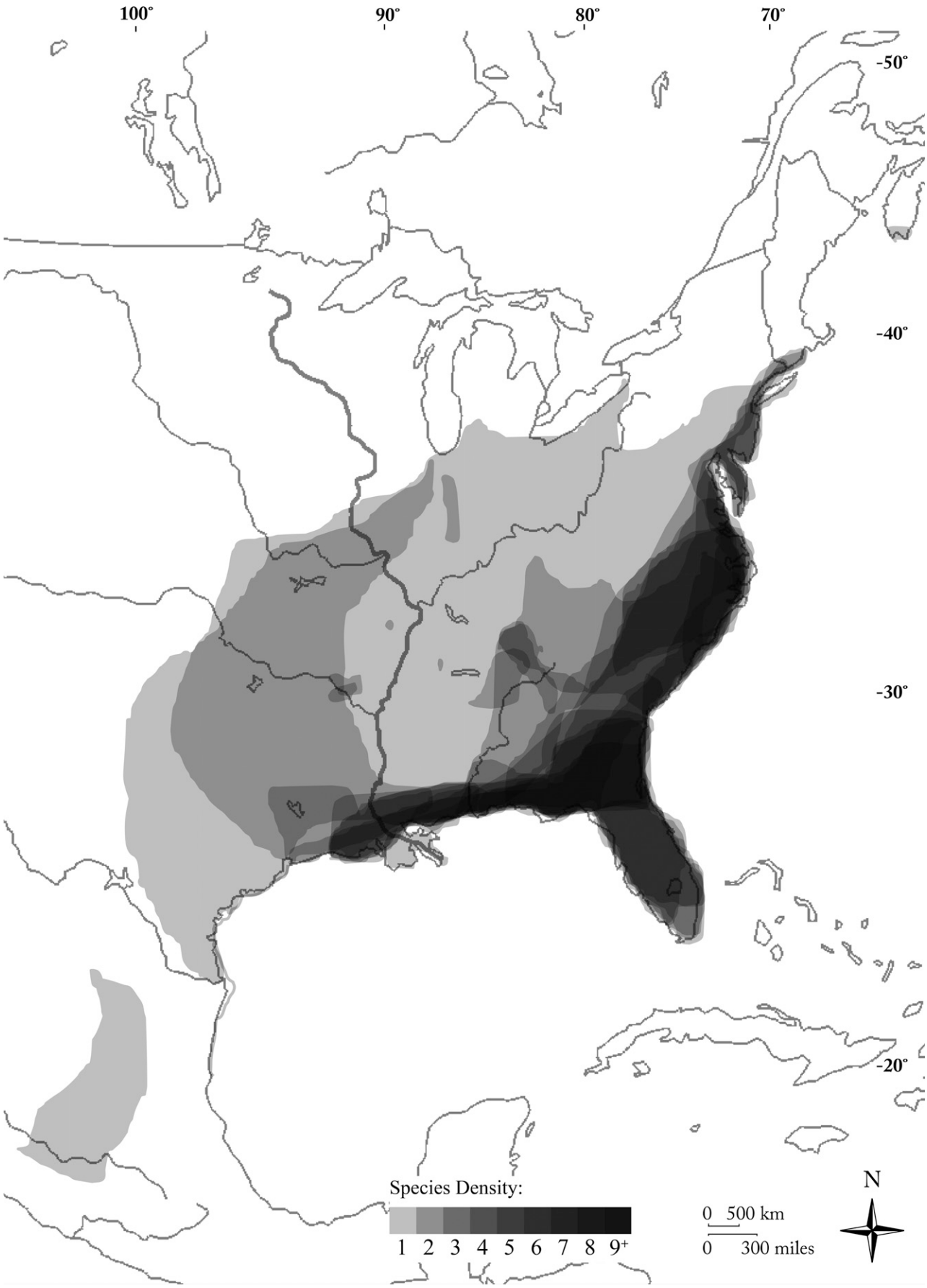

FIG. 1. Distribution of *Sabatia* species. Grey areas indicate areas of individual species distributions, with darker gray areas indicating areas of overlapping distributions with two or more species.

Nine different chromosome numbers both aneuploid and polyploid, from $n = 13$ to $n = 38$, have been reported within *Sabatia* (Perry 1971). Other species-rich members of subtribe Chroniinae, including *Centaurium* (20 species) and *Zeltnera* (25 species), also show a range of aneuploid numbers suggesting that aneuploidy is a trait associated with speciation in these related plant groups (Mansion and Zeltner 2004; Mansion et al. 2005).

The purpose of this study is to analyze the interspecific relationships of *Sabatia* and to examine patterns of biogeography, spatial structuring, and character evolution as they relate to timing and possible modes of speciation. We are specifically asking the following questions:

First, where and when did *Sabatia* first evolve, and where and when did most diversification take place? Perry (1971) proposed that *Sabatia* originated in the Interior Uplands and migrated to the Coastal Plain following the receding of the Mississippi Embayment during the Pliocene and Pleistocene periods. According to this hypothesis, ancestral lineages could be Appalachian or of the Western Uplands (Ozark plateau). *Sabatia campestris* ($n = 13$), primarily of the Ozark plateau, and the morphologically related *S. arenicola* ($n = 14$), of the western Gulf Coastal Plain, have the lowest chromosome numbers and are predicted to be ancestral in the genus.

Second, how are species of *Sabatia* distributed spatially in relation to phylogeny? More specifically, we want to know whether closely related species tend to be sympatric or not, and how phylogeny and sympatry correlate with habitat usage. As mentioned above, most species of *Sabatia* have overlapping distributions at least in part, especially on the coastal plain. Kozak et al. (2005) looked at multiple effects on community interactions in dusky salamanders with overlapping distributions in the southern Appalachians and found that closely related species co-occurred but segregated along a microhabitat gradient and by body size, which was related to resource use. Assuming most speciation in *Sabatia* occurred on the coastal plain where the majority of species are currently distributed, we predict that closely related species of *Sabatia* would co-occur but would not make use of the same ecological niches.

Third, how can character evolution and taxonomy be informed by phylogeny? The variable characters described above (chromosome number and ploidy level, flower color, floral merosity, breeding system, and longevity) may be associated with speciation and/or have been used to delimit infrageneric taxa in previous classification schemes in *Sabatia*. In this study, we map character states onto a phylogenetic tree to determine the synapomorphic value of the states for these characters to help derive hypotheses regarding their evolution. For instance, there could be selective pressure to evolve large, polymerous flowers in primarily outcrossing species like *Sabatia*, in which case polymery could have evolved multiple times. Alternatively, polymery could have evolved only once in *Sabatia*, with all the polymerous species being derived from a single pentamerous ancestor.

Materials and Methods

***Taxon Sampling*—**In this study, we followed the taxonomy of Wilbur (1955), Perry (1971) and Pringle (in prep). A total of 17 species, encompassing 85% of the described genus diversity, were sampled for this study. For nine species, multiple collections from different geographic localities were included. Two accessions of the sister genus *Gyrandra* (Mansion and Struwe 2004) were added to the analyses and two species of *Eustoma* were used as outgroup (Table 1; Appendix 1), giving a total of 30 accessions included in the study. The main characteristics for species included in this study are shown in Table 1.

Table 1. Taxonomic authorities, haploid chromosome number (n), geographic distribution, habitat, and wetland status for the 17 species of *Sabatia* and five outgroups included in this study. Geographic distributions are abbreviated as follows: AH = Appalachian highlands, ACP = Atlantic Coastal Plain, EGCP = eastern Gulf Coastal Plain, WGCP = western Gulf Coastal Plain, IH = interior highlands, PF = Peninsular Florida).

Species	*n*	Geographic Distribution	Habitat	Wetland Status
S. angularis Pursh	19	ACP,EGCP,WGCP,AH,IH	Dry woods, fields	Facultative hydrophyte
S. arenicola Greenm.	14	WGCP	Sand	Facultative hydrophyte
S. bartramii Wilbur	18	EGCP,PF	Wet savannah, pine barrens, Fresh marsh	Hydrophyte
S. brachiata Elliott	16	ACP,EGCP,WGCP,AH,IH	Dry woods, fields	Non-hydrophyte
S. brevifolia Raf.	16	ACP,EGCP,PF	Wet savannah, pine barrens	Hydrophyte
S. calycina (Lam.) A.Heller	32	ACP,EGCP,WGCP,PF	Fresh marsh, riverbank	Hydrophyte
S. campanulata Torr. Ex Griseb.	17	ACP,EGCP,WGCP,AH,IH	Bog, wet savannah, fresh marsh	Hydrophyte
S. campestris Nutt.	13	WGCP,IH	Wet savannah, fields, sand	Non-hydrophyte
S. capitata S.F.Blake	38	AH	Dry woods	Non-hydrophyte
S. difformis (L.) Druce	19	ACP,EGCP,PF	Wet savannah, pine barrens	Hydrophyte
S. dodecandra (L.) Britton, Sterns & Poggenb.	17	ACP,EGCP	Salt marsh, fresh marsh, Riverbank	Hydrophyte
S. gentianoides Elliott	14	ACP,EGCP,WGCP	Wet savannah, pine barrens	Hydrophyte
S. *grandiflora* (Gray) Small	18	PF	Sand, pine barrens, salt marsh	Hydrophyte
S. kennedyana Fernald	20	ACP	Fresh marsh, riverbank	Hydrophyte
S. *macrophylla* Hook var. *macrophylla*	19	EGCP	Wet savannah, pine barrens, Fresh marsh	Hydrophyte
S. quadrangula Wilbur	16,17	ACP,EGCP,AH	Dry woods, fields, wet Savannah (rock outcrop)	Hydrophyte
S. stellaris Pursh	18	ACP,EGCP,WGCP,MCA	Salt marsh, sand	Hydrophyte
Gyrandra brachycalyx (Standl. & L.O. Williams) Mansion	36	MCA	Open pine-oak forest communities (1500–2800 m)	
G. tenuifolia (M. Martens & Galeotti) Mansion	36	MCA	Pine forests (1300–2800 m)	
Eustoma exaltatum (L.) Salisb. Ex G. Don	36	EGCP,WGCP,PF,GP,MCA	Damp prairies or low open ground, alkaline or saline soil	
E. grandiflorum (Raf.) Shinners	36	GP	Damp prairies or low open ground, alkaline or saline soil	

Laboratory Protocol—DNA extraction was performed using leaf material either from silica gel-dried samples collected in the field or from herbarium specimens. Total DNA was extracted from 30–50 mg of leaf tissue using the DNEasy plant mini kit (Qiagen Inc., Valencia, California) or a modified CTAB extraction method (Doyle and Doyle 1987).

We employed five non-coding chloroplast DNA (cpDNA) regions and one nuclear marker (internal transcribed spacer, ITS). The *trnD-T* intergenic spacer, *trnS-G-G* intron and spacer, and *atpI-H* intergenic region were amplified with primers from Shaw et al. (2005; 2007). The *trnL* intron and *trnL-F* spacer were amplified separately using primers c and d (intron) and primers f and e (spacer) from Taberlet et al. (1991). Finally, ITS was amplified with primers 4 and 5 of White et al. (1990). All regions were selected based on their high variability and primer universality. We were unable to obtain readable sequence for some accessions in each dataset, but we included all samples we could obtain even if missing data for some. Recent studies have shown that adding more taxa with partial data, even up to 75% missing data, generally increases phylogenetic accuracy relative to excluding the incomplete taxa, and can especially help resolve poorly supported nodes (Weins and Tiu 2012; Jiang et al. 2014). None of our samples was missing more than 73% data, and only four of the 30 accessions were missing between 50–73% data. These four included one accession each of *S. calycina*, *S. differmis*, *S. kennedyana*, and *S. macrophylla*, taxa for which we also included an additional sample each with more complete data. Number of missing accessions and percent missing data for each gene region is as follows: *trnD-T*: 10/30 = 33%; *trnS-G-G*: 8/30 = 26%; *atpI-H*: 8/30 = 26%; *trnL* intron/*trnL-F* spacer: 4/30 = 13%; *ITS*: 4/30 = 13%. The total percent of missing data cells in the 30-taxon combined data matrix is 30%.

Samples were PCR-amplified in 25 μl reactions, including the following reagents: 12.5 μl PCR master mix (Promega, Madison, Wisconsin), containing *Taq* DNA polymerase, dNTPs, $MgCl_2$ and reaction buffers, 8.5 μl dH_2O, 1 μl bovine serum albumin (BSA, 1 mg/ml), 1 μl of template DNA, and 1 μl each of forward and reverse primers (10 μm). In some reactions, water was replaced with an additional 1.25 μl of $MgCl_2$ to improve amplification. The PCR cycling parameters for *trnD-T*, *trnS-G-G* and *atpI-H* are given in Shaw et al. (2005, 2007). For a few difficult to amplify samples, we used gradient PCR to find the ideal annealing temperature. For *trnLF* amplification we used an initial denaturation for two min at 95°C, followed by 35 cycles of 50 sec at 95°C, 50 sec at 50°C, and one min 50 sec at 72°C, followed by a final extension for five min at 72°C. For ITS amplification we used an initial denaturation for three min at 95°C, followed by 30 cycles of one min at 95°C, one min at 45°C, and one min 20 sec at 72°C, followed by a final extension for seven min at 72°C.

The PCR products were run on a 1% agarose gel stained with ethidium bromide to visualize the product. Successful reactions were cleaned using the QIAquick PCR purification kit (Qiagen Inc., Valencia, California) or ExoSAP-IT® (Affymetrix, Inc., Cleveland, Ohio) prior to sequencing reactions. Sequencing PCR was done in 5 μl reactions with the BigDye® Terminator 3.1 cycle sequencing kit (Applied Biosystems, Foster City, California). Both template strands were sequenced for each sample using the same PCR primers. Each sample consisted of 2 μl Big Dye, 1.6 μl of 1.0 μM primer, 0.4 μl of water, and 1.0 μl of purified PCR product. Completed sequencing reactions were purified with Centri_Sep™ spin columns (Princeton Separations, Adelphia, New Jersey), dried and resuspended in 10 μl of Hi-Di formamide (Applied Biosystems). Samples were then electrophoresed and analyzed on an AB3130 Genetic Analyzer (Applied Biosystems).

Phylogenetic Analyses—Sequences from each gene region were edited in Sequencher 4.2 (Gene Code, Ann Arbor, Michigan) to trim the ends and compare opposing strands for accuracy. Edited sequences were aligned using Muscle (Edgar 2004), with additional manual corrections in PhyDe (Müller et al. 2005). Indels were coded as binary characters using simple gap coding (Simmons and Ochoterena 2000). Phylogenetic analyses were first performed on 30-taxon ITS and cpDNA datasets separately using maximum parsimony, maximum likelihood, and Bayesian inference (see details below). Following Mansion and Struwe (2004), *Eustoma* was used as outgroup in all analyses. To explore each dataset's contribution and check for potential conflicting results, the individual matrices were further combined and an incongruence length difference test (ILD; Farris et al. 1995), implemented in PAUP* 4.0 as the "partition homogeneity test" (PHT; (Swofford 2002), was then performed with 100 replicates of heuristic searches, random sequence addition, and TBR branch swapping. The combined data matrix is available at TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S16940).

Maximum parsimony (MP) analyses were performed in PAUP* 4.0b10 (Swofford 2002) using a Fitch criterion. Heuristic searches were conducted with a ratchet batchfile generated with PRAP (Müller 2004), and including 200 iterations with 25% of the positions randomly weighted (weight = 2), and 100 random additions. Branch support was calculated with the bootstrap (BS) method, using 10,000 replicates, TBR branch swapping, 10 random-additions, multrees option OFF, and resampling all characters.

Maximum likelihood (ML) analyses were conducted with RAxML (Stamatakis 2006). One hundred runs with a fast hill-climbing algorithm and the default model of sequence evolution (option d with GTRGAMMA) were performed for the optimal ML tree calculation, and 1,000 BS replicates using a fast hill-climbing algorithm were used for branch-support calculation (option "a" with GTRCAT).

The Bayesian inference (BI) analyses were conducted with MRBAYES (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003), using six simultaneous runs of Metropolis-coupled Markov Chain Monte Carlo (MC3). The best-fit models of substitution for each individual partition, computed under the AIC implemented in MrModeltest (Nylander 2004), were as follow: ITS: GTR + gamma; *atpH1*: GTR + gamma, *trnDT*: GTR + gamma; *trnLF*: GTR + invariant; *trnSG*: GTR + invariant + gamma; coded indels: binary model (Lset coding = variable). Each chain was run in parallel for six million generations, saving one tree each 1,000th generation, keeping a default temperature parameter value of 0.2. The MC3 runs were repeated twice, and the first 10% of the saved trees were discarded as burn-in after checking for (i) stationarity on the log-likelihood curves; (ii) similarity of the respective majority-rule topologies and final likelihood scores; (iii) the values of standard deviation of split frequencies (< 0.001); and (iv) the value of the potential scale reduction factor (close to 1). The remaining trees were used to produce a majority-rule consensus tree and to calculate the posterior probability (pp) values.

Dating Analyses—To add a temporal dimension to our phylogenetic inference, we further performed dating analyses on the combined 26-taxon dataset. Divergence times were estimated under a relaxed clock method using the Bayesian software BEAST 1.4.7 (Drummond and Rambaut 2007), with an uncorrelated lognormal model, i.e. not assuming autocorrelation of molecular rate variation between ancestral and descendant lineages. The transformation of relative ages into absolute ages was performed using the following two-steps strategy. A preliminary BEAST analysis was performed for the whole Gentianaceae, including 79 ITS and *trnLF* accessions retrieved from GenBank, and overall encompassing most of the extant generic diversity in the family (Mansion, unpubl. data). The root age was set to 79 ± 10 Mega Annuum (Ma), the estimated age for the Gentianales following Janssens et al. (2009), and two internal calibration points were used, in agreement with Favre et al. (2010) and Merckx et al. (2013): (1) a potential *Lisianthius* fossil pollen from late Eocene (37 Ma; Graham 1984) was conservatively assigned to the crown node of the Potalieae (log normal distribution, offset = 37, mean = 5, SD = 1.6); and (2) an estimated age for the Swertiinae using ITS mutation rates (15 Ma; von Hagen and Kadereit 2001) was assigned to the crown of the subtribe (normal distribution, mean = 15, SD = 1). This preliminary analysis inferred respective ages of 27.73 Ma (95% confidence intervals, CI: 19.73–36.34) for crown node of the subtribe Chironiinae and 9.02 Ma (CI: 4.62–13.95) for the split between *Gyrandra* and *Sabatia*. Those estimates were then used as secondary calibration points in our 26-taxon dataset analysis by assigning normal distribution constraints to both the root (treeModel.rootHeight = 30 ± 1 Ma) and the *Gyrandra-Sabatia* split (9 ± 1.7 Ma).

The use of a normal prior distribution, which allows the inclusion of error associated with the primary molecular estimate in the subsequent analysis, is more appropriate for calibration points based on estimates from independent molecular studies (secondary calibrations) for it allows uncertainty to be equally distributed on either side of the mean age (Forest 2009).

The optimal substitution model was selected for the global dataset using MrModeltest (Nylander 2004). Posterior distributions of parameters were estimated by Markov Chain Monte Carlo (MCMC) sampling, each single chain being run for 10^8 generations, sampled every 10^4 steps, with a 10% burn-in. Analyses were repeated two times, and the posterior probability density was summarized in Tree-Annotator version 1.4.1 (Drummond and Rambaut 2007). The quality of posterior parameters sampling was checked with the software Tracer 1.5 (Rambaut and Drummond 2007).

Biogeographic and Spatial Analyses—Ancestral area reconstruction was performed using a statistical dispersal-vicariance analysis (S-DIVA), an implementation of the Bayes-DIVA method that integrates DIVA parsimony-based reconstructions over a posterior distribution of tree topologies (Nylander et al. 2008; Ronquist 1997; Yu et al. 2010). S-DIVA was conducted on the BEAST chronogram of the 26-taxon dataset, and optimized onto a random set of 1,000 BEAST trees.

Geographic distributions of *Sabatia* species were taken from Perry (1971) and Wilbur (1955). These distributions were assigned to one or more of eight physiogeographic areas defined by the USGS (2004) based on terrain, rock type, and geologic structure and history and which correspond to unique floristic provinces (e.g. Braun 1955; Sorrie and Weakley 2001). We abbreviated the areas as follows: A = Atlantic Coastal Plain (ACP), B = eastern Gulf Coastal Plain (EGCP), C = western Gulf Coastal Plain (WGCP), D = Appalachian Highlands (AH); E = interior Highlands (IH), F = Peninsular Florida (PF), G = Great Plains (GP) and H = Mexico and Central America (MCA). Reconstructions were done allowing either three (the mean number of areas for all extent species) or up to five (the maximum number of areas for all extant species) areas per node.

The spatial analysis program PASSaGE v.2 (Rosenberg and Anderson 2011) was used to conduct Mantel tests of correlation between environmental and phylogenetic matrices of the 15 *Sabatia* species included in the 26-taxon dataset. All but one accession of each species was pruned from the Bayesian chronogram (see below), which was then converted into an ultrametric tree in the program FigTree 1.4.0 (Rambaut 2012). The ultrametric pruned tree was converted into a pairwise phylogenetic distance matrix in PASSaGE. We also created three pairwise binary matrices, one for shared geographic distribution (overlapping in at least part of the species' ranges = 1, not overlapping = 0), one for habitat similarity (similarity in habitat at least in part = 1, no similarity in habitat = 0), and one for wetland status (same status = 1, different status = 0). The same eight geographic areas used in the S-DIVA analyses were used. Habitat preferences of *Sabatia* species followed Perry (1971) and Wilbur (1955). Eight habitat categories were defined as follows: sphagnum bog, wet savannah, pine barren, fresh marsh/pond, riverbank, salt marsh, sand (including beaches and dunes), dry (any setting not seasonally or permanently inundated, including dry fields, woods and roadsides in upland localities). Wetland status was taken from the USGS Plants Database (USDA and NRCS 2014). Matrices are available in Appendix 2. Two-way Mantel tests were conducted on each pair of matrices, and a three-way (partial) Mantel was performed on the geographic, habitat and wetland matrices while holding the phylogenetic distance matrix constant. Permutation tests were used with 9,999 iterations to obtain a significance value for the Mantel correlation coefficient.

Character Optimization—Character optimization of six selected diagnostic morphological and karyological characters (corolla color, floral merosity, plant longevity, reproductive system, haploid chromosome number, and ploidy level) was performed using a ML approach with a Markov k-state one-parameter model (Mk1) implemented in Mesquite v2.74 (Maddison and Maddison 2009). Ancestral states were reconstructed for 5,000 randomly chosen BEAST trees (combined 26-taxon dataset) and plotted onto the maximum clade credibility chronogram (see Results). Character coding for each species is given in Appendix 3. Regarding ploidy level coding, the following strategy was adopted. As frequent diploid numbers within tribe Chironiinae include $2n = 2x = 18$, 20 and 22 (Mansion and Struwe 2004; Mansion and Zeltner 2004; Mansion et al. 2005), we considered most of *Sabatia* species as tetraploids (n = 16–20), except *S. arenicola* and *S. campestris* (diploid, n = 13, 14) and *S. capitata* plus the outgroup (octoploid, n = 36, 38).

Results

Phylogenetic Analyses—Characteristics of the respective ITS, cpDNA, and combined datasets are shown in Table 2. Maximum parsimony (MP) analysis of the ITS dataset revealed 145 shortest trees of length L = 221 with the following statistics: consistency index (CI) = 0.79, retention index (RI) = 0.86, rescaled index (RC) = 0.68. Under the same criterion, 600 shortest trees (L = 919, CI = 0.81, RI = 0.71, RC = 0.57) and 52 shortest trees (L = 1,154, CI = 0.79, RI = 0.75, RC = 0.59) were inferred for the cpDNA and combined datasets, respectively. Maximum likelihood (ML) analyses performed under the GTR + GAMMA model of sequence evolution produced trees with the following scores: ITS: -ln = 1,696.0477; cpDNA: -ln = 11,060.6169; and combined: -ln = 12,866.4684 (Table 2). Finally, Bayesian inference (BI) analyses of the respective datasets yielded topologies congruent with the ones inferred under both MP and ML criteria.

The PHT first showed statistical topological differences between the respective nuclear and chloroplast partitions at an alpha of 0.05 ($p = 0.02$). The exclusion of four accessions, two of *S. calycina* and two of *S. quadrangula*, which were polyphyletic in the ITS tree, resulted in no significant differences ($p = 0.15$) between the two partitions in the resulting combined 26-taxon dataset. This pruned version of the combined dataset was further used to infer dating and biogeographic analyses, and optimize characters, since topology-based analyses are likely to be sensitive to incongruent phylogenetic position of taxonomic units. For phylogenetic purposes however, being aware of potential reticulation patterns in the group, we combined a 30-taxon dataset and followed the general picture inferred from the more conserved MP topology to describe the phylogenetic relationships below, mentioning when necessary minor discrepancies with the remaining analyses (Figs. 2–4).

The MP strict consensus tree (Fig. 2) rooted with *Eustoma* accessions, strongly support the monophyly of *Sabatia* (BS 100) and an early lineage split within *Sabatia* dividing a "western" *S. campestris*-*S. arenicola* clade (clade W: BS 100) from the remaining, primarily "eastern", *Sabatia* species (clade E: BS 100). Within the eastern clade, three sublineages can be described, namely an "eastern 1" (E1) clade (BS 99) encompassing *S. macrophylla*, *S. difformis*, and *S. quadrangula*, an "eastern 2" (E2) clade (BS 95) with *S. calycina*, *S. stellaris*, and *S. campanulata*, and an "eastern 3" (E3) clade, comprising eight species (*S. angularis*, *S. calycina*, *S. gentianoides*, *S. capitata*, *S. kennedyana*, *S. bartramii*, and *S. dodecandra*). The E3 clade showed no significant branch support in the MP or ML analyses (BS < 50, BS 64, respectively) but a marginally acceptable value (PP = 0.9) in the BI analysis (Fig. 4). Finally, the phylogenetic position of four species

Table 2. Characteristics of the respective phylogenetic analyses.

	30-taxon (combi)	30-taxon (nr)	30-taxon (cp)	26-taxon (combi)
Characters				
Total Aligned Length	5,397	482	4,915	5,357
Number of coded indels	169	17	152	161
Maximum Parsimony Analyses				
Parsimony-informative characters (%)	385 (7.2)	41 (8.5)	276 (5.6)	371 (6.9)
N trees	54	145	600	9
Length	1,154	221	919	1,042
Consistency Index	0.80	0.79	0.81	0.83
Retention Index	0.75	0.86	0.71	0.77
Rescaled Index	0.60	0.68	0.57	0.64
Maximum Likelihood Analyses				
Model of sequence evolution (default)	GTR + G	GTR + G	GTR + G	GTR + G
Likelihood score of best tree (-ln)	12,866.4684	1,696.0477	11,060.6169	12,416.7748

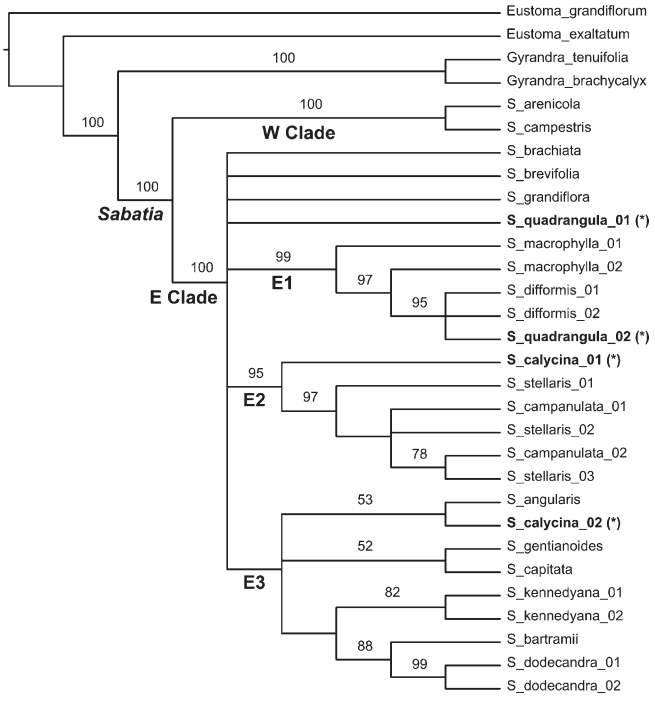

FIG. 2. Maximum parsimony strict consensus tree of *Sabatia* and relatives (30-taxon dataset). Values above branches indicate bootstrap support for sustained clade. W and E stand for the respective western and eastern clades, the latter subdivided into three subclades (E1 to E3). Accessions with incongruent topological position are indicated in bold.

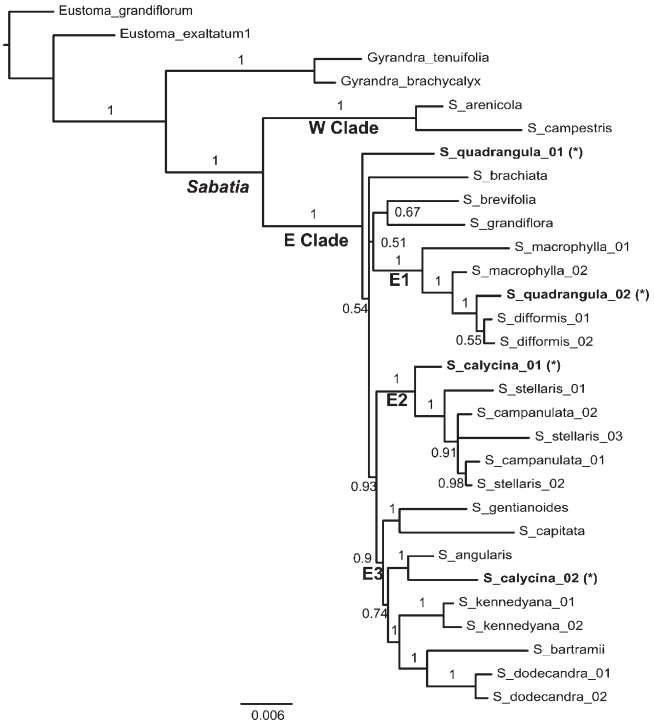

FIG. 4. Bayesian majority-rule phylogram of *Sabatia* and relatives (30-taxon dataset). Posterior probability values are indicated above branches. W and E stand for the respective western and eastern clades, the latter subdivided into three subclades (E1 to E3). Accessions with incongruent topological position are indicated in bold.

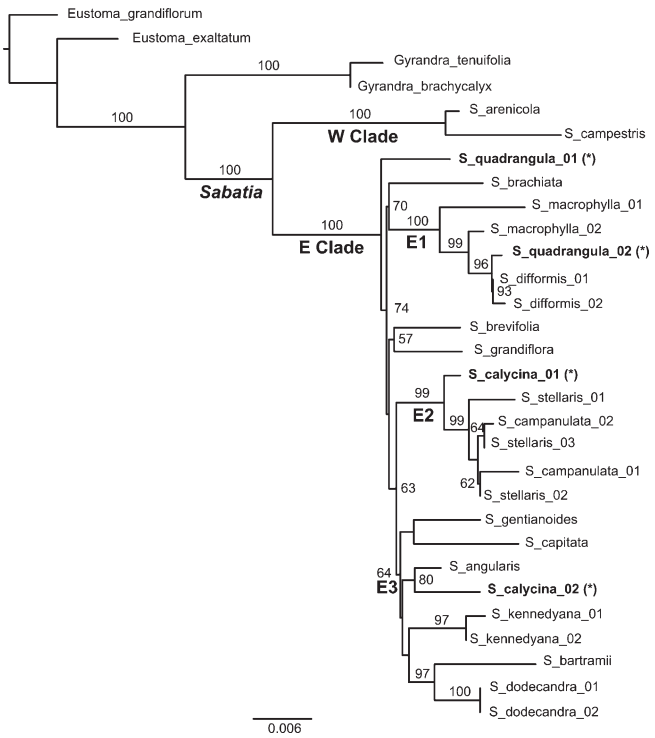

FIG. 3. Best maximum likelihood phylogram of *Sabatia* and relatives (30-taxon dataset). Values above branches indicate bootstrap support for sustained clade. W and E stand for the respective western and eastern clades, the latter subdivided into three subclades (E1 to E3). Accessions with incongruent topological position are indicated in bold.

(*S. brevifolia*, *S. brachiata*, *S. grandiflora*, and *S. quadrangula*) remained unresolved in the MP analysis. Similar analyses produced with the exclusion of four conflicting accessions (26-taxon dataset) produced congruent topologies, except for the exclusion of *S. angularis* from the E3 clade (data not shown).

Dating Analyses—Independent runs inferred similar node ages and confidence intervals, along with comparable posterior parameters with high ESS. The means (and 95% confidence intervals, CI) of the respective stem and crown ages for *Sabatia* were estimated to 11.26 Ma (CI: 8.16–14.29) and 8.87 Ma (CI: 5.72–11.98), the latter corresponding to the split between the W clade (crown age = 2.33 Ma; CI: 0.43–5.45) and the E clade (crown age = 6.17 Ma; CI: 3.51–9.09; Fig. 5). Finally, the crown ages of the respective E1-E3 clades were inferred by the end of the Miocene (E1 = 5.35 Ma, CI: 2.65–8.14; E3 = 4.34 Ma, CI: 2.09–6.78) or during the Pliocene (E2 = 2.04 Ma, CI: 0.66–3.96).

Biogeographic Analyses—The S-DIVA analyses performed on the BEAST chronogram of the 26-taxon dataset with "maxareas" = three inferred the origin of *Sabatia* stem lineage in several areas, all including H (MCA), with further diversification in either AC (ACP + WGCP; 46%) or BC (EGCP + WGCP; 46%), followed by a vicariant separation between the W clade (C = WGCP; 100%) and the E clade (A = ACP or B = EGCP; both 32.7%). The origin of the respective E1 to E3 clades was always inferred in A (ACP; 50%) or B (EGCP; 50%). When increasing the "maxareas" value to five, thus favoring the occurrence of vicariant events, no major changes were observed in the inferred ancestral areas (result not shown). Based on geographic proximity,

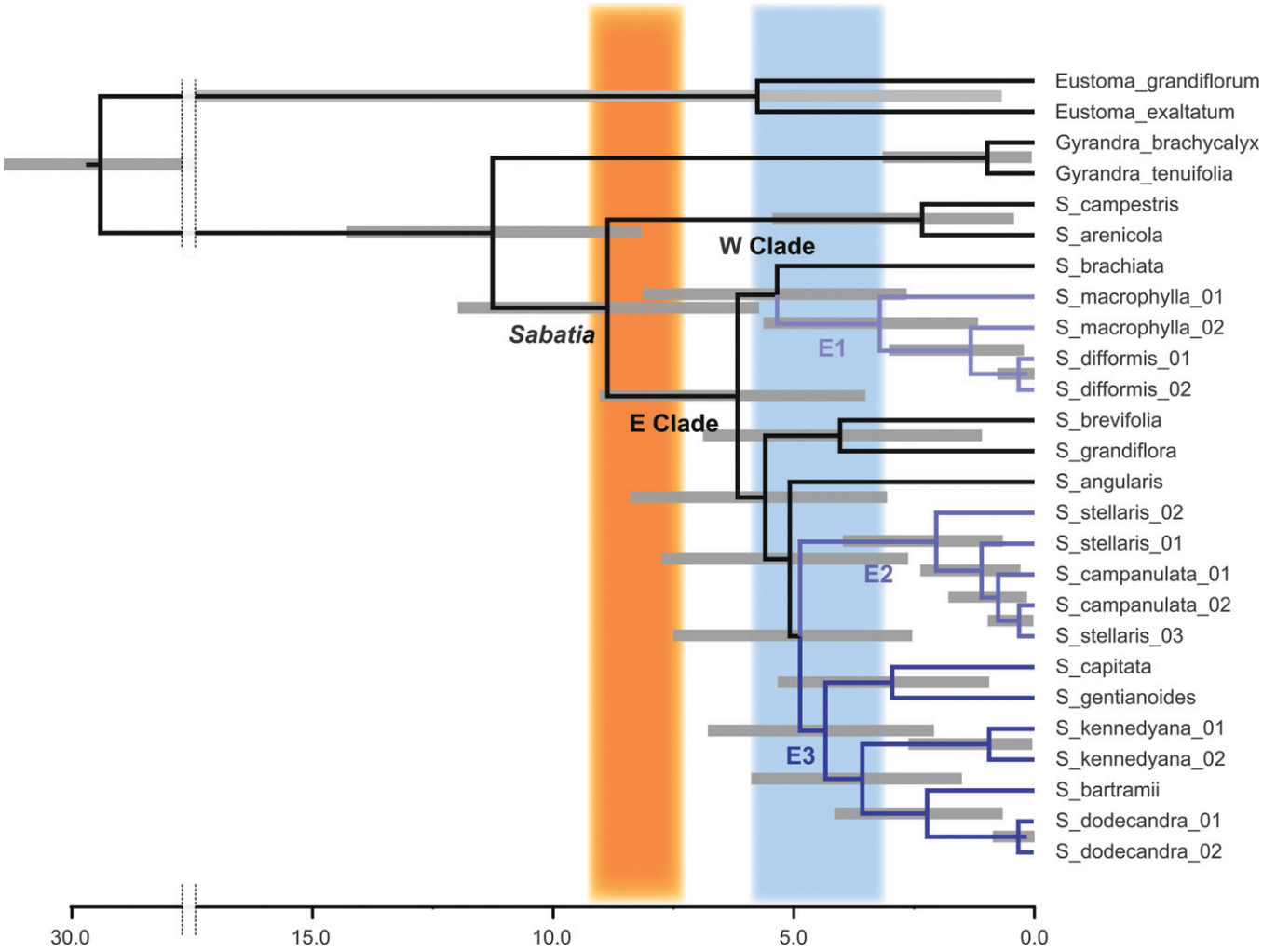

FIG. 5. Chronogram of *Sabatia* and relatives (26-taxon dataset) inferred from the relaxed clock method implemented in BEAST. W and E stand for the respective western and eastern clades, the latter subdivided into three subclades (E1 to E3). Grey bars represent 95% confidence intervals for the respective node ages. The orange (warm and xeric) and blue (cool and wet) zones indicate the important climatic changes in North America during the Middle-Late Miocene transition (see Text). Ma = Mega Annuum.

we overall favored the most likely scenario reconstructed by S-DIVA, and assuming an early vicariant event between H (MCA; *Gyrandra*) and BC (EGCP + WGCP; *Sabatia*), with further separation between the neighbor areas B (ECGP) and C (WGCP) (Fig. 6).

Spatial Analyses—We predicted that closely related species of *Sabatia* would co-occur geographically, but would not make use of the same habitat in those areas. Results from the Mantel tests showed non-significant negative correlations between phylogenetic distance and shared distribution (p = 0.08), phylogenetic distance and habitat similarity (p = 0.14), and phylogenetic distance and wetland status (p = 0.14; Table 3). We found significant positive correlations between shared distribution and habitat similarity (p = 0.001), shared distribution and wetland status (p = 0.002), and between habitat similarity and wetland status (p = 0.001). When phylogenetic distance was held constant, the association between shared distribution and wetland status became non-significant (p = 0.19), as shown by the three-way Mantel test result, indicating that some of the effect of species having both shared geographic distributions and shared wetland requirements can be accounted for by phylogenetic relationships.

Character Optimization—The "polymerous" character state is a potential synapomorphy for the poorly supported clade E3, encompassing *S. bartramii*, *S.capitata*, *S. dodecandra*, *S. gentianoides*, and *S. kennedyana* (Fig. 7B), the facultative polymerous *S. calycina* being excluded from the 26-taxon dataset. The "white-flowered" character state is synapomorphic for the strongly-supported clade E1 (*S. difformis*, *S. macrophylla*; PP = 1), *S. quadrangula* being excluded from the 26-taxon dataset, but is separately derived in *S. breviflora*, which is sister to the pink-flowered *S. grandiflora* (Fig. 7A). An annual or biennial longevity state is ancestral for *Sabatia*, with the perennial state derived, possibly multiple times (Fig. 7C). Autogamy has evolved separately in two species of *Sabatia* (only *S. arenicola* is shown in tree; *S. calycina* is the other autogamous species and never groups with *S. arenicola*) from the predominant allogamous state (Fig. 7D). Optimization of the chromosome valence reveals an equivocal ancestral state for the genus, a diploid origin for the W clade (PP = 1) and a tetraploid one for the E clade. Within the E clade, a switch to the octoploid levels

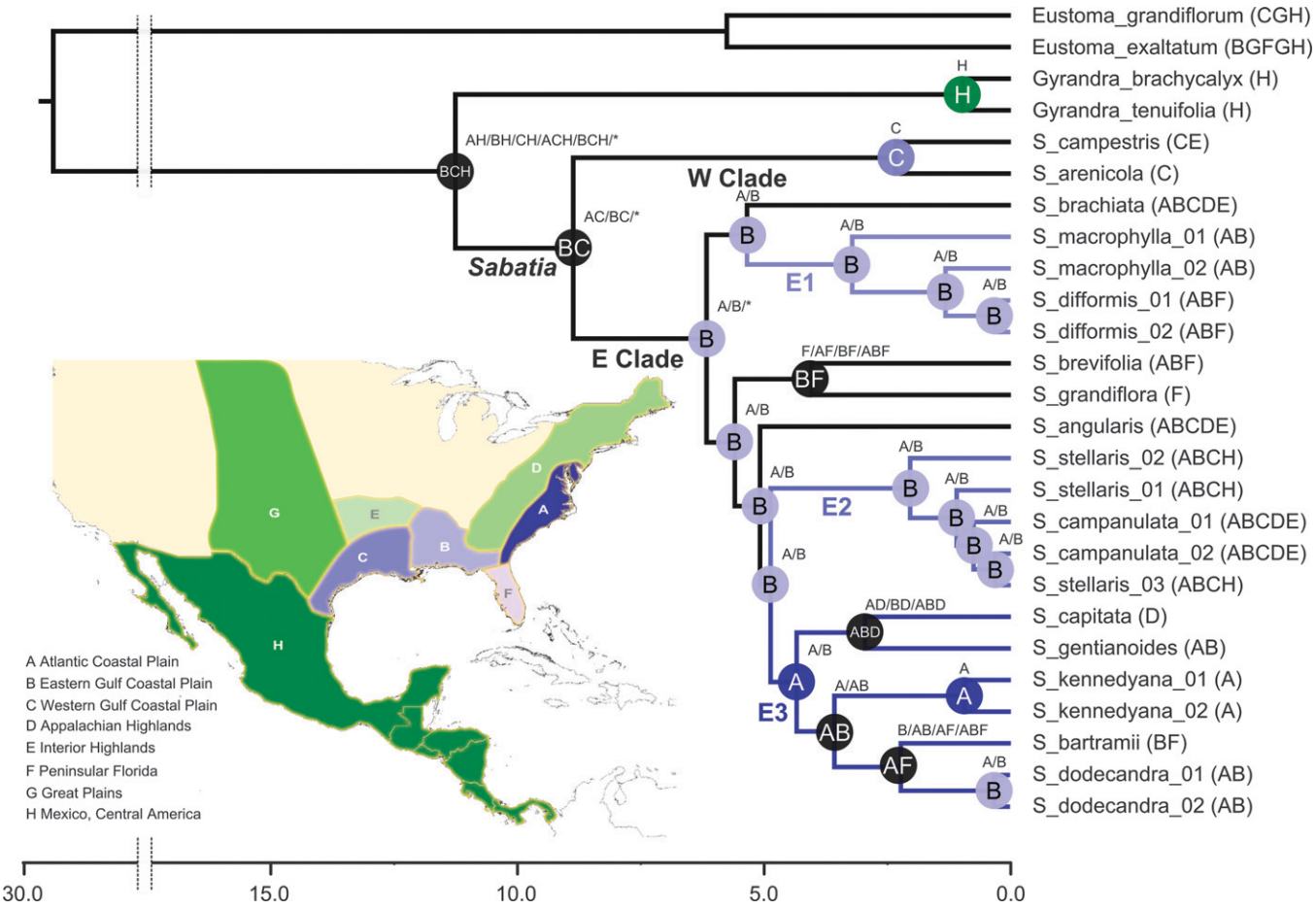

Fig. 6. Preferred biogeographic scenario for *Sabatia* inferred from S-DIVA analysis of the 26-taxon dataset. W and E stand for the respective western and eastern clades, the latter subdivided into three subclades (E1 to E3). Circles represent the most likely inference at a specific node using a maxarea value of 3. Alternative reconstructions, occurring at the same frequency, are shown above branches. An asterisk indicates the occurrence of less frequent inferences. The map shows the areas' circumscription used in the analyses. Areas are abbreviated as follows: A = Atlantic Coastal Plain (ACP), B = eastern Gulf Coastal Plain (EGCP), C = western Gulf Coastal Plain (WGCP), D = Appalachian highlands (AH), E = interior highlands (IH), F = Peninsular Florida (PF), G = Great Plains (GP), and H = Mexico and Central America (MCA).

occurred in *S. capitata*, whereas a reduction to the diploid condition occurred in the sister species *S. gentianoides*. Within the tetraploid E clade, haploid chromosome numbers range from $n = 16$ to $n = 20$, with no apparent pattern of reduction or increase, but with sister species having different chromosome numbers in all cases but one (*S. difformis-S. macrophylla*) (Figs. 7E and 7F).

Table 3. Mantel tests for association between distribution, habitat similarity, wetland status, and phylogenetic distance among 15 species of *Sabatia*. Significance of Mantel correlations coefficients (r) was evaluated using 9,999 permutations, using one-tailed tests to evaluate specific hypotheses. In the three-way (partial) Mantel test, the effect of phylogenetic relationship on geographical overlap, habitat similarity, and wetland status was held constant. **significant at <0.01, *significant at <0.05.

Matrices tested	r	p
Phylogenetic distance, shared distribution	−0.1874	0.0760
Phylogenetic distance, habitat similarity	−0.1183	0.1350
Phylogenetic distance, wetland status	−0.1212	0.1420
Shared distribution, habitat similarity	0.2055**	0.0010
Shared distribution, wetland status	0.0835**	0.0020
Habitat similarity, wetland status	0.2678**	0.0010
Phylogenetic distance, shared distribution, habitat similarity	0.1880*	0.0110
Phylogenetic distance, shared distribution, wetland status	0.0625	0.1900
Phylogenetic distance, habitat similarity, wetland status	0.2572**	0.0010

Discussion

Biogeography—Our temporal and spatial analyses (Figs. 5–6) suggest an origin of *Sabatia* in a wide area encompassing H (MCA), C (WGCP), and B (EGCP), during the late Middle Miocene (stem node CI: 8.16–14.29 Ma), followed by an early diversification in C (WGCP) and B (EGCP) during the early Late Miocene (crown node CI: 5.72–11.98 Ma). Major speciation events in the early history of the genus are likely allopatric or peripatric, involving vicariance between adjacent areas. During the inferred time frame, the southern part of North America was subject to repeated apparition/regression of numerous north to south fluvio-deltaic axes, including the Rio Bravo, Rio Grande, and Guadelupe river systems (Galloway et al. 2011), which could have acted as efficient barriers triggering the initial *Gyrandra* /*Sabatia* divergence. Alternatively, the currently Mexican-Central American genus *Gyrandra* could have adapted to mountain habitats already available in what would become the present Sierra Madre Occidental, while *Sabatia* could have diversified in lower and more arid lands

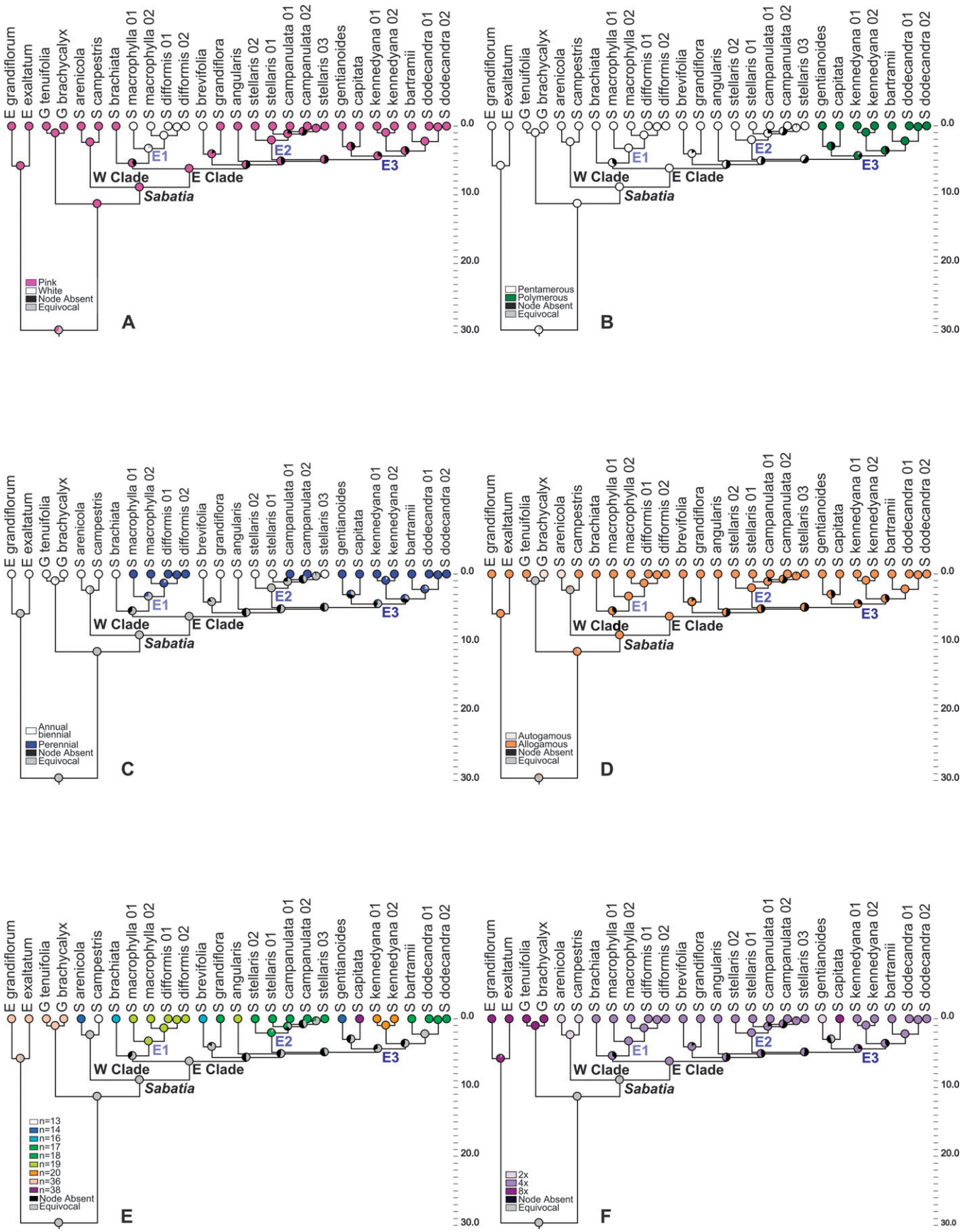

FIG. 7. Optimization of selected morphological and karyological characters onto the BEAST chromogram. A. Corolla color (0 = pink, 1 = white). B. Floral merosity (0 = pentamerous, 1 = polymerous). C. Plant longevity (0 = annual or biennial, 1 = perennial). D. Reproductive system (0 = autogamous, 1 = allogamous). E. Haploid chromosome number (0 = '*n* = 13'; 1 = '*n* = 14'; 2 = '*n* = 16'; 3 = '*n* = 17'; 4 = '*n* = 18'; 5 = '*n* = 19'; 6 = '*n* = 20'; 7 = '*n* = 36'; 8 = '*n* = 38'). F. Ploidy level (0 = 2x, 1 = 4x, 2 = 8x).

of the current Gulf Coastal Plain. Our spatial analyses indicate some support for niche conservatism in *Sabatia*, albeit not significant at the $p < 0.01$ level (Table 3).

The major split between the W and E clades of *Sabatia* during Middle-Late Miocene (CI: 5.72–11.98 Ma) is more difficult to explain in terms of paleotectonic events. Indeed, the formation of the Mississippi Embayment, which could have explained the current distribution of main *Sabatia* clades on both sides of the Mississippi River, dated back to the early Oligocene, well before the origin of the genus. The Middle-Late Miocene transition, however, was marked by important climate changes in North America, with the creation of a strong W-E zonal climate zone (Prothero 1998). Whereas the western Gulf Coastal Plain and the western interior remained hot and arid, subtropical to warm temperate conditions prevailed in the northeastern Gulf margin, with cooler conditions along the Appalachian trend (Galloway et al. 2011; Prothero 1998). This strong climatic frontier could explain the Middle-Late Miocene separation between a W clade, with more xerophytic elements (the succulent leaves observed in *S. arenicola* would indicate potential adaptation to xerophytic conditions), and an eastern clade rich in obligatory hydrophytes (USDA and NRCS 2014). Furthermore, our spatial analysis indicates that phylogenetically distant species (W clade vs. E clade, e.g.) tend to have different ecological niches, but this trend is not statistically supported at the $p < 0.01$ level. Genetic differentiation east and west of the Mississippi River has been found in other plant and animal groups, including in the plant genus *Spigelia* (Gould and Jansen 1999), and among populations of *Pinus taeda* (Al-Rabab'ah and Williams 2002), *Sarracenia alata* (Carstens and Satler 2013; Zellmer at al. 2012), and the lizard *Scincella lateralis* (Jackson and Austin 2010). However, while these studies lack absolute dates for divergence times, the divergences are all estimated to have originated more recently than *Sabatia*, mainly during the Pleistocene.

Finally, further diversification within the respective W and E clades mainly involves apparently sympatric speciation events of Late Miocene-Pliocene origin, with recent range expansion. Pliocene glacial/interglacial periods could have triggered range contraction/expansion in *Sabatia*, associated with polyploidy and aneuploidy events. Such pattern has been exemplified in North American groups of plants including *Primula* section *Aleuritia* or *Solidago* subsection *Humiles* (Guggisberg et al. 2006; Peirson et al. 2013). Further evidence for contraction and expansion of populations from glacial refugia has been found in the coastal plain plants *Trillium cuneatum* (Gonzalez et al. 2008) and the genus *Sarracenia* (Ellison et al. 2012).

Those recent (Pliocene) diversification events associated with several lines of evidence further support rapid speciation processes in the group. First, our molecular data have difficulty resolving relationships among the species, and branch lengths at the base of the eastern clade are short, in contrast to the longer branch separating the eastern and western clades. Second, Perry (1971) found high artificial crossability among *Sabatia* species, indicating that speciation has certainly taken place in the absence of intrinsic reproductive isolating barriers, despite the high degree of aneuploidy in the genus. This would suggest that species of *Sabatia* are still in the early stages of speciation and rely on other, possibly ecological or phenological factors, to maintain species boundaries. More research on the ecology of the sympatric species of *Sabatia* is needed to shed light on speciation mechanisms and maintenance of species boundaries.

Spatial Analysis—We hypothesized that closely related species of *Sabatia* would have overlapping distributions but not similar habitats compared to more distantly related taxa. This was not supported by Mantel tests, which showed there is a trend for closely related species to have both overlapping distributions and similar habitat usage. The significant positive correlations between species with overlapping distributions and habitat and wetland status similarity suggest that there is no resource partitioning within geographic areas among *Sabatia* species. Further, the three-way Mantel test results show that phylogenetic signal accounts for much of the relationship between geography and wetland status in *Sabatia* could be interpreted as phylogenetic niche conservatism.

Based on these results, divergence in *Sabatia* may have commonly taken place at the margins of ancestral species distributions, accompanied by aneuploid changes in chromosome number. This may help explain why more closely related species tend to overlap in at least part of their ranges. Looking more closely at sister groups resolved in the ML phylogenetic analysis, we see that sister species with geographic overlap typically only share a part of their range. For example, *S. dodecandra* overlaps with *S. bartramii* only along the southern Atlantic and eastern Gulf Coastal plains, whereas *S. bartramii* is primarily distributed in the panhandle of Florida and *S. dodecandra* extends into the northern Atlantic Coastal Plain. *Sabatia difformis* and *S. macrophylla* overlap on the Florida panhandle and southern Atlantic Coastal Plain, but *S. difformis* has a much more extensive distribution on the Atlantic Coastal Plain north to New Jersey and south to peninsular Florida.

Sister species with highly overlapping distributions, like *S. campanulata* and *S. stellaris*, tend to be isolated into distinct microhabitats (freshwater vs. brackish/sand dunes, respectively). Alternatively, *S. grandiflora* and *S. brevifolia* have largely overlapping distributions in peninsular Florida as well as similarity in microhabitat (wet savannahs and pine barrens), yet these two species are morphologically divergent (have different flower color) and have different flowering times (June–July vs. August–October; Perry 1971). Only one sister species pair shows complete geographic isolation, *S. capitata* (Appalachian Highlands) and *S. gentianoides* (Atlantic and Gulf Coastal Plains).

These observations therefore suggest that parapatric or sympatric speciation have been commonplace in *Sabatia* with lineage divergence perhaps occurring by ecological specialization along moisture or salinity gradients with diminished gene flow, then reinforced by chromosomal, morphological, and phenological changes. Perry's (1971) artificial crossing studies show high crossability among the younger aneuploid lineages of the eastern *Sabatia* clade, but sporadic to no seed production, depending on pollen parent, between the more ancient lineages of the west, *S. arenicola* and *S. campestris*.

Taxonomy and Character Evolution—Taxonomists have subdivided *Sabatia* into two sections or subgenera and five lower level taxa on the basis of morphology and artificial crossings (Blake 1915; Grisebach 1845; Perry 1971; Wilbur 1955). Our phylogenetic reconstructions are congruent with the lower level taxa (ranked at section or subsection) for the most part, although relationships among these taxa are mostly unresolved. One discrepancy is in the placement of *S. quadrangula* (one of two accessions), which groups with

members of subsection *Differmes* instead of *Angulares* (Wilbur 1955; Perry 1971; but see Blake 1915). *Sabatia quadrangula* is similar to other *Differmes* species (*S. differmis*, *S. macrophylla*) in having pentamerous flowers with white petals. White petals is a potential synapomorphy for the *Differmes* clade (clade E1) while separately derived in *S. breviflora*.

Analyses of the 26-taxa combined dataset consistently resolved a clade (though with little statistical support) containing all the polymerous taxa (clade E3), including members of *Dodecandrae* (*S. dodecandra*, *S. bartramii*, *S. kennedyana*) as well as *Pseudochironia* (*S. gentianoides*, *S. capitata*). This polymerous clade also included *S. angularis* (pentamerous) and *S. calycina* (facultatively polymerous) in the 30-taxon dataset. Perry (1971) placed *S. calycina* in subsection *Dodecandrae*, whereas Blake (1915) placed *S. calycina* in subsection *Campanulatae*. Our 30-taxon combined data analyses grouped one accession of *S. calycina* with the polymerous species and the other accession with *Campanulatae* (including *S. stellaris* and *S. campanulata*). Neither ITS nor cpDNA alone could resolve the positions of these accessions. These results point to a possible hybrid origin of *S. calycina*. In support of this hypothesis, *S. calycina* is polyploid (n = 32) and is one of the few species of *Sabatia* that is autogamous (the other being *S. arenicola*). In *S. arenicola*, autogamy may be related to better assurance for reproduction in its ephemeral, sand dune habitat, as plants in disturbance-prone habitats may be selected for self-fertilization (Eckert et al. 2010).

The most strongly supported discrepancy between the traditional classification of *Sabatia* and our analyses is the traditional primary subgeneric division into subgenus/section *Pseudochironia* (containing the two capitate species with large inflorescence bracts, *S. gentianoides* and *S. capitata*) and subgenus *Eusabatia*/section *Sabatia* (all other species). These two taxonomic groupings clearly do not have a basis in reciprocal monophyly. Rather, all molecular data show overwhelming support for an early lineage splitting event resulting in the western *Campestria* clade (clade W, including *S. campestris* and *S. arenicola*) and all the primarily eastern taxa (clade E). Perry (1971) noted that *S. campestris* and *S. arenicola* appear to represent a well-differentiated line of evolution in *Sabatia* because they are genetically distinct (having the lowest chromosome numbers, n = 13 and n = 14, respectively) and are geographically isolated from most other species in the genus.

Reconstruction of ancestral chromosome numbers in *Sabatia* is ambiguous, but there is an overall trend in increasing aneuploid numbers from the more ancestral to the more derived species. Moreover, species resolved as sister taxa have different chromosome numbers in all cases but one (*S. differmis* and *S. macrophylla* both have n = 19), indicating that speciation has been associated with changes in chromosome number.

Future Directions—Phylogenetic ambiguities in *Sabatia* could be resolved in the future with the addition of more samples of each taxon and full taxon sampling, including the western endemic *S. arkansana* and the disjunct Mexican population of *S. stellaris*. Better resolution could also be gained by sampling more characters, such as additional nuclear markers or targeted genome-wide sequencing to detect single-nucleotide polymorphisms. Additional population and genome sampling of problematic species (*S. calycina*, *S. quadrangula*) could also help disentangle potential reticulation events leading to their origin. Finally, careful fieldwork to establish fine-scale ecological niches for overlapping *Sabatia* species, measuring variables such as soil moisture content and salinity, could allow more accurate tests for niche conservatism vs. niche switching.

Acknowledgments. We thank Thomas Martin for assistance with interpretation of Mantel tests; Carolyn de Simone for lab assistance; Jim Allison, Kevin Fitch, and the staff of the University of South Alabama Herbarium for field assistance; Diane Ferguson and Jason Grant for collecting *Sabatia* species; James Pringle for guidance on *Sabatia* taxonomy; and Lena Struwe for supportive discussion and comments on an earlier version of this paper. This work was supported by a Western Carolina University Faculty Scholarship Award to K. G. M. and Highlands Biological Station funding to M. R.

Literature Cited

Adanson, M. 1763. *Familles de plantes*. Paris: Vincent.

Al-Rabab'ah, M. A. and C. G. Williams. 2002. Population dynamics of *Pinus taeda* L. based on nuclear microsatellites. *Forest Ecology and Management* 163: 263–271.

Avise, J. C. 2000. *Phylogeography - The history and formation of species*. Cambridge: Harvard University Press.

Blake, S. F. 1915. Notes on the genus *Sabatia. Rhodora* 17: 50–57.

Braun, E. L. 1955. The phytogeography of unglaciated eastern United States and its interpretation. *Botanical Review* 21: 297–375.

Carstens, B. C. and J. D. Satler. 2013. The carnivorous plant described as *Sarracenia alata* contains two cryptic species. *Biological Journal of the Linnean Society. Linnean Society of London* 109: 737–746.

Degner, J. F., D. M. Silva, T. D. Hether, J. M. Daza, and E. A. Hoffman. 2010. Fat frogs, mobile genes: Unexpected phylogeographic patterns for the ornate chorus frog (*Pseudacris ornata*). *Molecular Ecology* 19: 2501–2515.

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

Drummond, A. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

Eckert, C. G., S. Kalisz, M. A. Geber, R. Sargent, E. Elle, P.-O. Cheptou, C. Goodwillie, M. O. Johnston, J. K. Kelly, D. A. Moeller, E. Porcher, R. H. Ree, M. Vallejo-Marin, and A. A. Winn. 2010. Plant mating systems in a changing world. *Trends in Ecology & Evolution* 25.35-43

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

Ellison, A. M., E. D. Butler, E. J. Hicks, R. F. C. Naczi, P. J. Calie, C. D. Bell, and C. C. Davis. 2012. Phylogeny and biogeography of the carnivorous pant family Sarraceniaceae. *PLoS One* 7: e39291.

Farris, J. S., M. Kallersjö, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. *Cladistics* 10: 315–319.

Favre, A., Y.-M. Yuan, P. Kuepfer, and N. Alvarez. 2010. Phylogeny of subtribe Gentianinae (Gentianaceae): Biogeographic inferences despite limitations in temporal calibration points. *Taxon* 59: 1701–1711.

Forest, F. 2009. Calibrating the tree of life: Fossils, molecules and evolutionary timescales. *Annals of Botany* 104: 789–794.

Galloway, W. E., T. L. Whiteaker, and P. Ganey-Curry. 2011. History of Cenozoic North American drainage basin evolution, sediment yield, and accumulation in the Gulf of Mexico basin. *Geosphere* 7: 938–973.

Gonzales, E., J. L. Hamrick, and S.-M. Chang. 2008. Identification of glacial refugia in south-eastern North America by phylogeographical analyses of a forest understorey plant, *Trillium cuneatum*. *Journal of Biogeography* 35: 844-852.

Gould, K. R. and R. K. Jansen. 1999. Taxonomy and phylogeny of a Gulf Coast disjunct group of *Spigelia* (Loganiaceae *sensu lato*). *Lundellia* 2: 1–13.

Graham, A. 1984. *Lisianthius* pollen from the Eocene of Panama. *Annals of the Missouri Botanical Garden* 71: 987–993.

Gray, A. 1878. *Synoptical flora of North America*. New York: American Book Company.

Grisebach, A. H. R. 1839. Genera et species Gentianearum adjectis observationibus quibusdam phytogeographycis. Stuttgart and Tubingen: J. G. Cotta.

Grisebach, A. H. R. 1845. Gentianaceae. Pp. 39–141 in *Prodromus systematis naturalis regni vegetabilis* 9, ed. A. de Candolle. Paris: Fortin Masson Sociorum.

Guggisberg, A., G. Mansion, S. Kelso, and E. Conti. 2006. Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid-polyploid species complex of *Primula*. *The New Phytologist* 171: 617–632.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Jackson, N. J. and C. C. Austin. 2010. The combined effects of rivers and refugia generate extreme cryptic fragmentation within the common ground skink (*Scincella lateralis*). *Evolution* 64: 409–442.

Janssens, S. B., E. B. Knox, S. Huysmans, E. F. Smets, and V. Merckx. 2009. Rapid radiation of *Impatiens* (Balsaminaceae) during Pliocene and Pleistocene: Result of a global climate change. *Molecular Phylogenetics and Evolution* 52: 806–824.

Jiang, W., S.-Y. Chen, H. Wang, D.-Z. Li, and J. J. Wiens. 2014. Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution* 80: 308–318.

Kozak, K. H., A. Larson, R. M. Bonett, and L. J. Harmon. 2005. Phylogenetic analysis of ecomorphological divergence, community structure, and diversification rates in dusky salamanders (Plethodontidae: *Desmognathus*). *Evolution* 59: 2000–2016.

Maddison, W. P. and D. R. Maddison. 2009. Mesquite: A modular system for evolutionary analysis, v. 2.6. http://mesquiteproject.org.

Mansion, G. 2004. A new classification of the polyphyletic genus *Centaurium* Hill (Chironiinae, Gentianaceae): Description of the New World endemic *Zeltnera*, and reinstatement of *Gyrandra* Griseb. and *Schenkia* Griseb. *Taxon* 53: 719–740.

Mansion, G. and L. Struwe. 2004. Generic delimitation and phylogenetic relationships within the subtribe Chironiinae (Chironieae: Gentianaceae), with special reference to *Centaurium*: Evidence from nrDNA and cpDNA sequences. *Molecular Phylogenetics and Evolution* 32: 951–977.

Mansion, G. and L. Zeltner. 2004. Phylogenetic relationships within the new world endemic *Zeltnera* (Gentianaceae-Chironiinae) inferred from molecular and karyological data. *American Journal of Botany* 91: 2069–2086.

Mansion, G., L. Zeltner, and F. Bretagnolle. 2005. Phylogenetic patterns and polyploid evolution within the Mediterranean genus *Centaurium* (Gentianaceae-Chironieae). *Taxon* 54: 931–950.

Merckx, V. S. F., T., J. Kissling, H. Hentrich, S. B. Janssens, C. B. Mennes, C. D. Specht, and E. F. Smets. 2013. Phylogenetic relationships of the mycoheterotrophic genus *Voyria* and the implications for the biogeographic history of Gentianaceae. *American Journal of Botany* 100: 712–721.

Müller, K. 2004. PRAP-computation of Bremer support for large data sets. *Molecular Phylogenetics and Evolution* 31: 780–782.

Müller, K., D. Quandt, J. Müller, and C. Neinhuis. 2005. Phyde® - Phylogenetic Data Editor, v. 0.9971.

Nylander, J. 2004. MrModeltest, v. 2.1. Uppsala: Evolutionary Biology Centre.

Nylander, J. A. A., U. Olsson, P. Alstrom, and I. Sanmartin. 2008. Accounting for phylogenetic uncertainty in biogeography: A Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: *Turdus*). *Systematic Biology* 57: 257–268.

Peirson, J. A., C. W. Dick, and A. A. Reznicek. 2013. Phylogeography and polyploid evolution of North American goldenrods (*Solidago* subsect. *Humiles*, Asteraceae). *Journal of Biogeography* 40: 1887–1898.

Perry, J. D. 1971. Biosystematic studies in the North American genus *Sabatia* (Gentianaceae). *Rhodora* 73: 309–369.

Prothero, D. R. 1998. The chronostratigraphic, paleogeographic and paleoclimatic background to North American mammalian evolution. Pp. 9–36 in *Evolution of tertiary mammals of North America*, eds. C. Janis, K. M. Scott, and L. Jacobs. Cambridge: Cambridge University Press.

Pursh, F. 1814. *Sabatia*. Pp. 137–138 in *Flora Americae Septentrionalis* 1. London: White, Cochrane and Co.

Rambaut, A. 2012. FigTree. Tree Figure Drawing Tool, v. 1.4.0. Available from: http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut, A. and A. J. Drummond. 2007. Tracer, v. 1.5. Available from: http://tree.bio.ed.ac.uk/software/tracer/.

Ronquist, F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology* 46: 195–203.

Ronquist, F. and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Rosenberg, M. S. and C. D. Anderson. 2011. PASSaGE: Pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution* 2: 229–232.

Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small. 2005. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.

Shaw, J., E. B. Lickey, E. E. Schilling, and R. L. Small. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperms: The tortoise and the hare III. *American Journal of Botany* 94: 275–288.

Simmons, M. P. and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15: 4261–4293.

Sorrie, B. A. and A. S. Weakley. 2001. Coastal plain vascular plant endemics: Phytogeographic patterns. *Castanea* 66: 50–82.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

Swofford, D. L. 2002. PAUP* Phylogenetic analysis using parsimony (*and other methods), v. 4.0. Sunderland: Sinauer Associates.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

USDA and NRCS. 2014. The PLANTS database (http://plants.usda.gov, 28 March 2014). National Plant Data Team, Greensboro, Norht Carolina 27401–4901 U. S. A.

USGS. 2004. A tapestry of time and terrain (website). tapestry.usgs.gov. Last modified 14 December 2004. Accessed 10 March 2014.

von Hagen, K. B. and J. W. Kadereit. 2001. The phylogeny of *Gentianella* (Gentianaceae) and its colonization of the Southern hemisphere as revealed by nuclear and chloroplast DNA sequence variation. *Organisms, Diversity & Evolution* 1: 61–79.

Weins, J. J. and J. Tiu. 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PLoS One* 7: e42925, doi: 10.1371/journal.pone.0042925.

White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal DNA for phylogenetics. Pp. 315–322 in *PCR protocols: A guide to methods and applications*, eds. M. A. Innis, D. H. Gelfund, J. J. Sninsky, and T. J. White. San Diego: Academic Press.

Wilbur, R. L. 1955. A revision of the North American genus *Sabatia* (Gentianaceae). *Rhodora* 57: 1–104.

Wood, C. E. and R. E. Weaver. 1982. The genera of Gentianaceae in the southeastern United States. *Journal of the Arnold Arboretum* 63: 441–487.

Yu, Y., A. J. Harris, and X. He. 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): A tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution* 56: 848–850.

Zellmer, A. J., M. M. Hanes, S. M. Hird, and B. C. Carstens. 2012. Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. *Systematic Biology* 61: 763–777.

Appendix 1. Voucher specimens for tissue samples used in molecular phylogenetic analyses. Data entries are organized as follows: Species. – State: County: Collector and collection number (accession number if multiple accessions) (herbarium acronym), and GenBank accession numbers in the following order: *trnD-T*/ *atpH-I* / ITS1/ ITS2/ *trnL-F*/ *trnS-G-G*. A dash (–) indicates that the locus was not sequenced for that specimen. The range of GenBank numbers for all sequences included in the study are KP338623-KP338781.

Eustoma exaltatum*—**Mexico: *Zeltner 980528-1* (NEU) (MG98711) –/ –/ KP338693/ KP338724/ KP338752/ KP338778. ***E. grandiflorum – Horticultural origin, *Mansion 010834* (NEU) –/ –/ KP338694/ KP338725/ KP338753/ KP338779. ***Gyrandra brachycalyx*** – Mexico: Chiapas: *Mansion & Zeltner MG990205* KP338642/ KP338664/ KP338695/ KP338726/ KP338754/ KP338780. ***G. speciosa*** – Mexico: Jalisco: *Mansion & Zeltner 990228* KP338643/ KP338665/ KP338696/ KP338727/ KP338755/ KP338781. ***Sabatia angularis*** – Tennessee: Coffee Co.: *Mathews 335* (WCUH) KP338623/ KP338644/ KP338666/ KP338697/ KP338728/ KP338756; ***S. arenicola*** – Louisiana: Jefferson Parish: *Ferguson 1699* (WCUH, LSU) KP338626/ KP338645/ KP338668/ KP338699/ KP338729/ –. ***S. bartramii*** – Florida: Escambia Co.: *Mathews 343* (WCUH) –/ KP338646/ KP338669/ KP338700/ KP338730/ KP338759. ***S. brachiata*** – Tennessee: Franklin Co.: *Fitch 1452* (WCUH) –/ –/ –/ –/ KP338731/ KP338760. ***S. brevifolia*** – Alabama: Baldwin Co.: *Mathews 350* (WCUH) –/ KP338648/ KP338671/ KP338702/ KP338733/ KP338762. ***S. calycina* (1)** – Alabama: Baldwin Co.: *Mathews 345a* (WCUH) –/ –/ KP338673/ KP338704/ KP338735/ –; ***S. calycina* (2)** – Alabama: Baldwin Co.: *Mathews 344a* (WCUH) KP338628/ –/ KP338672/ KP338703/ KP338734/ KP338763. ***S. campanulata* (1)** – Alabama: Baldwin Co.:

Mathews 341c (WCUH) KP338630/ KP338651/ KP338675/ KP338706/ KP338737/ KP338765; ***S. campanulata*** **(2)** – Alabama: Baldwin Co.: *Mathews 341a2* (WCUH) KP338629/ KP338650/ KP338674/ KP338705/ KP338736/ KP338764. ***S. campestris*** – Texas: Waller Co.: *Mansion & Zeltner 97706* (NEU) KP338631/ KP338652/ KP338676/ KP338707/ KP338738/ KP338766. ***S. capitata*** – Tennessee: Marion Co.: *Mathews 346* (WCUH) KP338632/ KP338653/ KP338677/ KP338708/ KP338739/ KP338767. ***S. difformis*** **(1)** – New Jersey: Burlington Co.: *Abbott RS 1889* (MI) (MG02045) –/ –/ KP338678/ KP338709/ KP338740/ –; ***S. difformis*** **(2)** – New Jersey: Burlington Co.: *Palmer s. n.* (CHRB) KP338633/ KP338654/ KP338679/ KP338710/ KP338741/ KP338768. ***S. dodecandra*** **(1)** – South Carolina: Orangeburg Co.: *Mathews 360* (WCUH) –/ KP338655/ KP338680/ KP338711/ –/ –; ***S. dodecandra*** **(2)** – Maryland: Dorschester Co.: *J. R. Grant & W. Longbottom 97-02858* (NEU) (MG98509) KP338634/ KP338656/ KP338681/ KP338712/ KP338742/ KP338769. ***S. gentianoides*** – Alabama: Mobile Co.: *Mathews 338* (WCUH) KP338635/ KP338657/ KP338682/ KP338713/ KP338743/ KP338770. ***S. grandiflora*** – Florida: Hillborough Co: *Lakela 24094* (MI) (MG020406) KP338636/ –/ KP338683/ KP338714/ KP338744/ KP338771. ***S. kennedyana*** **(1)** – Massachusetts: Barnstable: *Neel & Pollini s. n.* (MG001023) KP338637/ KP338658/ KP338684/ KP338715/ –/ KP338772; ***S. kennedyana*** **(2)** – South Carolina: Horry Co.: *Wilbur 6895* (MI) (MG020408) –/ –/ KP338685/ KP338716/ KP338745/ –. ***S. macrophylla*** **(1)** – Alabama: Baldwin Co.: *Mathews 345* (WCUH) KP338638/ KP338659/ KP338686/ KP338717/ –/ KP338773; ***S. macrophylla*** **(2)** – Florida: Leon Co.: *Godfrey 84639* (MI) (MG020407) –/ –/ KP338687/ KP338718/ KP338746/ –. ***S. quadrangula*** **(1)** – Georgia: Newton Co.: *Mathews 348* (WCUH) –/ KP338661/ KP338689/ KP338720/ KP338748/ KP338775; ***S. quadrangula*** **(2)** – Georgia: Rockdale Co.: *Allison 1375* (WCUH) –/ KP338660/ KP338688/ KP338719/ KP338747/ KP338774. ***S. stellaris*** **(1)** – Louisiana: Jefferson Parish: *Ferguson 1690* (WCUH, LSU) KP338639/ KP338662/ KP338690/ KP338721/ KP338749/ –; ***S. stellaris*** **(2)** – Maryland: Dorschester Co.: *J. R. Grant & W. Longbottom 97-02871* (NEU) (MG98510) KP338641/ –/ KP338692/ KP338723/ KP338751/ KP338777; ***S. stellaris*** **(3)** – Alabama: Mobile Co.: *Mathews 340* (WCUH) KP338640/ KP338663/ KP338691/ KP338722/ KP338750/ KP338776.

Appendix 2.

Distribution matrix—

	S_campestris	*S_arenicola*	*S_kennedyana*	*S_bartramii*	*S_dodecandra*	*S_capitata*	*S_gentianoides*	*S_stellaris*	*S_campanulata*	*S_angularis*	*S_brevifolia*	*S_grandiflora*	*S_macrophylla*	*S_difformis*	*S_brachiata*
S_campestris		1	0	0	0	0	1	1	1	1	0	0	0	0	1
S_arenicola			0	0	0	0	1	1	1	1	0	0	0	0	1
S_kennedyana				0	1	0	1	1	1	1	1	0	1	1	1
S_bartramii					1	0	1	1	1	1	1	1	1	1	1
S_dodecandra						0	1	1	1	1	1	0	1	1	1
S_capitata							0	0	1	1	0	0	0	0	1
S_gentianoides								1	1	1	1	0	1	1	1
S_stellaris									1	1	1	1	1	1	1
S_campanulata										1	1	0	1	1	1
S_angularis											1	0	1	1	1
S_brevifolia												1	1	1	1
S_grandiflora													0	1	0
S_macrophylla														1	1
S_difformis															1
S_brachiata															

Microhabitat matrix—

	S_campestris	*S_arenicola*	*S_kennedyana*	*S_bartramii*	*S_dodecandra*	*S_capitata*	*S_gentianoides*	*S_stellaris*	*S_campanulata*	*S_angularis*	*S_brevifolia*	*S_grandiflora*	*S_macrophylla*	*S_difformis*	*S_brachiata*
S_campestris		1	0	0	0	0	0	0	0	1	0	0	0	0	1
S_arenicola			0	0	1	0	0	1	0	0	1	1	0	0	0
S_kennedyana				1	1	0	0	0	1	0	0	0	1	0	0
S_bartramii					1	0	1	0	1	0	1	1	1	1	0
S_dodecandra						0	0	1	1	0	0	1	1	0	0
S_capitata							0	0	0	1	0	0	0	0	1
S_gentianoides								0	1	0	1	1	1	1	0
S_stellaris									0	0	0	1	0	0	0
S_campanulata										0	1	1	1	1	0
S_angularis											0	0	0	0	1
S_brevifolia												1	1	1	0
S_grandiflora													1	1	0
S_macrophylla														1	0
S_difformis															0
S_brachiata															

(Continued)

Appendix 2. (Continued).

Wetland status matrix—

	S_campestris	*S_arenicola*	*S_kennedyana*	*S_bartramii*	*S_dodecandra*	*S_capitata*	*S_gentianoides*	*S_stellaris*	*S_campanulata*	*S_angularis*	*S_brevifolia*	*S_grandiflora*	*S_macrophylla*	*S_difformis*	*S_brachiata*
S_campestris		0	0	0	0	1	0	0	0	0	0	0	0	0	1
S_arenicola			0	0	0	0	0	0	1	0	1	1	1	0	0
S_kennedyana				1	1	0	1	1	0	0	0	0	0	1	0
S_bartramii					1	0	1	1	0	0	0	0	0	1	0
S_dodecandra						0	1	1	0	0	0	0	0	1	0
S_capitata							0	0	0	0	0	0	0	0	1
S_gentianoides								1	0	0	0	0	0	1	0
S_stellaris									0	0	0	0	0	1	0
S_campanulata										0	1	1	1	0	0
S_angularis											0	0	0	0	0
S_brevifolia												1	1	0	0
S_grandiflora													1	0	0
S_macrophylla														0	0
S_difformis															0
S_brachiata															

Appendix 3.

Coding of six morphological and karyological characters used in this study. Numbers zero to eight correspond to the respective character states. Character 1, corolla color: 0 = pink, 1 = white. Character 2, floral merosity: 0 = pentamerous, 1 = polymerous. Character 3, plant longevity: 0 = annual or biennial, 1 = perennial. Character 4, reproductive system: 0 = autogamous, 1 = allogamous. Character 5, haploid chromosome number: 0 = 'n = 13'; 1 = 'n = 14'; 2 = 'n = 16'; 3 = 'n = 17'; 4 = 'n = 18'; 5 = 'n = 19'; 6 = 'n = 20'; 7 = 'n = 36'; 8 = 'n = 38'. Character 6, ploidy level: 0 = 2x, 1 = 4x, 2 = 8x.

Species / Characters	1	2	3	4	5	6
Eustoma grandiflorum	0	0	0	1	7	2
Eustoma exaltatum	0	0	0	1	7	2
Gyrandra tenuifolia	0	0	0	1	7	2
Gyrandra brachycalyx	0	0	0	0	7	2
Sabatia angularis	0	0	0	1	5	1
Sabatia arenicola	0	0	0	0	1	0
Sabatia bartramii	0	1	1	1	4	1
Sabatia brachiata	0	0	0	1	2	1
Sabatia brevifolia	1	0	0	1	2	1
Sabatia campanulata	0	0	1	1	3	1
Sabatia campestris	0	0	0	1	0	0
Sabatia difformis	1	0	1	1	5	1
Sabatia dodecandra	0	1	1	1	3	1
Sabatia gentianoides	0	1	1	1	1	0
Sabatia capitata	0	1	1	1	8	2
Sabatia grandiflora	0	0	0	1	4	1
Sabatia kennedyana	0	1	1	1	6	1
Sabatia macrophylla	1	0	1	1	5	1
Sabatia stellaris	0	0	0	1	4	1