Chapter

THE ECOLOGY, ASSEMBLY AND EVOLUTION OF GYPSOPHILE FLORAS

Michael J. Moore[1]*, Juan F. Mota[2], Norman A. Douglas[1], Hilda Flores Olvera[3] and Helga Ochoterena[3]

[1]Department of Biology, Oberlin College, Oberlin, OH, USA
[2]Departamento de Biología y Geología, Universidad de Almería, Almería, España
[3]Instituto de Biología, Universidad Nacional Autónoma de México, México, DF, México

ABSTRACT

Gypsum ($CaSO_4 \cdot 2H_2O$) exposures and gypseous soils occupy over 100 million ha worldwide, primarily in arid and semiarid regions, with particularly large areas of surface gypsum in southwestern Asia, the Mediterranean region, the Horn of Africa and southwestern North America. Each of these areas hosts a diverse assemblage of gypsum endemic plant taxa, known as gypsophiles. Although plant biologists have been interested in the causes of gypsophily for well over a century, it has only been over the past few decades that gypsophile floras have received sustained ecological and evolutionary study. Recent work, principally in Spain, has revealed that both physical (*e.g.*, gypsum crusts, soil porosity) and chemical (*e.g.*, high Ca and S, low cation exchange capacity) factors may control community structure on highly gypseous substrates. Plant-fungal interactions may also play a key role in plant establishment on gypsum, although few studies have examined this subject. Molecular systematic and population genetic studies over the past two decades have revealed several key similarities in the assembly and evolution of gypsophile floras and taxa. These studies imply that gypsophile lineages have frequently appeared multiple times within clades that are ancestrally tolerant of gypsum, that speciation has been common in the most widespread lineages of gypsophiles, and that most gypsophile lineages first appeared no earlier than the latest Miocene. Population genetic studies have revealed generally higher levels of among-population genetic differentiation and isolation-by-distance within gypsophile taxa, in line with expectations for taxa that are restricted to substrate archipelagoes such as gypsum. Despite these advances in our understanding of gypsophily, gypsum floras remain much more poorly studied compared to other important

* Corresponding author: michael.moore@oberlin.edu

edaphic endemic communities, such as serpentine and halophilic floras, highlighting the need for additional work.

INTRODUCTION

Surface gypsum ($CaSO_4 \cdot 2H_2O$) deposits and gypsisols occur worldwide in arid and semi-arid regions, covering 100-207 million ha worldwide (Eswaran & Gong, 1991; Herrero, 2004; Herrero & Porta, 2000; Verheye & Boyadgiev, 1997). For example, large areas of exposed gypsum characterize parts of the Horn of Africa region (*e.g.*, Ethiopia, Somalia), North Africa (*e.g.*, Tunisia, Algeria), western Asia (*e.g.*, Iran, Iraq, Turkey), Australia, eastern Spain, and the Chihuahuan Desert region of North America (Escudero et al., 2014; Food and Agriculture Organization, 1998; Merlo et al., 2011). In contrast to most NaCl-rich soils, which are mainly concentrated along sea shores or less commonly in interior deserts and endorheic basins (Merlo et al., 2011), gypsum bedrock exists primarily in interior deposits and is derived from ancient, shallow hypersaline lagoons (Mota et al., 2011). Gypsum may also form in hot springs from volcanic vapors (Herrero et al., 2009) and can form pedogenically (Eswaran & Gong, 1991). Although subsurface gypsum deposits occur worldwide, the high solubility of gypsum means that it persists at the surface for evolutionarily meaningful times almost exclusively in arid and semiarid regions (Escudero et al., 2014; Parsons, 1976).

Gypsum outcrops can be relatively pure or may be combined with other salts, such as sodium chloride. Because of its high solubility, bedrock gypsum often becomes intermixed with surrounding soils, creating mosaics of soils with differing gypsum contents. Gypsum soils (or gypsisols) are characterized by gypsum contents > 5% and the presence of a gypsic horizon in which gypsum is accumulated (Food and Agriculture Organization, 1990). Gypsum outcrops can have different physical characteristics, as they can be exposed as massive gypsum evaporite bedrock, crystalline selenite, anhydrite, secondary evaporites or even sand dunes (Figure 1). Physical surface crusts commonly contain > 25% gypsum (Verheye & Boyadgiev, 1997).

Plants living on gypsum soils show varying degrees of fidelity to gypsum and employ a variety of survival strategies, both of which have been used as bases for ecological classification (*e.g.*, Davis et al., 1986; Duvigneaud & Denaeyer-de Smet, 1968). The vegetation of gypsum soils includes substrate generalist taxa that grow on and off of gypsum, taxa that grow mostly on gypsum, and taxa that are endemic to gypsum. In recent literature, these three groups of taxa have generally been referred to as gypsovags, gypsoclines, and gypsophiles, respectively (Meyer, 1986), although it is important to note that in older literature the word gypsophile had a much more variable meaning, often referring to any species commonly encountered on gypsum, regardless of its overall fidelity to the substrate (*e.g.*, Johnston, 1941; Powell & Turner, 1977). We follow Meyer's definitions for the purposes of this chapter.

Not coincidentally, the regions with the most extensive gypsum outcrops host the largest assemblages of gypsophiles. Particularly species-rich gypsophile floras exist in the Chihuahuan Desert (at least 200 species; *e.g.*, Johnston, 1941; Moore & Jansen, 2007; Powell & Turner, 1977), Somalia and Ethiopia (at least 50 species; Thulin, 1993; 1995; 1999; 2006), Turkey (at least 40 species; *e.g.*, Akpulat & Celik, 2005), and Spain (at least 40 species; Mota et al., 2009; 2011), with smaller gypsophile floras in Iran (*e.g.*, Akhani, 2004), North Africa (Le Houérou, 1969), Australia (Symon, 2007), the Mojave Desert and Intermountain West of the United

States (*e.g.*, Forbis de Queiroz et al., 2012; Meyer, 1986), Cyprus (Hadjikyriakou & Hand, 2011), and Yemen (Petrusson & Thulin, 1996).

Figure 1. Different physical characteristics of gypsum outcrops, as encountered in the Chihuahuan Desert of northern Mexico: A) rocks (Sierra Tlahualilo, Durango); B) crystals (Puerto de Lobos, Chihuahua); C) crusts (Sierra Roque, Chihuahua); D) sand dunes (Bolsón de Cuatro Ciénegas, Coahuila).

With the exception of the gypsum flora of Spain (Mota et al., 2011), gypsophiles have been poorly studied in most areas of the world, especially compared to serpentine and halophilic vegetation. For example, in most of the above regions, but particularly in western Asia and the Horn of Africa, gypsum habitats have been underexplored botanically, and it is likely that many more gypsophile taxa remain to be discovered and described. Even in the relatively well-botanized gypsum areas of Spain and the United States, more than a dozen new gypsophile taxa have been described in the past decade (*e.g.*, Atwood & Welsh, 2005; Erben & Arán, 2005; Sivinski & Howard, 2011). While great strides have been made in understanding the physiological and community ecology of gypsophile floras in Spain over the past 20 years, little or no corresponding research has been conducted in other gypsum environments, many of which have much different climates and/or rainfall regimes compared to Spain. Furthermore, it has only been over the 15 years that researchers have begun to assess the phylogenetic and population-level histories of gypsophiles. The present chapter reviews the current state of knowledge for gypsophile ecology and evolution, and identifies areas where additional research is needed to understand this globally important edaphic community.

GYPSOPHILE ECOLOGY

For well over a century, plant biologists have sought to understand the ecological controls on gypsum plant communities (*e.g.*, Contejan, 1881; Macchiati, 1888). Historically, ecologists have focused on physical (*e.g.*, Johnston, 1941; Meyer, 1986) and/or chemical (*e.g.*, Boukhris & Lossaint, 1970; Duvigneaud & Denaeyer-de Smet, 1968) causes for gypsophily, although more recent debates on the assembly of gypsophile floras have attempted to discriminate between two reference models: the specialist and refuge models (Escudero et al., 2014; Merlo et al., 1998; Palacio et al., 2007). These models closely link ecology with evolutionary processes and apply to other unusual geological substrates such as serpentine as well (Harrison & Rajakaruna, 2011). Below we review current understanding of the physical and chemical aspects of gypsophile ecology, as well as fungal-plant interactions.

Physical Soil Factors

Several physical characteristics have been posited to influence plant growth in soils with high gypsum content, including soil crusts, density and porosity, and associated phenomena such as water holding capacity. Among these factors, the crust that characterizes most gypsum soils has received perhaps the most attention as a physical soil attribute controlling gypsum endemism (*e.g.*, Cañadas et al., 2013; Romão & Escudero, 2005). In arid soils in general, traditionally two types of soil crusts, physical and biological, have been differentiated (but see Gil de Carrasco & Ramos, 2011). Reprecipitation of gypsum creates a physical crust in gypsum soils that contributes significantly to the formation of the structure of gypsic horizons (Daniells, 2012). However, gypsum soils are also frequently characterized by cryptogamic crusts that also influence soil chemistry and texture (Anderson et al., 1982; for more on cryptogamic crusts, see the section below). Hence it is not easy to separate the relative effects of physical vs. biotic crusts on germination and seedling establishment in gypsum environments. This is a clear example of how difficult it can be to separate the chemical, physical and biological factors when explaining gypsophily. Moreover, physical crusts are not exclusive to gypsum (*e.g.*, Anderson et al., 1982). A search in SCOPUS (10 April 2014) using the terms "soil crusts" and "arid" produced 388 results. Of these, only 39 included the word "gypsum."

While soil crusts have received the bulk of attention from ecologists, the hard upper soil horizons (gypsic and petrogypsic; Herrero & Porta, 2000) in highly gypseous soils also likely influence community composition. The gypsum content of soils influences porosity and root penetration capacity (Poch, 1998). Although gypsisols contain > 5% gypsum, much higher amounts of gypsum tend to characterize gypsophile floras. For example, Salmerón et al. (2014) found an average gypsum content approaching 60% in soils dominated by the gypsocline *Jurinea pinnata* in Spain, which qualifies such soils as hypergypsics (Herrero, 2004). Unfortunately, as Drohan & Merkler (2009) have noted, gypsum content of gypseous soils is rarely provided in most studies. Although there are not many field data, those that are available show that gypsum is a difficult environment for plant roots (Guerrero-Campo et al., 1999). Several studies, mainly of cultivated plants, have noted that gypsum contents > 25% hinder root development (*e.g.*, Boyadgiev, 1974; Mashali, 1996). Poch (1998) found that roots are seldom found in horizons with gypsum content > 60%, and when that percentage exceeds 80%,

roots only grow through preexisting cracks or faunal channels and will otherwise form a mat above the upper boundary of these horizons. Poch (1998) also notes that gypsum soil pores may be irregular and discontinuous, which would seriously affect root development in plants whose roots are concentrated in shallow, highly gypseous horizons. Poch & Verplancke (1997) showed that gypsum content was positively correlated with penetration resistance, although they note that this alone does not explain the poor growth of roots in hypergypsic soils. Furthermore, the resistance of soil to root penetration may be increased upon drying, which may help explain why the effect of gypsum on plants appears much greater in arid and semiarid climates. Gibbens & Lenz (2001) reported that petrogypsid soils in the Chihuahuan Desert restricted rooting depth of shrubs to less than 1 m and thus contributed to vegetative sparseness. Nevertheless, some gypsophiles have been found to possess relatively deep roots (*e.g.*, Mota et al., 2011), and hence the effects of gypsum content on root penetration may not be universal.

Water holding capacity of gypsum soils is also likely to influence gypsum floras. However, here too, the data are contradictory. Several authors claim that gypsum soils have lower water holding capacity (*e.g.*, Meyer & García-Moya, 1989), whereas others have suggested the opposite (Hiouani, 2006). It has also been observed that gypsum soils are moist at depth even when surrounding soils dry completely (Meyer & García-Moya, 1989), and according to Hiouani (2006), moisture tends to increase as the percentage of gypsum increases. These apparent contradictions may be related to the irregular distribution of water in these soils, especially when their gypsum contents are very high (Food and Agriculture Organization, 1990). In these cases the pores in the gypsum may become plugged by the precipitation of leached gypsum (Poch, 1998), which may cause high mortality in the fine roots and limit their performance. Precipitation of gypsum and calcium carbonate around roots has also been reported to occur as a consequence of high calcium concentrations in the rhizosphere (Hinsinger et al., 2009). Further investigation of gypsum particle size and micromorphology, including their influence on soil matric potential, may reveal additional influences on community structure in gypsum soils.

Soil Chemistry

Although much of the ecological research into gypsophily recognizes that physical and chemical constraints may exist, chemical factors have largely been treated as secondary and have therefore been underexplored (Escudero et al., 2014; Romão & Escudero, 2005). The fact that many gypsophiles, and particularly those that are regionally dominant, seem to be characterized by certain nutritional or chemical profiles, strongly suggests that unusual soil chemistry of gypseous substrates has influenced the evolution of such taxa. Below we summarize the chemical aspects of gypsum soils that plants typically must contend with, with a focus on how gypsophiles deal with excess levels of calcium and sulfur.

In general, gypsum soils are characterized by alkaline pH, high content of carbonates, the dominance of Ca and Mg ions, low NaCl, and above all, reduced fertility (Salmerón-Sánchez et al., 2014). The pH of the gypsum soils varies between slightly and moderately alkaline (Drohan & Merkler, 2009) and is not very different from other calcareous soils (Salmerón-Sánchez et al., 2014). Gypsum does not significantly increase osmotic potential despite its high contents of certain salts and ions (Herrero et al., 2009). Electrical conductivity of these soils is usually below 3 dS m^{-1} (*e.g.*, Herrero et al., 2009; Salmerón-Sánchez et al., 2014). For Spanish

gypsum soils, Gil de Carrasco & Ramos (2011) provide an average value of 2.76 dS m^{-1} and Herrero et al. (2009) provide a value of 2.25 dS m^{-1}.

Gypsum soils are characterized by their lowered fertility. Highly gypseous soils have very little organic matter and a low cation exchange capacity (CEC). The high pH and high concentrations of Ca promote rapid insolubilization of nutrients released by weathering (Gil de Carrasco & Ramos, 2011), and reduce the availability of key macro- and micronutrients such as Fe, K, Mg, Mn, P and Zn (Boscaiu et al., 2013; Oyonarte et al., 2002).

Plants growing in high Ca environments must also deal with the cytotoxicity of this element. Although Ca is an essential element for numerous biological functions, it is toxic at high concentrations in the cytoplasm (Hawkesford et al., 2012). Physiological mechanisms, such as sequestering Ca within cells or in the apoplast via oxalate crystallization, allow plants growing on calcium-rich soils to avoid this toxicity (*e.g.*, Fink, 1991; Franceschi & Nakata, 2005). Plants tolerant of gypsum soils pose no exception. In their study of gypsovags from White Sands, New Mexico, USA, Borer et al. (2012) found that plants have different strategies that allow them to cope with the Ca excess, including the prevention of Ca uptake, the sequestration of foliar Ca in chemically unavailable forms (calcium oxalate), and the maintenance of foliar Ca in labile forms, which may allow it to be excreted from foliar salt glands. These mechanisms largely coincide with the four strategies that allow plants to deal with excess Ca and S found by Duvigneaud & Denaeyer-de Smet (1968; 1973) and Merlo et al. (1998; 2001) among plants growing on gypsum in Spain: the accumulator, the extruder, the assimilator, and the avoider. The first group includes species that accumulate large amounts of Ca, and often S and Mg; slight foliar succulence is characteristic of many of these plants (*e.g.*, *Gypsophila*, *Ononis tridentata*). The extruders contain species from primarily halophilic lineages that possess secretory glands, including *Frankenia* and some *Limonium* (Kleinkopf & Wallace, 1974). The assimilators include groups with S-rich secondary metabolites, including the many taxa of Brassicales (*e.g.*, the families Brassicaceae, Capparaceae, and Resedaceae) that are found on gypsum around the world (see below), all of which may be physiologically preadapted to gypsum. The avoiders are able to finely control ionic import and hence are able to survive on very poor and oligotrophic soils; Duvigneaud & Denaeyer-de Smet (1968) note that most avoider taxa on gypsum are gypsovags.

Since the seminal work by Duvigneaud & Denaeyer-de Smet (1966), gypsophiles have been viewed in general as Ca, Mg and S accumulators. This pattern is evident in Table 1, which summarizes foliar nutrient concentrations for various gypsophiles, gypsoclines, and gypsovags. Values for Ca concentration in the leaves of most plants typically range from 0.5-2.5% (Jones, 2012; Kalra, 1997; Parsons, 1976). Among Spanish gypsophiles, highly elevated levels of Ca (> 5%) have been found in *Gypsophila struthium*, *G. hispanica*, *Ononis tridentata*, *Frankenia thymifolia*, and *Sedum gypsicola* (Table 1). All of these species have slightly succulent leaves (Merlo et al., 1998; 2001), with the exception of *F. thymifolia*, which is an extruder. Another group of Iberian gypsophiles also exhibit above average values (> 3%) of foliar Ca: *Helianthemum squamatum*, *Lepidium subulatum*, *Herniaria fruticosa*, *Coris hispanica*, and *Santolina viscosa* (Table 1). Two Spanish gypsovags, *Helianthemum syriacum* and *Sedum sediforme*, also have Ca levels above 3%, whereas locally endemic gypsophiles such as *Centaurea hyssopifolia*, *Thymus lacaitae* or *Teucrium turredanum* have lower values (Table 1).

The widely distributed Spanish gypsocline *Jurinea pinnata*, which grows on both gypsum and dolomite, also has relatively low levels of Ca (2.6%), although those values are higher on

gypsum than on dolomites (Table 1). This behavior is very similar to that exhibited by the Iberian gypsovags *Rosmarinus officinalis*, *Linum suffruticosum* or *Salvia lavandulifolia* (Palacio et al., 2007). Although little nutrient data from other gypsophile floras are available, Ca concentrations above 5% were found in the Tunisian gypsoclines *Erodium glaucophyllum*, *Zygophyllum album,* and *Moricandia suffruticosa* (Boukhris & Lossaint, 1970; 1972). No data are available for the large and diverse gypsophile flora of the Chihuahuan Desert region, although almost all regionally dominant gypsophiles in that area have slightly succulent leaves (*e.g*. gypsophile species of *Dicranocarpus, Sartwellia, Acleisanthes, Nama, Tiquilia*, and *Nerisyrenia*), suggesting that these taxa are also likely accumulators.

Many of the gypsophile taxa with elevated Ca concentration in Table 1 also possess elevated S and Mg concentration, although the pattern is less consistent for Mg. For example, the Ca accumulators *Gypsophila struthium*, *G. hispanica*, and *Ononis tridentata* possess the highest known S contents of any plants growing on gypsum, and have elevated Mg concentrations as well (Table 1). Other Spanish gypsophiles, such as *Helianthemum squamatum* and *Lepidium subulatum*, have elevated S but much lower Mg concentrations. In contrast, the narrowly distributed gypsophile *Helianthemum conquense* has relatively low foliar concentrations of Ca, Mg, and S (Table 1).

Palacio et al. (2007) suggest that there are two broad categories of gypsophile species: those that are dominant on gypsum and broadly distributed geographically (the regionally dominant gypsophiles) and those that are narrowly distributed. The former group is composed of taxa that are typically succulent-leaved and often show a remarkable ability to accumulate Ca, Mg, and S, as well as the macronutrients that are scarce in gypsum soils such as N and P (Table 1). Whereas many narrowly distributed gypsophiles like *Helianthemum conquense* seem to behave more like gypsovags in terms of nutrient accumulation, some locally distributed gypsophiles such as *Coris hispanica* and *Santolina viscosa* behave similarly to the "stockpiling" regional dominants (Table 1). Even the gypsovag *Helianthemum syriacum* is difficult to separate from the latter two species based on Ca concentration (Table 1). Although not perfect, the relatively strong correlation between regional dominance, foliar succulence, and the strategy of accumulating certain nutrients suggests a syndrome of common adaptations to gypsum soil chemistry, which deserves much further physiological and ecological study. Indeed, as Merlo et al. (2011) have noted, foliar Ca, Mg, and S concentration, as well as Ca:Mg ratio, seem to be useful parameters for establishing differences in the nutritional behavior of plants growing on gypsum, dolomite, and serpentine.

Table 1. Community characteristics, succulence, and foliar nutrient content for selected gypsophiles, gypsoclines, and gypsovags. All nutrient values are mean percentages; dashes indicate that values were not available. Taxa in bold are gypsophiles; all other taxa are gypsovags, except for the gypsocline *Jurinea pinnata*. Average values for halophytes are provided at the bottom of the table. Key to references: (1) Drohan & Merkler (2009); (2) Duvigneaud & Denaeyer de Smet (1966); (3) Duvigneaud & Denaeyer de Smet (1968); (4) Escudero et al. (2014); (5) M. Merlo et al. (unpublished); (6) Salmerón-Sánchez et al. (2014)

Species	Population growing on gypsum?	Taxon dominant on gypsum?	Taxon widespread on gypsum?	Succulent?	Ca	Mg	S	Na	N	P	K	References
Arctomecon californica	yes	?	no	yes	3.83	2.47	0.33	0.29	–	0.06	1.88	[1]
Artemisia herba-alba	yes	no	yes	no	1.20	0.33	0.20	0.04	3.94	0.27	1.45	[4]
Centaurea hyssopifolia	yes	no	no	no	2.60	0.49	0.80	0.06	4.02	0.27	2.33	[4]
Coris hispanica	yes	no	no	no	3.72	0.13	–	< 0.10	1.65	0.02	0.68	[5]
Eriogonum corymbosum	yes	?	no	no	0.84	2.47	0.31	0.18	–	0.08	2.27	[1]
Frankenia thymifolia	yes	yes	yes	no	11.00	1.22	1.15	0.13	2.00	0.10	1.80	[2,3]
Frankenia thymifolia	yes	yes	yes	no	10.66	0.91	–	0.10	1.50	0.04	0.37	[5]
Gypsophila hispanica	yes	yes	yes	yes	7.83	2.23	4.99	0.03	1.75	0.10	0.93	[2,3]
Gypsophila hispanica	yes	yes	yes	yes	7.40	1.21	3.00	0.06	2.49	0.19	1.18	[4]
Gypsophila struthium	yes	yes	yes	yes	6.13	3.94	3.64	1.33	1.55	1.26	0.68	[2,3]
Gypsophila struthium	yes	yes	yes	yes	8.17	0.83	–	< 0.10	1.26	0.08	0.80	[5]
Helianthemum alypoides	yes	yes	no	no	1.83	0.25	–	< 0.10	1.08	0.07	0.28	[5]
Helianthemum conquense	yes	no	yes	no	1.90	0.26	0.10	0.03	1.68	0.11	0.39	[4]

Table 1. (Continued)

Species	Population growing on gypsum?	Taxon dominant on gypsum?	Taxon widespread on gypsum?	Succulent?	Ca	Mg	S	Na	N	P	K	References
Helianthemum squamatum	yes	yes	yes	yes	3.43	0.65	2.90	0.08	1.65	0.12	0.62	[4]
Helianthemum squamatum	yes	yes	yes	yes	3.15	0.78	2.48	0.08	1.37	0.09	0.75	[2,3]
Helianthemum squamatum	yes	yes	yes	yes	2.62	0.42	–	< 0.10	1.14	0.05	0.28	[5]
Helianthemum syriacum	yes	yes	yes	no	3.10	0.50	1.30	0.02	1.10	0.07	0.70	[2,3]
Helianthemum syriacum	yes	yes	yes	no	3.00	0.31	1.00	0.02	1.76	0.11	0.50	[4]
Helianthemum syriacum	yes	yes	yes	no	3.02	0.20	–	< 0.10	1.30	0.08	0.36	[5]
Herniaria fruticosa	yes	no	yes	no	2.90	0.77	1.10	0.05	2.53	0.11	0.92	[4]
Herniaria fruticosa	yes	no	yes	no	3.00	1.30	0.81	0.01	1.00	0.04	0.89	[2,3]
Jurinea pinnata (on dolomite)	no	no	yes	no	2.20	0.71	0.35	0.04	1.47	0.06	1.05	[6]
Jurinea pinnata (on gypsum)	yes	no	yes	no	2.62	0.46	0.51	0.06	2.12	0.04	1.37	[6]
Lepidium subulatum	yes	yes	yes	no	1.80	0.46	2.80	0.02	3.20	0.16	1.40	[2,3]
Lepidium subulatum	yes	yes	yes	no	2.70	0.38	2.30	0.06	5.12	0.25	0.97	[4]
Lepidium subulatum	yes	yes	yes	no	1.83	0.11	–	< 0.10	2.12	0.08	0.36	[5]
Linum suffruticosum	yes	no	yes	no	2.65	2.45	0.06	0.06	2.80	0.17	0.92	[4]
Linum suffruticosum	no	no	yes	no	2.70	0.33	0.08	0.06	2.31	0.14	0.73	[4]
Ononis tridentata	yes	yes	yes	yes	5.57	2.52	6.07	0.03	2.31	0.10	0.68	[2,3]

Species	Population growing on gypsum?	Taxon dominant on gypsum?	Taxon widespread on gypsum?	Succulent?	Ca	Mg	S	Na	N	P	K	References
Ononis tridentata	yes	yes	yes	yes	5.75	1.84	4.50	0.09	2.42	0.12	0.28	[4]
Ononis tridentata	yes	yes	yes	yes	4.37	0.86	–	0.20	1.31	0.04	0.24	[5]
Rosmarinus officinalis	yes	no	yes	no	1.15	0.25	0.22	0.04	0.97	0.06	1.29	[2,3]
Rosmarinus officinalis	yes	no	yes	no	1.40	0.28	0.10	0.06	1.09	0.07	0.80	[4]
Rosmarinus officinalis	no	no	yes	no	1.20	0.17	0.10	0.06	1.17	0.09	1.01	[4]
Salvia lavandulifolia	yes	no	yes	no	1.95	0.33	0.10	0.05	1.77	0.10	0.52	[4]
Salvia lavandulifolia	no	no	yes	no	1.50	0.30	0.05	0.05	1.51	0.09	0.58	[4]
Santolina viscosa	yes	no	no	no	3.01	0.12	–	0.29	1.34	0.06	0.60	[5]
Sedum gypsicola	yes	no	yes	yes	8.18	0.18	–	< 0.10	2.76	0.05	0.49	[5]
Sedum sediforme	yes	no	yes	yes	4.31	0.10	–	< 0.10	0.69	0.04	0.55	[5]
Teucrium capitatum	yes	no	yes	no	1.90	0.24	0.06	0.04	2.65	0.13	0.72	[4]
Teucrium capitatum	no	no	yes	no	1.80	0.26	0.05	0.05	2.30	0.15	0.63	[4]
Teucrium polium	yes	no	yes	no	2.00	0.61	0.60	0.07	1.67	0.07	0.76	[4]
Teucrium turredanum	yes	yes	no	no	1.37	0.22	–	< 0.10	0.99	0.03	0.60	[5]
Thymus lacaitae	yes	no	no	no	1.60	0.40	0.04	0.05	1.42	0.11	0.56	[4]
halophytes (several species)	no	no	yes	yes	1.10	1.64	2.36	9.37	2.28	0.20	1.98	[2,3]

Fungal-Plant Interactions

Mycorrhizal and endophytic fungal interactions with gypsophiles are poorly understood but may play an important role in structuring gypsophile plant communities. A handful of recent studies have begun to shed light on the community composition of these fungi in gypsum environments. In Spain, Alguacil et al. (2009a; b; 2012) have found an unusually diverse arbuscular mycorrhizal fungus (AMF) community on Spanish gypsum, comparable to that found on non-gypseous sites with much higher plant density. A total of 21 AMF types were found in association with four Spanish gypsophiles: *Gypsophila struthium*, *Teucrium libanitis*, *Helianthemum squamatum*, and *Ononis tridentata* (Alguacil et al., 2009b). As Alguacil et al. (2009a) note, this appears to be the first report of AMF in the genus *Gypsophila* (Wang & Qiu, 2006). Moreover, Alguacil et al. (2009a) found novel AMF sequences among roots of *G. struthium*, suggesting the presence of undescribed species. This new fungal type was found mainly in the less altered gypsum zone, raising the possibility that it could be associated with survival or proliferation of *G. struthium* on gypsum, which could be among the factors underlying the great colonizing power that this species exhibits in abandoned gypsum quarries, where it becomes almost monospecific (Mota et al., 2004). Alguacil et al. (2012) found a higher diversity of AMF in perennial gypsophiles and gypsovags vs. an annual gypsovag, and Porras-Alfaro et al. (2014) also found that regionally dominant gypsophile taxa in New Mexico have generally higher overall levels of AMF colonization than nearby non-gypseous grasslands. The same authors also found high levels of colonization by dark septate fungi and hyaline septate endophytic fungi.

All of these results accord with the a priori prediction of Palacio et al. (2012) that gypsophiles have a higher degree of mycorrhizal infection than gypsovags, although it is important to note that these authors did not find higher rates of AMF colonization in gypsophiles vs. gypsovags in their own study, nor did they find support for the hypothesis that AMF are responsible for the high levels of soil macronutrients that characterize such taxa. To explain both the high diversity of AMF on gypsophiles and the presence of potentially undescribed taxa, Alguacil et al. (2009a) postulate the existence of strong selective pressures that have been able to promote the specialization of symbiotic microorganisms, helping vascular gypsophiles to proliferate under heavy stress. This hypothesis adds a possible coevolutionary dimension to the mechanisms involved in gypsophily.

THE ASSEMBLY AND EVOLUTION OF GYPSOPHILE FLORAS

A comparison of existing, albeit incomplete, checklists and other related literature concerning gypsophiles reveals several interesting patterns relevant to the assembly of gypsophile floras worldwide. First, it is clear that each of the major gypsophile floras evolved independently, drawing their constituent taxa from local plant lineages. For example, all of the common gypsophiles in the Chihuahuan Desert (*e.g.*, species of *Tiquilia*, *Acleisanthes*, *Nerisyrenia*, *Nama*, etc.; Figure 2), Spain (*e.g.*, species of *Helianthemum*, *Ononis*, *Teucrium*, *Limonium*, etc.; Figure 3) and Somalia (*e.g.*, species of *Commiphora*, *Euphorbia*, *Kleinia*, etc.) are members of larger genera or species groups with centers of diversity in the same region (Mota et al., 2011; Thulin, 1993; 1995; 1999; 2006; Turner & Powell, 1979). Although some

individual gypsophile taxa may be widely distributed within a particular gypsum region (*e.g.*, *Dicranocarpus parviflorus* in the Chihuahuan Desert), there is no evidence of direct long-distance dispersal of gypsophiles among major gypsum regions, with the possible exception of *Campanula fastigiata*, which is found in both Spain and Cyprus (Hadjikyriakou & Hand, 2011; Mota et al., 2011). Even in larger cosmopolitan genera like *Euphorbia, Helianthemum*, and *Campanula*, which have different gypsophiles in multiple major gypsum regions of the world (*e.g.*, gypsophile taxa in *Euphorbia* exist in both Somalia and the Chihuahuan Desert, but these taxa are not shared between the two regions), it is clear based on morphological and/or molecular evidence that the gypsophiles within each genus are locally derived rather than the result of long-distance dispersal (Mota et al., 2011; Thulin, 1993; 1995; 1999; 2006; Turner & Powell, 1979).

A preliminary review of floristic literature also reveals that the overwhelming majority of gypsophiles fall within just a few major flowering plant clades. For example, of 44 Spanish taxa that ranked highest (a rating ≥ 4) on the lists of gypsum plant taxa from Mota et al. (2009) and Mota et al. (2011), and thus may be considered gypsophiles, 18 are asterids, 9 are Caryophyllales, 6 are Brassicales, and 11 belong to other groups (clade membership sensu Angiosperm Phylogeny Group, 2009). Although species lists are incomplete or absent for other regions of the world, patterns of clade membership appear similar to those seen in Spain. The same preponderance of asterids and Caryophyllales characterizes the gypsophile floras of the Chihuahuan Desert, Somalia and Australia, with key Brassicales groups in the former two regions as well [*e.g.*, *Nerisyrenia* (Brassicaceae) in the Chihuahuan Desert, and *Cleome* (Cleomaceae) and *Reseda* (Resedaceae) in Somalia] (Thulin, 1993; Turner & Powell, 1979). For example, 58% of the taxa listed as gypsophiles in Powell & Turner (1977) are asterids and 27% are Caryophyllales, while 9 of the 13 gypsophile taxa listed by Symon (2007) from southern Australia are asterids. This global bias toward clade membership in such groups as asterids, Caryophyllales and Brassicales likely reflects underlying predispositions for gypsum tolerance within these groups. To examine these clade membership patterns more rigorously requires a more thorough global checklist of gypsophiles, which we are currently assembling.

Recent phylogenetic studies that have included gypsophile taxa have also revealed several trends in the origin and evolution of gypsophiles. The overwhelming majority of such studies to date have examined Chihuahuan Desert gypsophiles, with several clear patterns having emerged from these studies. First, multiple origins of gypsophily are typical within plant lineages that appear to be ancestrally tolerant of gypsum. Excellent examples of this phenomenon have been documented in recent studies of regionally dominant gypsophile taxa in the Chihuahuan Desert: Marlowe & Hufford (2007) found three independent origins of gypsophily within *Gaillardia* (Asteraceae), Moore & Jansen (2007) found two origins of gypsophily in *Tiquilia* subg. *Eddya* (Ehretiaceae), Douglas & Manos (2007) found at least four origins of gypsophily in tribe Nyctagineae (Nyctaginaceae) [although not available to Douglas & Manos (2007), the inclusion of two more gypsophile species of Nyctagineae from Somalia (*Acleisanthes somalensis* and *Commicarpus reniformis*) raises the number of origins in this clade to at least six (Levin, 2000; M. Thulin, pers. comm.)], McKown et al. (2005) implied at least three origins of gypsophily in subtribe Flaveriinae (Asteraceae), Taylor (2012) found three origins of gypsophily in *Nama* (Hydrophyllaceae) and Schenk (2013)

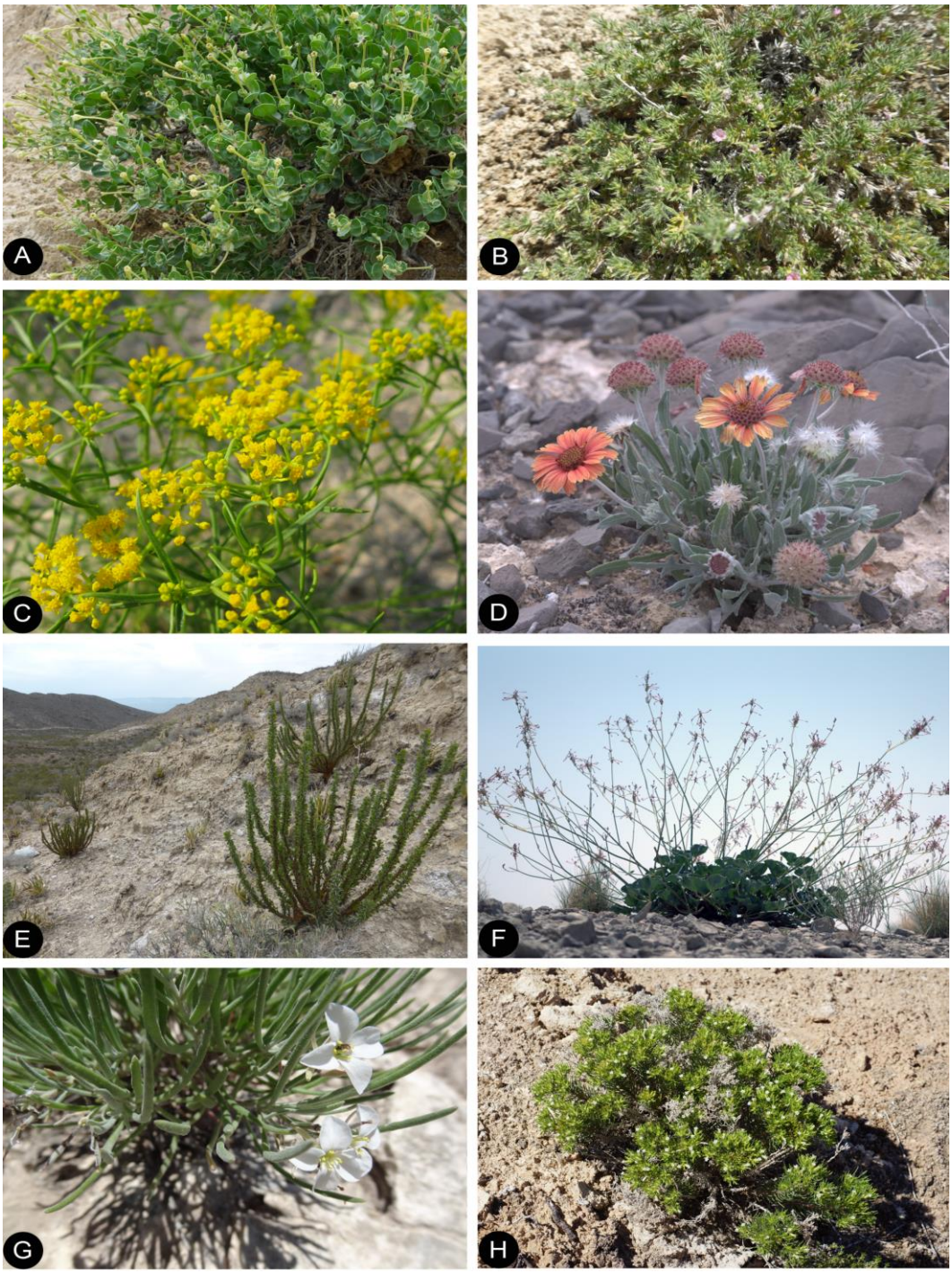

Figure 2. Examples of Chihuahuan Desert gypsophiles: A) *Acleisanthes lanceolata* var. *megaphylla* (Nyctaginaceae); B) *Tiquilia hispidissima* (Ehretiaceae); C) *Sartwellia flaveriae* (Asteraceae); D) *Gaillardia henricksonii* (Asteraceae); E) *Fouquieria shrevei* (Fouquieriaceae); F) *Anulocaulis leiosolenus* var. *howardii* (Nyctaginaceae); G) *Nerisyrenia gracilis* (Brassicaceae); H) *Nama carnosum* (Hydrophyllaceae).

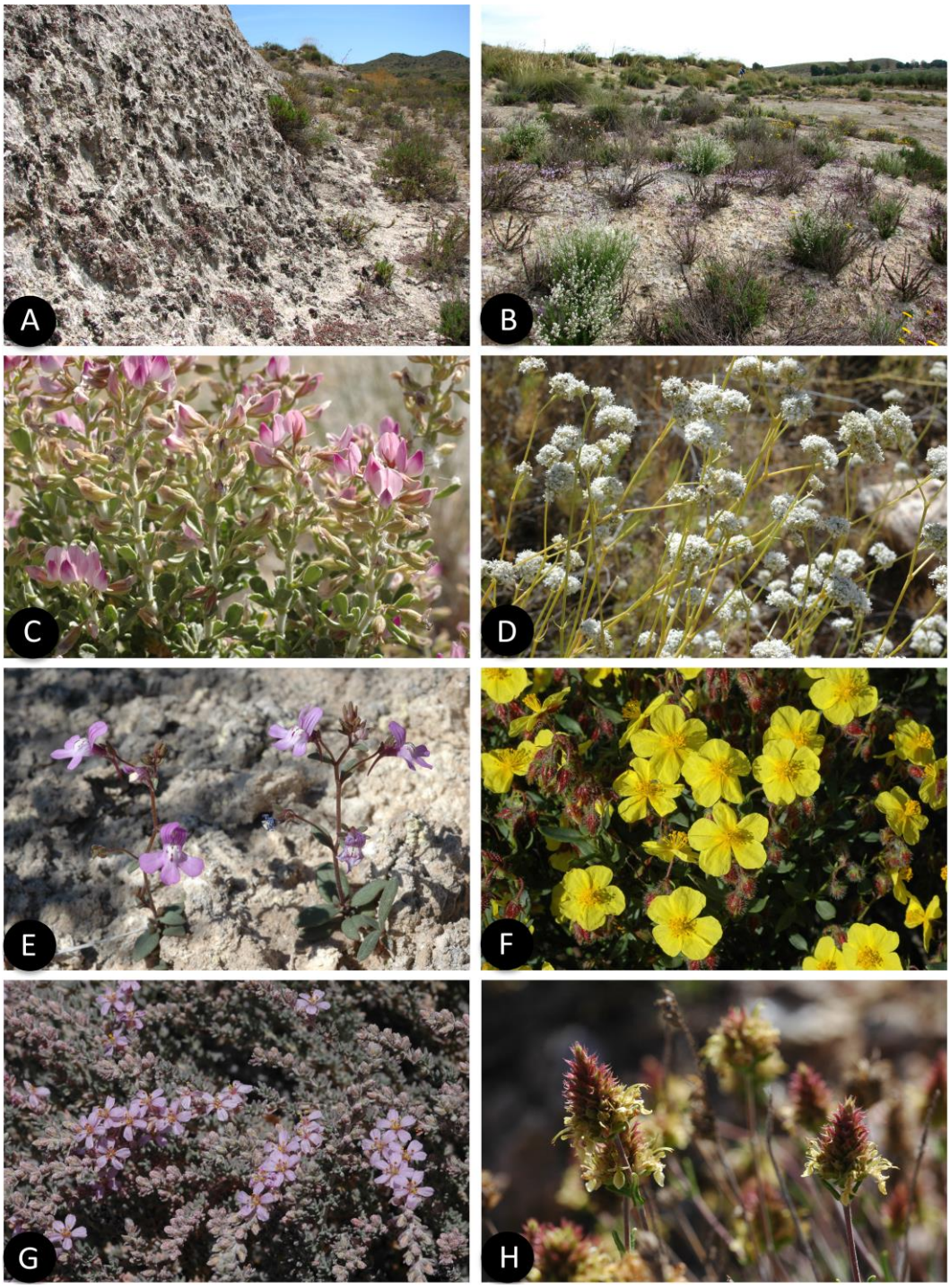

Figure 3. Examples of gypsum habitats and gypsophiles in Spain: A) Gypsum scarp with *Sedum gypsicola* (Crassulaceae) and the lichen *Parmelia pokorny* (Parmeliaceae); B) Gypsum scrubland at Venta de los Yesos, Almería; C) *Ononis tridentata* (Fabaceae); D) *Gypsophila struthium* subsp. *struthium* (Caryophyllaceae); E) *Chaenorhinum grandiflorum* (Plantaginaceae); F) *Helianthemum alypoides* (Cistaceae); G) *Frankenia thymifolia* (Frankeniaceae); H) *Teucrium lepicephalum* (Lamiaceae).

documented up to five origins of gypsophily in *Mentzelia* sect. *Bartonia* (Loasaceae). In all of these examples, the larger group containing the gypsophiles possesses numerous other taxa that are gypsovags. For example, all non-gypsophile taxa in *Tiquilia* subg. *Eddya* grow both on and off of gypsum (Moore & Jansen, 2007; Richardson, 1977), and numerous members of tribe Nyctagineae (*e.g.*, *Anulocaulis eriosolenus,* all non-gypsophile species of *Allionia* and *Cyphomeris* and many non-gypsophile taxa of *Acleisanthes, Boerhavia*, and *Mirabilis*), *Gaillardia* (*e.g.*, *G. pulchella, G. spathulata*, and *G. parryi*) and *Mentzelia* (*e.g.*, *M. nuda, M. mexicana*, and *M. saxicola*) are also gypsovags (Douglas & Manos, 2007; Schenk, 2013; Thompson & Powell, 1981; Turner & Watson, 2007). Although phylogenetic studies including gypsophiles from other regions of the world are scarcer, those that have been completed support the results from in the Chihuahuan Desert. For example, at least three origins of gypsophily have been confirmed or implied in Spanish *Helianthemum* (Cistaceae; leading to the gypsophiles *H. squamatum, H. alypoides*, and *H. conquense*) and Mediterranean *Campanula* (Campanulaceae; leading to the Spanish/Cypriot gypsophile *C. fastigiata*, the North African gypsophile *C. filicaulis* subsp. *reboudiana* and the Turkish gypsophile *C. pinnatifida* var. *germanicopolitana*) and can be expected in Spanish *Limonium* (Plumbaginaceae) (Mota et al., 2009; 2011; Parejo-Farnés et al., 2013; Roquet et al., 2008). Each of these genera is characterized by numerous other gypsovag taxa as well.

Within the gypsophile flora of the Chihuahuan Desert region, existing phylogenetic studies further suggest that speciation has occurred frequently after the acquisition of gypsophily, particularly in those lineages that comprise the regionally dominant taxa on gypsum. Clades of regionally dominant gypsophiles have been documented in phylogenetic studies of *Gaillardia* (which has two gypsophilic clades; Marlowe & Hufford, 2007), *Tiquilia* subg. *Eddya* (Moore & Jansen, 2007), *Acleisanthes* (Levin, 2000), *Nama* (Taylor, 2012), *Mentzelia* sect. *Bartonia* (Schenk & Hufford, 2011), *Leucophyllum* (Scrophulariaceae; Gándara & Sosa, 2013), and *Argemone* (Papaveraceae; Schwarzbach & Kadereit, 1999), and unpublished data in the senior author's lab suggest that clades of gypsophiles exist in *Sartwellia* (Asteraceae), *Haploësthes* (Asteraceae), *Nerisyrenia* and *Anulocaulis* (Nyctaginaceae). Although not typically dominant on gypsum, the gypsophile Cactaceae genera *Aztekium* (2 species) and *Geohintonia* (1 species) form a clade and have also speciated on gypsum (Hernández-Hernández et al., 2011). In several of these gypsophile clades, extensive speciation has occurred. For example, the Chihuahuan Desert gypsophile clade of *Acleisanthes* comprises six taxa; the principal clade of *Nama* gypsophiles comprises 10 taxa, of which 8 taxa are gypsophiles; while *Nerisyrenia* is composed almost entirely of gypsophiles, with all but one of 12 described taxa being gypsophiles (Bacon, 1978; Fowler & Turner, 1977; Taylor, 2012). Most of the gypsophile clades, and all such clades with the largest number of taxa, are broadly distributed across the Chihuahuan Desert, despite the island-like nature of gypsum exposures. Although these clades as a whole are broadly distributed, individual taxa within them generally occupy much narrower geographic ranges that are usually allopatric from one another, suggesting that allopatric speciation is typically responsible for taxon boundaries within these gypsophile lineages. A good example of this phenomenon is provided by the gypsophile clade of *Nama*, the distribution of which is illustrated in Figure 4.

While the lack of phylogenetic studies in other gypsophile floras prevents firm conclusions, it is possible that similar phylogenetic and biogeographic patterns may also characterize some of the other more broadly distributed gypsophile floras. For example, possible clades of gypsophiles may exist within *Ononis* (Fabaceae), *Teucrium* (Lamiaceae), *Orobanche*

(Orobanchaceae) and *Chaenorhinum* (Plantaginaceae) in Spain, within *Psephellus* (Asteraceae) in Turkey, and within *Pseudoblepharispermum* (Asteraceae) and *Xylocalyx* (Orobanchaceae) in the Horn of Africa region (Mota et al., 2011; Thulin, 2006; Wagenitz & Kandemir, 2008).

Molecular evidence indicates that many gypsophile lineages around the globe may have appeared no earlier than the late Miocene (*ca*. 8-5.3 mya). Using molecular dating techniques, Moore & Jansen (2006; 2007) found that the two origins of gypsophily in *Tiquilia* subg. *Eddya* dated most likely to the early Pliocene and early-to-mid Pleistocene, respectively, with the earlier origin leading to the geographically widespread and regionally dominant *T*. *hispidissima* taxon complex, and the later origin leading to the geographically restricted clade of *T*. *turneri* and *T*. *tuberculata*. A late Miocene or early Pliocene divergence time was also favored for the split of the gypsophile (and morphologically quite distinctive) cactus genera *Aztekium* and *Geohintonia* (mean age = 5.67 mya), suggesting gypsophily is at least that old in that lineage (Hernández-Hernández et al., 2014).

In a molecular dating analysis of Cornales (which includes *Mentzelia* of the Loasaceae), Schenk & Hufford (2010) recovered a Pleistocene origin for *Mentzelia* sect. *Bartonia*, which includes numerous gypsophile taxa. Although they did not perform a separate dating analysis of these gypsophile lineages, it is clear from studies with more complete taxon sampling that the regionally dominant and geographically widespread Chihuahuan Desert gypsophile *Mentzelias* (*M*. *perennis, M*. *todiltoensis*, and *M*. *humilis*) diverged early in the history of the section, implying that they are older than the other gypsophile taxa of *Mentzelia*, which are all in more recently derived positions, have narrow distributions outside the Chihuahuan Desert, and represent distinct origins of gypsophily (Schenk, 2013; Schenk & Hufford, 2011). Gándara et al. (2014) recovered a late Miocene divergence time between the morphologically distinctive and monotypic gypsophile genus *Jaimehintonia* (Amaryllidaceae) and its nearest relative, suggesting that gypsophily arose in *Jaimehintonia* after that point. Wagstaff & Tate (2011) found a similar late Miocene divergence time between the Australian gypsophile *Lawrencia helmsii* (Malvaceae) and its congeners, again placing a late Miocene upper bound on the origin of gypsophily in this lineage.

In contrast, there are numerous other gypsophile lineages composed of single species that have restricted geographic ranges and are morphologically much more similar to their non-gypsophile relatives. Examples include *Tiquilia turneri, Mirabilis nesomii* (Nyctaginaceae), *Abronia nealleyi* (Nyctaginaceae), *Nama stevensii*, and *Gaillardia gypsophila*, all of which have been found to have very little phylogenetic distance separating them from morphologically very similar congeners, implying a very recent origin (Marlowe & Hufford, 2007; Moore & Jansen, 2007; Taylor, 2012; unpublished data). Presumably these taxa appeared in the Pleistocene, as was inferred for *Tiquilia turneri* in the molecular dating analyses of Moore & Jansen (2006; 2007). The existence of a mix of older and younger gypsophile lineages is important because it suggests that modern gypsophile floras have assembled gradually over the last several million years.

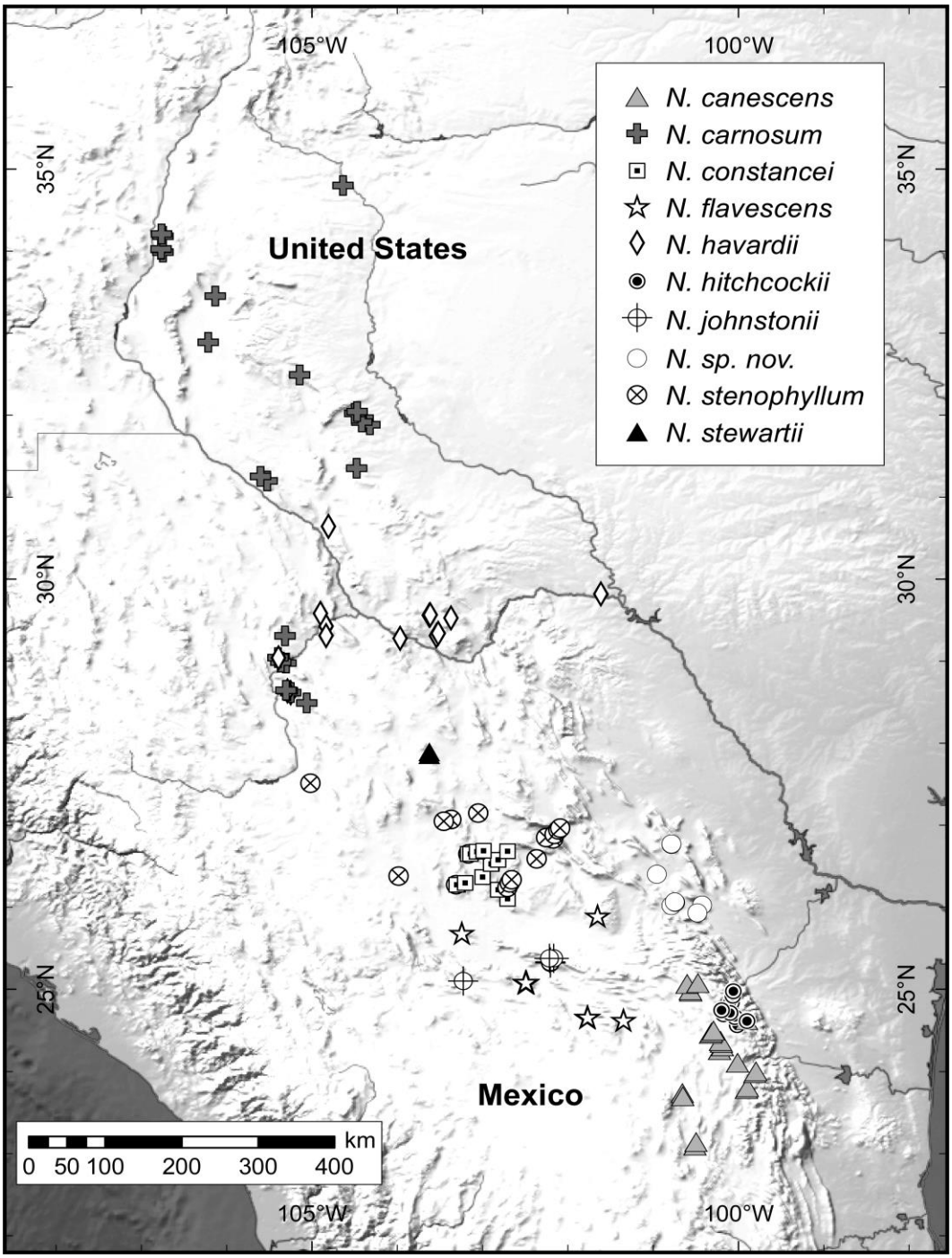

Figure 4. Distribution of taxa within the gypsophile clade of *Nama* (Hydrophyllaceae). All of these taxa are gypsophiles, with the exception of the gypsovags *N. johnstonii* and *N. havardii*.

The post-Miocene assembly of gypsophile floras corresponds well with the current hypotheses concerning the spread of semi-arid and arid habitats during the Cenozoic. After a peak of global average temperature and precipitation in the late Paleocene and early Eocene, the Earth experienced several major episodes of cooling and drying, culminating in the most recent major episode during the latest Miocene and Pliocene (Graham, 2011; Zachos et al., 2008). Available paleoclimatic evidence suggests that it was not until this period that arid and semi-arid regions began to occupy relatively large portions of the Earth's surface (Arakaki et al., 2011; Axelrod, 1979; Graham, 2011; Hernández-Hernández et al., 2014; Salzmann et al., 2008). Given that all of the world's gypsophile plant assemblages occur in such habitats, it is unlikely that gypsum habitats dry enough and extensive enough to support gypsophile floras existed prior to the latest Miocene. Additional molecular dating analyses will be necessary to test this hypothesis further.

EVOLUTION AT THE POPULATION LEVEL

The population structure of gypsophiles should be largely determined by the island-like distribution of gypsum outcrops across the landscape. This edaphic restriction places an upper limit on the population size any gypsophile species can achieve. Once such a species has colonized a particular gypsum "island," however, it may remain on that outcrop indefinitely. Hence allele frequencies in gypsophiles should reach an equilibrium reflecting the combined effects of migration, mutation and genetic drift. The close correspondence of the island-like distributions of gypsophiles to the assumptions of well-studied theoretical models of population genetic structure (*e.g.*, the stepping-stone model; Kimura & Weiss, 1964) provides an opportunity to infer aspects of their demographic and evolutionary history from parameters commonly estimated in population genetic studies.

In gypsophiles, as with many desert and island taxa (Filner & Shmida, 1981), we would expect migration to be generally quite limited (and selection may actually favor reduced dispersibility; Schenk, 2013). Thus, isolation-by-distance should be evident. While low population size has no effect on mutation rates, at least one aspect of the biology of gypsophiles may serve to increase the effective population size, thus reducing the rate at which genetic diversity is lost due to genetic drift: almost all gypsophiles are perennials, with overlapping generations. Genetic diversity may also be maintained by outcrossing. Only a few gypsophiles are obvious selfers [though mixed mating systems may be common; for example *Acleisanthes* produces both cleistogamous and chasmogamous flowers (Douglas & Manos, 2007)]. On the other hand, biparental inbreeding in small populations may have the opposite effect. Finally, during the climatic oscillations of the Pleistocene, gypsophiles may not have been subjected to repeated genetic bottlenecks as severe as those suffered by plants in other habitats (*e.g.*, alpine taxa), because community composition on unusual substrates such as gypsum may be more stable than substrate generalist communities over a broader range of climates (Damschen et al., 2012; Harrison et al., 2009; Tapper et al., 2014).

In general, endemic taxa tend to have lower genetic diversity than widespread taxa, but measures of population structure do not seem to differ greatly between rare and common species (see reviews by Cole, 2003; Gitzendanner & Soltis, 2000; Hamrick & Godt, 1989).

However, the number of migrants is typically much reduced in rare species (Cole, 2003) as compared to common ones.

To examine whether genetic variation and population structure in gypsophile taxa differ from that in "ordinary" endemics in predictable ways, we followed the example of these three reviews of genetic variation in plants with contrasting life histories (Cole, 2003; Gitzendanner & Soltis, 2000; Hamrick & Godt, 1989). We summarized available estimates of population genetic parameters from five studies that focused on gypsophile taxa; we also included five other studies of Spanish gypsoclines and a study of the gypsovag *Arctomecon californica*, a close congener of the gypsophile *A. humilis,* for comparison (Table 2). We report the following statistics that reflect genetic diversity of these taxa: percentage of polymorphic loci, *P*; number of alleles per locus, *A*; Nei's total gene diversity, H_t; average genetic diversity within populations, H_s; effective number of alleles, A_e; and observed heterozygosity, H_o. We also tabulated estimates of population differentiation, including G_{st}, F_{st}, or among-population variance from analysis of molecular variance (AMOVA).

Some studies estimated the effective number of migrants, N_m. For the sake of comparison, we estimated N_m from F_{st} or G_{st} for the remaining studies where this was possible. Finally, we report the degree to which populations exhibited isolation-by-distance. Averages discussed in the following section exclude diversity parameters estimated from known polyploids, which typically have larger numbers of alleles, and from haploid chloroplast data. Comparisons of parameter values to those obtained from the three reviews should be viewed qualitatively, since the small number of studies of gypsophiles precludes rigorous statistical analysis.

Genetic Diversity

The taxa in Table 2 tend to show levels of genetic variation similar to that expected for endemic (Hamrick & Godt, 1989) or rare (Cole, 2003; Gitzendanner & Soltis, 2000) plant species, although by some measures, they exceed the genetic diversity typical of widespread species. The percentage of polymorphic loci ranged from 0.10 to 0.82, with a mean of 0.53, intermediate to the averages for rare and widespread taxa in the three reviews cited above. The number of alleles per locus varied from 1.43 to 1.71 (mean 1.59). Though this was reported in only three enzyme studies of diploid taxa, the value lies slightly below averages for rare or endemic taxa in the three reviews. Nei's total gene diversity (H_t) averaged 0.24 (range 0.15-0.34), slightly below the estimated means of endemics in Hamrick & Godt (1989), but actually greater than later estimates for widespread species (Cole, 2003; Gitzendanner & Soltis, 2000). Average genetic diversity within populations is 0.17, slightly higher than found for endemics in general (0.16; Hamrick & Godt, 1989). Perhaps more significantly, the estimated effective number of alleles (Kimura & Crow, 1964), which we estimated as $1/(1- H_t)$, averaged 1.33 (1.18-1.51), while this statistic (which depends on total

Table 2. Population genetic parameters estimated for gypsophile (in bold) and selected gypsocline species, as well as the gypsovag *Arctomecon californica*. Abbreviations: Pops = number of populations, Inds = number of individuals, P = proportion of loci that are polymorphic, A = mean number of alleles per locus, H_o = observed heterozygosity, H_e = expected heterozygosity, H_t = Nei's gene diversity, A_e = effective number of alleles (calculated from H_o), H_s = mean within-population gene diversity, N_m = effective number of migrants, IBD = isolation-by-distance. For population differentiation, the following indicators apply: † = G_{st} or F_{st}; ‡ = among-population variance from AMOVA. Average parameter values calculated from diploid data only; # identifies polyploid taxa or haploid genomes excluded from parameter averages. Significant R^2 values identified by an asterisk (*). The final three entries report parameter averages from published reviews of population parameters in plants, for comparison

Citation	Taxon	Data Type	Pops	Inds	P	A	H_o	H_e	A_e	H_t	H_s	Population Differentiation	N_m	R^2 (IBD)
Allphin et al., 1998	***Arctomecon humilis***	isozyme	6	163	0.104	1.43	0.100	0.103	1.51	0.339		0.620†	0.16	
Hickerson & Wolf, 1998	*Arctomecon californica*	allozyme	16	480	0.554	1.71	0.158	0.163		0.239	0.163	0.320†	0.54	0.048*
Aguirre-Liguori et al., 2014	***Fouquieria shrevei***	cpDNA#	5	94								0.850†, 0.709‡		0.532*
Pérez-Collazos & Catalán, 2008	***Ferula loscosii***	allozyme	11	330	0.327	1.62	0.164	0.125	1.18	0.152	0.125	0.134†	1.62	0.839*
Pérez-Collazos et al., 2009	***Ferula loscosii***	AFLP	12	342	0.523						0.171	0.440‡	0.32	0.811*
Salmerón-Sánchez et al., 2014	*Jurinea pinnata*	AFLP	16	160								0.370‡	0.42	
Martínez-Nieto et al., 2013	***Gypsophila struthium* subsp. *hispanica***	AFLP	7	82	0.573				1.35	0.258	0.200	0.226†, 0.280‡	0.86	0.038
		cpDNA#								0.810	0.381	0.530†		
	Gypsophila struthium* subsp. *struthium	AFLP	16	185	0.562				1.29	0.224	0.160	0.286†, 0.334‡	0.62	0.128
		cpDNA#								0.827	0.292	0.647†		
Jiménez & Sánchez-Gómez, 2012	*Moricandia moricandioides* subsp. *pseudofoetida*	ISSR	1	50	0.817						0.213	0.533†, 0.665‡	0.19	
	Moricandia moricandioides subsp. *moricandioides*	ISSR	1	30	0.790						0.213			
Pérez-Collazos et al., 2008	*Boleum asperum*#	AFLP	10	240	0.913				3.91	0.744		0.202‡	0.99	0.443*

Citation	Taxon	Data Type	Pops	Inds	*P*	*A*	H_o	H_e	A_e	H_t	H_s	Population Differentiation	N_m	R^2 (IBD)
Pérez-Collazos & Catalán, 2006	*Vella pseudocytisus* subsp. *paui*[#]	allozyme	6	162	0.636				2.23	0.552		0.200‡	1.00	
		AFLP	**6**	**162**	**0.625**				**2.39**	**0.581**		**0.219‡**	**0.89**	**0.560***
López-Pujol et al., 2004	*Thymus loscosii*[#]	allozyme	8	257	0.850	3.00	0.472	0.422	1.80	0.444	0.429	0.033†	7.33	
	Average				**0.531**	**1.59**	**0.141**	**0.130**	**1.33**	**0.242**	**0.169**	**0.423†, 0.466‡**	**0.589**	**0.425**
Hamrick & Godt, 1989	**Endemic**				0.400	1.80		0.096	1.15	0.263	0.163	0.248		
	Widespread				0.589	2.29		0.202	1.31	0.347	0.267	0.210		
Gitzendanner & Soltis, 2000	**Endemic**				0.367	1.94				0.219		0.206		
	Widespread				0.449	2.23				0.242		0.224		
Cole, 2003	**Rare**				0.407	1.74	0.100	0.113		0.142		0.212	1.190	
	Common				0.588	2.34	0.139	0.150		0.199		0.198	2.240	

heterozygosity) averaged 1.31 among widespread species in Hamrick & Godt (1989). Thus, while gypsophiles tend to possess fewer alleles per locus than most rare or endemic plant species, they do not show obviously reduced heterozygosity. Finally, in the three studies that reported observed heterozygosity, H_o averaged 0.14, nearly equal to the figure reported for common taxa in Cole (2003). It is likely that the old ages of these populations, and their long-term stability, have allowed allele frequencies to reach equilibrium.

Population Differentiation

Measures of population differentiation (F_{st} or G_{st}: mean 0.42; or proportion of variation explained by differences between populations from AMOVA: 0.47) were in general higher than those found for rare or common species (Table 2) (Cole, 2003). This is not surprising given the static, island-like distribution of their habitats through time. In *Gypsophila struthium* subsp. *struthium* and *G. s.* subsp. *hispanica* (Martínez-Nieto et al., 2013), chloroplast haplotypes were used in addition to nuclear markers (AFLPs) to estimate population differentiation. The estimated values of among-population chloroplast haplotype differentiation were 0.65 and 0.53 in these two taxa respectively, compared to differentiations of 0.33 and 0.28, as estimated by AMOVA on the AFLP markers. A similarly high value (0.65) was obtained from an AMOVA of chloroplast haplotypes from *Fouquieria shrevei*, the only gypsophile taxon from the Chihuahuan Desert that has been studied at the population level (Aguirre-Liguori et al., 2014). The fact that chloroplast differentiation in *Gypsophila* is roughly twice that of the nuclear genome results from inherent differences in effective population size between genomic compartments, which for chloroplasts in hermaphroditic plants is expected to be ½ that of nuclear loci (Birky et al., 1989). Interestingly, there is no indication of greater differentiation in chloroplast data as one might expect if seed dispersal were more limited than pollen dispersal, or lesser differentiation, which could indicate additional nuclear gene flow via pollen dispersal. In the absence of nuclear data, there is no way to evaluate this in *Fouquieria shrevei*, but it is important to recognize that seemingly very high population differentiation values for chloroplast data do not necessarily imply that seed dispersal is necessarily more restricted than gene flow through pollen.

Migration and Isolation-By-Distance

Migration was estimated by some authors (Allphin et al., 1998; Hickerson & Wolf, 1998; López-Pujol et al., 2004; Pérez-Collazos & Catalán, 2006; Pérez-Collazos et al., 2009), who generally based their estimates on the value of F_{st}. While estimates of the number of effective migrants based on population differentiation must be viewed with extreme caution (Whitlock & McCauley, 1999), for the sake of comparison, we calculated values for the gypsophile species based on the reported among-population variation (Table 2). N_m values thus obtained averaged only 0.59, much lower than the mean value for rare species in Cole (2003). This may reflect the highly discontinuous nature of gypsum outcrops in Spain, where the majority of these population genetic surveys have been conducted. Additional reports from different areas may shed light on how much the patchiness of gypsum outcrops affects migration.

These studies often examined whether genetic distance was correlated with geographic distance, in other words, whether isolation-by-distance (IBD) was evident in their datasets. A stepping-stone model, in which gene flow is a function of geographic distance, is likely to produce such a pattern if populations are at equilibrium. However, if populations have recently expanded into new areas, such a pattern may not have had time to emerge (Slatkin, 1993). Moderate to strong IBD was manifest up to a distance of 24.8 km in the locally distributed gypsoclines *Vella pseudocytisus* subsp. *paui* (Pérez-Collazos & Catalán, 2006) and *Boleum asperum* (Pérez-Collazos et al., 2008). Isolation-by-distance is also evident in the gypsophiles *Gypsophila struthium* sensu lato (Martínez-Nieto et al., 2013) and *Ferula loscosii* (Pérez-Collazos & Catalán, 2008; Pérez-Collazos et al., 2009), which occur in multiple gypsum areas of Spain. In the Chihuahuan Desert, the regionally dominant gypsophile *Fouquieria shrevei* shows IBD as well (Aguirre-Liguori et al., 2014). IBD in these gypsophiles is obviously driven largely by the geographic separation of discrete populations with limited gene flow between them, rather than genetic structure within continuous habitat. In contrast, IBD appears to be weak within subspecies of *Gypsophila struthium* (Martínez-Nieto et al., 2013) and in the gypsovag *Arctomecon californica* (Hickerson & Wolf, 1998), perhaps limited by sustained high gene flow in comparatively continuous habitat, or by recent population expansion.

General Phylogeographic Patterns

Few of these studies explicitly test a phylogeographic model; however, Pérez-Collazos et al. (2009) discerned a Pliocene colonization of the Iberian Peninsula from north Africa in *Ferula loscosii*, followed by south-to-north dispersal through the Pleistocene. In *Gypsophila struthium* (Martínez-Nieto et al., 2013), chloroplast data suggest that central and eastern Spain represents the ancestral range, which has expanded, and given rise to *G. struthium* subsp. *hispanica* in eastern and, more recently, northern Spain, specifically the Ebro Valley, which is home to several unique gypsophiles. Finally, Aguirre-Liguori et al. (2014) determined that chloroplast haplotypes in *Fouquieria shrevei* are invariant at low elevation sites in western Coahuila, Mexico, which were inundated during pluvials in the Pleistocene, whereas montane gypsum sites show greater diversity, consistent with longer residence of these populations *in situ*. As the number of phylogeographic investigations of gypsophiles increases, we will be able to better characterize the response of populations to historical climate fluctuations, which is key to understanding the diversification of gypsophile floras.

CONCLUSION

Although great progress has been made in understanding the ecology, assembly and evolution of gypsophile floras worldwide, much remains to be explored in this major but underappreciated edaphic community. The ecological mechanisms controlling the establishment of gypsophile floras deserve further study, especially with regard to interactions among physical, chemical, and biological factors operating in the rhizosphere. Ecological and floristic studies would be particularly welcome in areas with different climates than Spain, especially in places like the Chihuahuan Desert, Iran, and Somalia, all of which have important

differences in rainfall amounts and seasonality compared to each other and to Spain, and which are likely to yield many additional gypsophile taxa. Finally, further phylogeographic and phylogenetic studies are needed in gypsum environments throughout the world to assess whether island biogeographic patterns are typical of gypsum archipelagoes, both at the community and genetic level, and to confirm whether different gypsophile communities share similar ages and assembly characteristics.

ACKNOWLEDGMENTS

The authors thank Sara Palacio and two additional reviewers for helpful reviews that improved this manuscript, and Arianna Goodman for technical help with figures. This work was supported by grants from Oberlin College, the US National Science Foundation (DEB-1054539) and the National Geographic Society (8873-11) to MJM, and by grants from the Consejería de Economía, Innovación y Ciencia of the Junta de Andalucía (co-financed with FEDER funds) to JFM.

REFERENCES

Aguirre-Liguori, J. A., Scheinvar, E. & Eguiarte, L. E. (2014) Gypsum soil restriction drives genetic differentiation in *Fouquieria shrevei* (Fouquieriaceae). *American Journal of Botany 101*, 730–736.

Akhani, H. (2004) A new spiny, cushion-like *Euphorbia* (Euphorbiaceae) from south-west Iran with special reference to the phytogeographic importance of local endemic species. *Botanical Journal of the Linnean Society 146*, 107–121.

Akpulat, H. A. & Celik, N. (2005) Flora of gypsum areas in Sivas in the eastern part of Cappadocia in Central Anatolia, Turkey. *Journal of Arid Environments 61*, 27–46.

Alguacil, M. M., Roldán, A. & Torres, M. P. (2009a) Assessing the diversity of AM fungi in arid gypsophilous plant communities. *Environmental Microbiology 11*, 2649–2659.

Alguacil, M. M., Roldán, A. & Torres, M. P. (2009b) Complexity of semiarid gypsophilous shrub communities mediates the AMF biodiversity at the plant species level. *Microbial Ecology 57*, 718–727.

Alguacil, M. M., Torrecillas, E., Roldán, A., Díaz, G. & Torres, M. P. (2012) Perennial plant species from semiarid gypsum soils support higher AMF diversity in roots than the annual *Bromus rubens*. *Soil Biology and Biochemistry 49*, 132–138.

Allphin, L., Windham, M. D. & Harper, K. T. (1998) Genetic diversity and gene flow in the endangered dwarf bear poppy, *Arctomecon humilis* (Papaveraceae). *American Journal of Botany 85*, 1251–1261.

Anderson, D. C., Harper, K. T. & Holmgren, R. C. (1982) Factors influencing development of cryptogamic soil crusts in Utah deserts. *Journal of Range Management 35*, 180–185.

Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society 161*, 105–121.

Arakaki, M., Christin, P.-A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R. M., Spriggs, E., Moore, M. J. & Edwards, E. J. (2011) Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences, USA 108*, 8379–8384.

Atwood, N. D. & Welsh, S. L. (2005) New species of *Mentzelia* (Loasaceae) and *Phacelia* (Hydrophyllaceae) from New Mexico. *Western North American Naturalist 65*, 365–370.

Axelrod, D. I. (1979) Age and origin of Sonoran Desert vegetation. *Occasional Papers of the California Academy of Sciences*, *No. 132*.

Bacon, J. D. (1978) Taxonomy of *Nerisyrenia* (Cruciferae). *Rhodora 80*, 159–227.

Birky, C. W., Fuerst, P. & Maruyama, T. (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics 121*, 613–627.

Borer, C. H., Hamby, M. N. & Hutchinson, L. H. (2012) Plant tolerance of a high calcium environment via foliar partitioning and sequestration. *Journal of Arid Environments 85*, 128–131.

Boscaiu, M., Bautista, I., Lidón, A., Llinares, J., Lull, C., Donat, P., Mayoral, O. & Vicente, O. (2013) Environmental-dependent proline accumulation in plants living on gypsum soils. *Acta Physiologiae Plantarum 35*, 2193–2204.

Boukhris, M. & Lossaint, P. (1970) Sur la teneur en soufre de quelque plantes gypsophiles de Tunisie. *Oecologia Plantarum 5*, 345–354.

Boukhris, M. & Lossaint, P. (1972) Spécificité biogeochimique des plantes gypsophiles de Tunisie. *Oecologia Plantarum 7*, 45–68.

Boyadgiev, T. G. (1974) *Contribution to the knowledge of gypsiferous soils*. Rome: Food and Agriculture Organization of the United Nations.

Cañadas, E. M., Ballesteros, M., Valle, F. & Lorite, J. (2013) Does gypsum influence seed germination? *Turkish Journal of Botany 38*, 141–147.

Cole, C. T. (2003) Genetic variation in rare and common plants. *Annual Review of Ecology, Evolution, and Systematics 34*, 213–237.

Contejan, C. (1881) *Géographie botanique: Influence du terrain sur la végétation*. Paris: Librairie J. B. Baillere et fils.

Damschen, E. I., Harrison, S., Ackerly, D. D., Fernandez-Going, B. M. & Anacker, B. L. (2012) Endemic plant communities on special soils: Early victims or hardy survivors of climate change? *Journal of Ecology 100*, 1122–1130.

Daniells, I. G. (2012) Hardsetting soils: A review. *Soil Research 50*, 349–359.

Davis, D. S., Droop, S. J. M., Gregerson, P., Henson, L., Leon, C. J., Villa-Lobos, J. L., Synge, H. & Zantovska, J. (1986) *Plants in danger: What do we know?* Cambridge: International Union for Conservation of Nature and Natural Resources.

Douglas, N. A. & Manos, P. S. (2007) Molecular phylogeny of Nyctaginaceae: Taxonomy, biogeography, and characters associated with a radiation of xerophytic genera in North America. *American Journal of Botany 94*, 856–872.

Drohan, P. J. & Merkler, D. J. (2009) How do we find a true gypsophile? *Geoderma 150*, 96–105.

Duvigneaud, P. & Denaeyer-de Smet, S. (1966) Accumulation du soufre dans quelques espèces gypsophiles d'Espagne. *Bulletin de la Société Royale de Botanique de Belgique 99*, 263–269.

Duvigneaud, P. & Denaeyer-de Smet, S. (1968) Essai de classification chimique (éléments minéraux) des plantes gypsicoles du bassin de l'Ebre. *Bulletin de la Société Royale de Botanique de Belgique 101*, 279–291.

Duvigneaud, P. & Denaeyer-de Smet, S. (1973) Considerations sur l'ecologie de la nutrition minerale des tapis vegetaux naturels. *Oecologia Plantarum 8*, 219–246.

Erben, M. & Arán, V. J. (2005) *Limonium mateoi* (Plumbaginaceae), a new species from Central Spain. *Anales del Jardín Botánico de Madrid 62*, 3–7.

Escudero, A., Palacio, S., Maestre, F. T. & Luzuriaga, A. L. (2014) Plant life on gypsum: A review of its multiple facets. *Biological Reviews*, DOI: 10.1111/brv.12092

Eswaran, H. & Gong, Z.-T. (1991) Properties, genesis, classification, and distribution of soils with gypsum. In: W. D. Nettleton (Ed.). *Occurrence, characteristics, and genesis of carbonate, gypsum, and silica accumulations in soils* (pp. 89–119). SSSA Special Publication. Madison, WI: Soil Science Society of America.

Filner, S. & Shmida, A. (1981) Why are adaptations for long-range seed dispersal rare in desert plants? *Oecologia 51*, 133–144.

Fink, S. (1991) The micromorphological distribution of bound calcium in needles of Norway spruce [*Picea abies* (L.) Karst.]. *New Phytologist 119*, 33–40.

Food and Agriculture Organization (1990) *Management of gypsiferous soils (FAO Soils Bulletin 62)*. Rome: Food and Agriculture Organization of the United Nations.

Food and Agriculture Organization (1998) *World reference base for soil resources*. *World Soil Resources Report 84*. Rome: Food and Agriculture Organization of the United Nations.

Forbis de Queiroz, T., Baughman, C., Baughman, O., Gara, M. & Williams, N. (2012) Species distribution modeling for conservation of rare, edaphic endemic plants in White River Valley, Nevada. *Natural Areas Journal 32*, 149–158.

Fowler, B. A. & Turner, B. L. (1977) Taxonomy of *Selinocarpus* and *Ammocodon* (Nyctaginaceae). *Phytologia 37*, 177–208.

Franceschi, V. R. & Nakata, P. A. (2005) Calcium oxalate in plants: formation and function. *Annual Review of Plant Biology 56*, 41–71.

Gándara, E. & Sosa, V. (2013) Testing the monophyly and position of the North American shrubby desert genus *Leucophyllum* (Scrophulariaceae: Leucophylleae). *Botanical Journal of the Linnean Society 171*, 508–518.

Gándara, E., Specht, C. D. & Sosa, V. (2014) Origin and diversification of the *Milla* Clade (Brodiaeoideae, Asparagaceae): a Neotropical group of six geophytic genera. *Molecular Phylogenetics and Evolution 75*, 118-125.

Gibbens, R. P. & Lenz, J. M. (2001) Root systems of some Chihuahuan Desert plants. *Journal of Arid Environments 49*, 221–263.

Gil de Carrasco, C. & Ramos, J. J. (2011) Los suelos yesíferos (gypsisoles) en España. In: J. F. Mota Poveda, P. Sánchez Gómez & J. S. Guirado Romero (Eds.). *Diversidad vegetal de las Yeseras Ibéricas: El reto de los archipiélagos edáficos para la biología de la conservación* (pp. 33–50). Almería, Spain: ADIF-Mediterráneo Asesores Consultores.

Gitzendanner, M. A. & Soltis, P. S. (2000) Patterns of genetic variation in rare and widespread plant congeners. *American Journal of Botany 87*, 783–792.

Graham, A. (2011) The age and diversification of terrestrial New World ecosystems through Cretaceous and Cenozoic time. *American Journal of Botany 98*, 336–351.

Guerrero-Campo, J., Alberto, F., Hodgson, J., García-Ruiz, J. M. & Montserrat-Martí, G. (1999) Plant community patterns in a gypsum area of NE Spain. I. Interactions with topographic factors and soil erosion. *Journal of Arid Environments 41*, 401–410.

Hadjikyriakou, G. & Hand, R. (2011) *Teucrium salaminium* Hadjik. & Hand (Lamiaceae, *Teucrium* sect. *Polium*), a new species from Cyprus. *Candollea 66*, 341–351.

Hamrick, J. L. & Godt, M. J. W. (1989) Allozyme diversity in plant species. In: A. H. D. Brown, M. T. Clegg, A. L. Kahler & B. S. Weir (Eds.). *Plant population genetics, breeding, and genetic resources* (pp. 43–63). Sunderland, MA: Sinauer Associates.

Harrison, S., Damschen, E. & Going, B. M. (2009) Climate gradients, climate change, and special edaphic floras. *Northeastern Naturalist 16*, 121–130.

Harrison, S. & Rajakaruna, N. (Eds.) (2011) *Serpentine: The evolution and ecology of a model system*. Berkeley, CA: University of California Press.

Hawkesford, M., Horst, W., Kichey, T., Lambers, H., Schjoerring, J., Skrumsager Møller, I. & White, P. (2012) Functions of macronutrients. In: P. Marschner (Ed.). *Marschner's mineral nutrition of higher plants*. Third edition (pp. 135–190). London: Academic Press.

Hernández-Hernández, T., Hernández, H. M., De-Nova, J. A., Puente, R., Eguiarte, L. E. & Magallón, S. (2011) Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *American Journal of Botany 98*, 44–61.

Hernández-Hernández, T., Brown, J. W., Schlumpberger, B. O., Eguiarte, L. E. & Magallón, S. (2014) Beyond aridification: Multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytologist 202*, 1382-1397.

Herrero, J. & Porta, J. (2000) The terminology and the concepts of gypsum-rich soils. *Geoderma 96*, 47–61.

Herrero, J. (2004) Revisiting the definitions of gypsic and petrogypsic horizons in Soil Taxonomy and World Reference Base for Soil Resources. *Geoderma 120*, 1–5.

Herrero, J., Artieda, O. & Hudnall, W. H. (2009) Gypsum, a tricky material. *Soil Science Society of America Journal 73*, 1757–1763.

Hickerson, L. L. & Wolf, P. G. (1998) Population genetic structure of *Arctomecon californica* Torrey & Frémont (Papaveraceae) in fragmented and unfragmented habitat. *Plant Species Biology 13*, 21–33.

Hinsinger, P., Bengough, A. G., Vetterlein, D. & Young, I. M. (2009) Rhizosphere: Biophysics, biogeochemistry and ecological relevance. *Plant and Soil 321*, 117–152.

Hiouani, F. (2006) *Influence de la teneur en gypse et de la taille de ses grains sur la capacité de retention en eau sur les sols de zone de Ain Benoui -Biskra*. MS Thesis: Algiers, Ministere de L'Enseignement Superieur et de la Recherche.

Jiménez, J. F. & Sánchez-Gómez, P. (2012) Molecular taxonomy and genetic diversity of *Moricandia moricandioides* subsp. *pseudofoetida* compared to wild relatives. *Plant Biosystems 146*, 99–105.

Johnston, I. M. (1941) Gypsophily among Mexican desert plants. *Journal of the Arnold Arboretum 22*, 145–170.

Jones, J. B. (2012) *Plant nutrition and soil fertility manual*. Second edition. Boca Raton, FL: CRC Press.

Kalra, Y. (1997) *Handbook of reference methods for plant analysis*. Boca Raton, FL: CRC Press.

Kimura, M. & Crow, J. F. (1964) The number of alleles that can be maintained in a finite population. *Genetics 49*, 725–738.

Kimura, M. & Weiss, G. H. (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics 49*, 561–576.

Kleinkopf, G. E. & Wallace, A. (1974) Physiological basis for salt tolerance in *Tamarix ramosissima*. *Plant Science Letters 3*, 157–163.

Le Houérou, H. N. (1969) La végétation de la Tunisie steppique. *Annales de l'Institut National de la Recherche Agronomique de Tunisie 42*, 1–622.

Levin, R. A. (2000) Phylogenetic relationships within Nyctaginaceae tribe Nyctagineae: evidence from nuclear and chloroplast genomes. *Systematic Botany 25*, 738–750.

López-Pujol, J., Bosch, M., Simon, J. & Blanché, C. (2004) Allozyme diversity in the tetraploid endemic *Thymus loscosii* (Lamiaceae). *Annals of Botany 93*, 323–332.

Macchiati, L. (1888) Contribuzione alla flora del gesso. *Nuevo Giornale Botanico Italiano 20*, 418–422.

Marlowe, K. & Hufford, L. (2007) Taxonomy and biogeography of *Gaillardia* (Asteraceae): A phylogenetic analysis. *Systematic Botany 32*, 208–226.

Martínez-Nieto, M. I., Segarra-Moragues, J. G., Merlo, E., Martínez-Hernández, F. & Mota, J. F. (2013) Genetic diversity, genetic structure and phylogeography of the Iberian endemic *Gypsophila struthium* (Caryophyllaceae) as revealed by AFLP and plastid DNA sequences: Connecting habitat fragmentation and diversification. *Botanical Journal of the Linnean Society 173*, 654–675.

Mashali, A. M. (1996) Soil management practices for gypsiferous soils. In: R. M. Poch (Ed.). *Proceedings of the international symposium on soils with gypsum* (pp. 34–52). Lleida: Edicions Universitat de Lleida.

McKown, A. D., Moncalvo, J.-M. & Dengler, N. G. (2005) Phylogeny of *Flaveria* (Asteraceae) and inference of C4 photosynthesis evolution. *American Journal of Botany 92*, 1911–1928.

Merlo, M. E., Mota, J. F., Cabello, J. & Alemán, M. M. (1998) La gipsofilia en plantas: Un apasionante edafismo. *Investigación y Gestión 3*, 103–112.

Merlo, M. E., Rodríguez-Tamayo, M. L., Jiménez, M. L. & Mota, J. F. (2001) Recapitulación sobre el comportamiento biogeoquímico de algunos gipsófitos y halófitos ibéricos. *Monografías Flora Vegetación Béticas 12*, 97–106.

Merlo, M. E., Mota, J. F. & Sánchez-Gómez, P. (2011) Ecofisiología y adaptaciones de las plantas vasculares a las características físicas y químicas de sustratos especiales. In: J. F. Mota, P. Sánchez-Gómez & J. S. Guirado Romero (Eds.). *Diversidad vegetal de las Yeseras Ibéricas: El reto de los archipiélagos edáficos para la biología de la conservación* (pp. 53–73). Almería: ADIF-Mediterráneo Asesores Consultores.

Meyer, S. E. (1986) The ecology of gypsophile endemism in the eastern Mojave Desert. *Ecology 67*, 1303–1313.

Meyer, S. E. & García-Moya, E. (1989) Plant community patterns and soil moisture regime in gypsum grasslands of north central Mexico. *Journal of Arid Environments 16*, 147–155.

Moore, M. J. & Jansen, R. K. (2006) Molecular evidence for the age, origin, and evolutionary history of the American desert plant genus *Tiquilia* (Boraginaceae). *Molecular Phylogenetics and Evolution 39*, 668–687.

Moore, M. J. & Jansen, R. K. (2007) Origins and biogeography of gypsophily in the Chihuahuan Desert plant group *Tiquilia* subg. *Eddya* (Boraginaceae). *Systematic Botany 32*, 392–414.

Mota, J. F., Sola, A. J., Jiménez-Sánchez, M. L., Pérez-García, F. & Merlo, M. E. (2004) Gypsicolous flora, conservation and restoration of quarries in the southeast of the Iberian Peninsula. *Biodiversity and Conservation 13*, 1797–1808.

Mota, J. F., Sánchez Gómez, P., Merlo Calvente, M. E., Catalán Rodríguez, P., Laguna Lumbreras, E., De la Cruz Rot, M., Navarro Reyes, F. B., Marchal Gallardo, F., Bartolomé Esteban, C., Martínez Labarga, J. M., Sainz Ollero, H., Valle Tendero, F., Serra Laliga, L., Martínez Hernández, F., Garrido Becerra, J. A. & Pérez García, F. J. (2009) Aproximación a la checklist de los gipsófitos ibéricos. *Anales de Biología 31*, 71–80.

Mota, J. F., Sánchez Gómez, P. & Guirado Romero, J. S. (Eds.) (2011) *Diversidad vegetal de las Yeseras Ibéricas: El reto de los archipiélagos edáficos para la biología de la conservación*. Almería, Spain: ADIF-Mediterráneo Asesores Consultores.

Oyonarte, C., Sánchez, G., Urrestarazu, M. & Alvarado, J. J. (2002) A comparison of chemical properties between gypsophile and nongypsophile plant rhizospheres. *Arid Land Research and Management 16*, 47–54.

Palacio, S., Escudero, A., Montserrat-Martí, G., Maestro, M., Milla, R. & Albert, M. J. (2007) Plants living on gypsum: Beyond the specialist model. *Annals of Botany 99*, 333–343.

Palacio, S., Johnson, D., Escudero, A. & Montserrat-Martí, G. (2012) Root colonisation by AM fungi differs between gypsum specialist and non-specialist plants: Links to the gypsophile behaviour. *Journal of Arid Environments 76*, 128–132.

Parejo-Farnés, C., Albaladejo, R. G., Arroyo, J. & Aparicio, A. (2013) A phylogenetic hypothesis for *Helianthemum* (Cistaceae) in the Iberian Peninsula. *Botanica Complutensis 37*, 83–92.

Parsons, R. F. (1976) Gypsophily in plants—a review. *American Midland Naturalist 96*, 1–20.

Pérez-Collazos, E. & Catalán, P. (2006) Palaeopolyploidy, spatial structure and conservation genetics of the narrow steppe plant *Vella pseudocytisus* subsp. *paui* (Vellinae, Cruciferae). *Annals of Botany 97*, 635–647.

Pérez-Collazos, E. & Catalán, P. (2008) Conservation genetics of the endangered Iberian steppe plant *Ferula loscosii* (Apiaceae). *Plant Biology 10*, 492–501.

Pérez-Collazos, E., Segarra-Moragues, J. P. & Catalán, P. (2008) Two approaches for the selection of relevant genetic units for conservation in the narrow European endemic steppe plant *Boleum asperum* (Brassicaceae). *Biological journal of the Linnean Society 94*, 341–354.

Pérez-Collazos, E., Sánchez-Gómez, P., Jiménez, J. F. & Catalán, P. (2009) The phylogeographical history of the Iberian steppe plant *Ferula loscosii* (Apiaceae): A test of the abundant-centre hypothesis. *Molecular Ecology 18*, 848–861.

Petrusson, L. & Thulin, M. (1996) Taxonomy and biogeography of *Gymnocarpos* (Caryophyllaceae). *Edinburgh Journal of Botany 53*, 1–26.

Poch, R. M. (1998) Pore space characteristics as indicators of soil behaviour in gypsiferous soils. *Geoderma 87*, 87–109.

Poch, R. M. & Verplancke, H. (1997) Penetration resistance of gypsiferous horizons. *European Journal of Soil Science 48*, 535–543.

Porras-Alfaro, A., Raghavan, S., Garcia, M., Sinsabaugh, R. L., Natvig, D. O. & Lowrey, T. K. (2014) Endophytic fungal symbionts associated with gypsophilous plants. *Botany 92*, 295–301.

Powell, A. M. & Turner, B. L. (1977) Aspects of the plant biology of the gypsum outcrops of the Chihuahuan Desert. In: R. H. Wauer & D. H. Riskind (Eds.). *Transactions of the*

symposium on the biological resources of the Chihuahuan Desert region, United States and Mexico, Sul Ross State University, Alpine, Texas, 17-18 October 1974 (pp. 315–325). Washington, DC: U.S. Department of the Interior, National Park Service Transactions and Proceedings Series, Number 3.

Richardson, A. T. (1977) Monograph of the genus *Tiquilia* (*Coldenia*, sensu lato), Boraginaceae: Ehretioideae. *Rhodora 79*, 467–572.

Romão, R. L. & Escudero, A. (2005) Gypsum physical soil crusts and the existence of gypsophytes in semi-arid central Spain. *Plant Ecology 181*, 127–137.

Roquet, C., Sáez, L., Aldasoro, J. J., Susanna, A., Alarcón, M. L. & García-Jacas, N. (2008) Natural delineation, molecular phylogeny and floral evolution in *Campanula*. *Systematic Botany 33*, 203–217.

Salmerón-Sánchez, E., Martínez-Nieto, M. I., Martínez-Hernández, F., Garrido-Becerra, J. A., Mendoza-Fernández, A. J., de Carrasco, C. G., Ramos-Miras, J. J., Lozano, R., Merlo, M. E. & Mota, J. F. (2014) Ecology, genetic diversity and phylogeography of the Iberian endemic plant *Jurinea pinnata* (Lag.) DC. (Compositae) on two special edaphic substrates: Dolomite and gypsum. *Plant and Soil 374*, 233–250.

Salzmann, U., Haywood, A. M., Lunt, D. J., Valdes, P. J. & Hill, D. J. (2008) A new global biome reconstruction and data-model comparison for the middle Pliocene. *Global Ecology and Biogeography 17*, 432–447.

Schenk, J. J. (2013) Evolution of limited seed dispersal ability on gypsum islands. *American Journal of Botany 100*, 1811–1822.

Schenk, J. J. & Hufford, L. (2010) Effects of substitution models on divergence time estimates: simulations and an empirical study of model uncertainty using Cornales. *Systematic Botany 35*, 578–592.

Schenk, J. J. & Hufford, L. (2011) Phylogeny and taxonomy of *Mentzelia* section *Bartonia* (Loasaceae). *Systematic Botany 36*, 711–720.

Schwarzbach, A. E. & Kadereit, J. W. (1999) Phylogeny of prickly poppies, *Argemone* (Papaveraceae), and the evolution of morphological and alkaloid characters based on ITS nrDNA sequence variation. *Plant Systematics and Evolution 218*, 257–279.

Sivinski, R. & Howard, M. O. (2011) A new species of *Linum* (Linaceae) from the northern Chihuahuan Desert. *Phytoneuron 2011-33*, 1–7.

Slatkin, M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution 47*, 264–279.

Symon, D. E. (2007) Lists of gypsophilous plants from southern Australia. *Journal of the Adelaide Botanic Gardens 21*, 45–54.

Tapper, S.-L., Byrne, M., Yates, C. J., Keppel, G., Hopper, S. D., Van Niel, K., Schut, A. G. T., Mucina, L. & Wardell-Johnson, G. W. (2014) Prolonged isolation and persistence of a common endemic on granite outcrops in both mesic and semi-arid environments in south-western Australia. *Journal of Biogeography*, DOI: 10.1111/jbi.12343

Taylor, S. E. (2012) *Molecular systematics and the origins of gypsophily in* Nama *L. (Boraginaceae)*. PhD Dissertation. Austin, TX: University of Texas at Austin.

Thompson, H. J. & Powell, A. M. (1981) Loasaceae of the Chihuahuan Desert region. *Phytologia 49*, 16–32.

Thulin, M. (1993) *Flora of Somalia, Volume 1: Pteridophyta; Gymnospermae; Angiospermae (Annonaceae–Fabaceae)*. London: Royal Botanic Gardens, Kew.

Thulin, M. (1995) *Flora of Somalia, Volume 4: Angiospermae (Hydrocharitceae–Pandanaceae)*. London: Royal Botanic Gardens, Kew.

Thulin, M. (1999) *Flora of Somalia, Volume 2: Angiospermae (Tiliaceae–Apiaceae)*. London: Royal Botanic Gardens, Kew.

Thulin, M. (2006) *Flora of Somalia, Volume 3: Angiospermae (Ericaceae–Asteraceae)*. London: Royal Botanic Gardens, Kew.

Turner, B. L. & Powell, A. M. (1979) Deserts, gypsum, and endemism. In: J. R. Goodin & D. K. Northington (Eds.). *Arid land plant resources: Proceedings of the international arid lands conference on plant resources, Texas Tech University* (pp. 96–116). Lubbock, TX: International Center for Arid and Semi-Arid Land Studies.

Turner, B. L. & Watson, T. (2007) Taxonomic revision of *Gaillardia* (Asteraceae). *Phytologia Memoirs 13*. Gruver, TX; Texensis Publishing.

Verheye, W. H. & Boyadgiev, T. G. (1997) Evaluating the land use potential of gypsiferous soils from field pedogenic characteristics. *Soil Use and Management 13*, 97–103.

Wagenitz, G. & Kandemir, A. (2008) Two new species of the genus *Psephellus* (Compositae, Cardueae) from eastern Turkey. *Willdenowia 38*, 521–526.

Wagstaff, S. J. & Tate, J. A. (2011) Phylogeny and character evolution in the New Zealand endemic genus *Plagianthus* (Malveae, Malvaceae). *Systematic Botany 36*, 405–418.

Wang, B. & Qiu, Y.-L. (2006) Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza 16*, 299–363.

Whitlock, M. C. & McCauley, D. E. (1999) Indirect measures of gene flow and migration: FST|[ne]|1/(4Nm+1). *Heredity 82*, 117–125.

Zachos, J. C., Dickens, G. R. & Zeebe, R. E. (2008) An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature 451*, 279–283.

LCH

A Targeted Enrichment Strategy for Massively Parallel Sequencing of Angiosperm Plastid Genomes

Author(s): Gregory W. Stull , Michael J. Moore , Venkata S. Mandala , Norman A. Douglas , Heather-Rose Kates , Xinshuai Qi , Samuel F. Brockington , Pamela S. Soltis , Douglas E. Soltis , and Matthew A. Gitzendanner

Source: Applications in Plant Sciences, 1(2) 2013.
Published By: Botanical Society of America
DOI: http://dx.doi.org/10.3732/apps.1200497
URL: http://www.bioone.org/doi/full/10.3732/apps.1200497

Applications in Plant Sciences

Applications in Plant Sciences 2013 1(2): 1200497

Protocol Note

A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes[1]

Gregory W. Stull[2,3,8], Michael J. Moore[4], Venkata S. Mandala[4], Norman A. Douglas[4], Heather-Rose Kates[3,5], Xinshuai Qi[6], Samuel F. Brockington[7], Pamela S. Soltis[3,5], Douglas E. Soltis[2,3,5], and Matthew A. Gitzendanner[2,3,5]

[2]Department of Biology, University of Florida, Gainesville, Florida 32611-8525 USA; [3]Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611-7800 USA; [4]Department of Biology, Oberlin College, Oberlin, Ohio 44074-1097 USA; [5]Genetics Institute, University of Florida, Gainesville, Florida 32611 USA; [6]Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, and Laboratory of Systematic and Evolutionary Botany and Biodiversity, College of Life Sciences, Zhejiang University, Hangzhou 310058 People's Republic of China; and [7]Department of Plant Sciences, University of Cambridge, Cambridge CB13EA United Kingdom

- *Premise of the study:* We explored a targeted enrichment strategy to facilitate rapid and low-cost next-generation sequencing (NGS) of numerous complete plastid genomes from across the phylogenetic breadth of angiosperms.
- *Methods and Results:* A custom RNA probe set including the complete sequences of 22 previously sequenced eudicot plastomes was designed to facilitate hybridization-based targeted enrichment of eudicot plastid genomes. Using this probe set and an Agilent SureSelect targeted enrichment kit, we conducted an enrichment experiment including 24 angiosperms (22 eudicots, two monocots), which were subsequently sequenced on a single lane of the Illumina GAIIx with single-end, 100-bp reads. This approach yielded nearly complete to complete plastid genomes with exceptionally high coverage (mean coverage: 717×), even for the two monocots.
- *Conclusions:* Our enrichment experiment was highly successful even though many aspects of the capture process employed were suboptimal. Hence, significant improvements to this methodology are feasible. With this general approach and probe set, it should be possible to sequence more than 300 essentially complete plastid genomes in a single Illumina GAIIx lane (achieving ~50× mean coverage). However, given the complications of pooling numerous samples for multiplex sequencing and the limited number of barcodes (e.g., 96) available in commercial kits, we recommend 96 samples as a current practical maximum for multiplex plastome sequencing. This high-throughput approach should facilitate large-scale plastid genome sequencing at any level of phylogenetic diversity in angiosperms.

Key words: next-generation sequencing; phylogenomics; plastid genomes.

Over the past few years, complete plastid genome sequencing has emerged as a powerful and increasingly accessible tool for plant phylogenetics, facilitated by rapid advances in next-generation sequencing (NGS) technologies (e.g., Moore et al., 2006, 2007, 2010; Jansen et al., 2007; Cronn et al., 2008, 2012). Many aspects of the plastid genome, including its structural simplicity, relatively small size, and highly conserved gene content, make it ideally suited for next-generation sequencing and assembly. Additionally, its wealth of characters, useful across many taxonomic levels, makes it an excellent resource for phylogenetic studies across the plant branch of the tree of life. Plastome-scale phylogenetic studies have, for example, clarified relationships among major angiosperm lineages (Moore et al., 2007, 2010; Jansen et al., 2007) and resolved recent, rapid radiations in *Pinus* (Parks et al., 2009). Plastid genomes also have great potential for population genetic and phylogeographic studies (e.g., Whittall et al., 2010), particularly as a complement to multiple unlinked nuclear loci, although this application of large-scale plastid data sets has been underexplored compared to deeper-level phylogenetic studies.

The ever-increasing capacities of next-generation sequencers, particularly the Illumina platforms, coupled with the high-copy nature of the plastid genome, have made it possible to multiplex numerous samples of whole-genomic DNA (gDNA) on a single lane and still recover sufficient coverage to assemble complete or nearly complete plastid genomes (e.g., Cronn et al., 2008, 2012; Steele et al., 2012; Straub et al., 2012). However, given that plastid DNA typically constitutes only ~0.5–13% of gDNA samples (Steele et al., 2012; Straub et al., 2012), this approach expends much of the sequencing capacity on nuclear reads, significantly reducing the number of plastomes that can be sequenced in parallel. Consequently, this limits the scalability

[1]Manuscript received 17 September 2012; revision accepted 9 December 2012.

The authors thank K. Kupsch for sharing material of *Monococcus echinophorus*, M. Heaney for providing DNA of *Nolina brittoniana*, M. Croley and R. Mostow for assistance with library construction, Andy Crowl and Nico Cellinese for proving the *Campanula* sample, and R. Cronn for helpful advice. We also thank K. E. Holsinger and two anonymous reviewers for their helpful comments on the manuscript. This research was funded by National Science Foundation grants DBI-0735191 and EF-0431266 and the Oberlin College Office of Sponsored Programs.

[8]Author for correspondence: gwstull@ufl.edu

doi:10.3732/apps.1200497

of plastid genome sequencing for large-scale phylogenetic and phylogeographic studies when funding is limited.

By increasing the abundance of plastid DNA relative to the nuclear and mitochondrial genomes, targeted enrichment strategies for the plastid genome offer a promising means of vastly increasing the number of plastomes that can be multiplexed on a single lane. Some researchers have used long-range PCR to amplify segments of the plastid genome as one enrichment strategy (e.g., Cronn et al., 2008; Njuguna et al., 2013). However, such methods are more time-intensive and require appropriate primer design as well as high-quality DNA to ensure amplification of the long segments. Another method of enriching for plastids is through sucrose gradient centrifugation during DNA extraction (e.g., Moore et al., 2006), but this requires large amounts (frequently >5 g) of fresh tissue. In contrast, hybridization-based methods of plastid enrichment, which use oligonucleotide probes (or "baits") to capture plastid targets, show considerable potential for broad applicability given their ability to enrich degraded samples (e.g., DNA from herbarium material) and their utility across large phylogenetic distances (when the probe design incorporates sequences from phylogenetically diverse samples) (e.g., Cronn et al., 2012). However, these plastid capture methods, while promising, have until now only been developed for *Pinus* (Cronn et al., 2012; Parks et al., 2012). Designing a plastid probe set of broad phylogenetic applicability has not been attempted.

Several commercial kits have been developed for hybridization-based targeted enrichment using custom probe sets (e.g., Agilent SureSelect, Roche Nimblegen, MYcroarray), and the offerings are rapidly changing. Here we present a hybridization-based method for targeted enrichment of angiosperm plastid genomes, using a custom set of RNA probes designed from 22 previously sequenced eudicot plastomes (see Table 1) and an early version of the Agilent SureSelect technology. We demonstrate the utility of this probe-based approach with results from an enrichment experiment that involved 24 angiosperms (22 species of eudicots and two species of monocots) multiplexed on a single lane of the Illumina GAIIx (Illumina Inc., San Diego, California, USA) subsequent to enrichment. The success of this experiment illustrates the utility of the capture method in general and the broad applicability of the probe set in particular. This capture method, or improvements thereto, will enable a significant increase in the number of angiosperm plastid genomes that can be multiplexed on the Illumina platform. This, in turn, will dramatically decrease per-genome sequencing costs, making large-scale sequencing of plastid genomes a feasible option for any phylogenetic or phylogeographic study. Furthermore, the broad phylogenetic utility of the probe set employed here makes this method applicable for plastome-based evolutionary studies across not only eudicots, but also monocots and potentially all angiosperms.

METHODS AND RESULTS

***Probe design*—**RNA probes ("baits") were designed by Genotypic Technology Ltd. (Bangalore, India) from the complete plastid genomes of 22 eudicot species, selected to represent much of the phylogenetic breadth of eudicots (Table 1). We chose to limit bait design to eudicots to maximize the utility of the bait array for plastid phylogenomics throughout this clade, which includes approximately 75% of angiosperm diversity (Drinnan et al., 1994; Soltis et al., 2005) and has been the subject of ongoing research in our laboratories (e.g., Jian et al., 2008; Wang et al., 2009; Brockington et al., 2009; Moore et al., 2010; Arakaki et al., 2011). For each input genome, 120-bp baits were designed, with 50-bp overlap (~2× tiling). To minimize representational bias of highly conserved regions of the plastid genome (e.g., rRNA genes) during hybridization capture, bait sequences for all genomes were compared using BLAST, and only baits with <90% identity to all other baits were retained in the final bait design. In all, ~55 000 baits were included in the final design. The bait sequences and coordinates are available in Appendix S1.

***Sampling*—**To test the efficacy of the bait array for plastome capture, we constructed Illumina libraries for 24 species (Table 2), representing 22 eudicots and two monocots. The 22 eudicots span the phylogenetic diversity of the clade, including species from *Rosidae*, *Asteridae*, and *Caryophyllales* (sensu Cantino et al., 2007). These species were also selected to test the effects on plastome capture of increasing phylogenetic distance from the sequences included in the bait design. For example, we constructed libraries for one species that was part of the bait design (*Cucumis sativus*), one species (*Oenothera hartwegii*) that is

Table 1. Eudicot plastomes used for probe design.

Taxon	Family (Order)	GenBank accession no.
Antirrhinum majus L.	Plantaginaceae (Lamiales)	Unpublished data (M. J. Moore)
Arabidopsis thaliana (L.) Heynh.	Brassicaceae (Brassicales)	NC_000932
Citrus sinensis (L.) Osbeck	Rutaceae (Sapindales)	NC_008334
Cornus florida L.	Cornaceae (Cornales)	Unpublished data (M. J. Moore)
Cucumis sativus L.	Cucurbitaceae (Cucurbitales)	NC_007144
Dillenia indica L.	Dilleniaceae (Dilleniales)	Unpublished data (M. J. Moore)
Ficus sp.	Moraceae (Rosales)	Unpublished data (M. J. Moore)
Gossypium hirsutum L.	Malvaceae (Malvales)	NC_007944
Helianthus annuus L.	Asteraceae (Asterales)	NC_007977
Ilex cornuta Lindl. & Paxton	Aquifoliaceae (Aquifoliales)	Unpublished data (M. J. Moore)
Liquidambar styraciflua L.	Altingiaceae (Saxifragales)	Unpublished data (M. J. Moore)
Lonicera japonica Thunb.	Caprifoliaceae (Dipsacales)	Unpublished data (M. J. Moore)
Nandina domestica Thunb.	Berberidaceae (Ranunculales)	NC_008336
Nerium oleander L.	Apocynaceae (Gentianales)	Unpublished data (M. J. Moore)
Oenothera biennis L.	Onagraceae (Myrtales)	NC_010361
Oxalis latifolia Kunth	Oxalidaceae (Oxalidales)	Unpublished data (M. J. Moore)
Platanus occidentalis L.	Platanaceae (Proteales)	NC_008335
Plumbago auriculata Lam.	Plumbaginaceae (Caryophyllales)	Unpublished data (M. J. Moore)
Populus trichocarpa Torr. & A. Gray	Salicaceae (Malpighiales)	NC_009143
Spinacia oleracea L.	Amaranthaceae (Caryophyllales)	NC_002202
Staphylea colchica Steven	Staphyleaceae (Crossosomatales)	Unpublished data (M. J. Moore)
Ximenia americana L.	Olacaceae (Santalales)	Unpublished data (M. J. Moore)

Table 2. Eudicot and monocot species included in this study, with voucher information and assembly statistics.

Taxon	Family (Order)	Voucher (Herbarium)	No. of plastid reads/ total reads	% Plastid reads	% Plastid reads (unenriched)*	% Plastome recovered	Mean coverage
Acleisanthes lanceolata (Wooton) R. A. Levin	Nyctaginaceae (Caryophyllales)	R. Merkel 8 (OC)	1 478 311/2 001 153	73.9	17.7	99.83	1091
Campanula erinus L.	Campanulaceae (Asterales)	A. Crowl 42 (FLAS)	292 895/833 412	35	4.1	95.8	176
Cucumis sativus L.	Cucurbitaceae (Cucurbitales)	cv. 'Calypso' (Seminis Vegetable Seeds)	2 131 764/2 495 346	85.4	N/A	100	1408
Dicranocarpus parviflorus A. Gray	Asteraceae (Asterales)	M. Moore 655 (OC)	1 660 972/2 408 461	69	14.5	99.99	1250
Frankenia L. sp.	Frankeniaceae (Caryophyllales)	S. F. Brockington (s.n.)	2 658 678/3 839 653	69	N/A	84.5	2088
Glinus dahomensis (Fenzl) A. Chev.	Molluginaceae (Caryophyllales)	S. F. Brockington (cultivated from seed, s.n.)	152 181/413 956	36.8	N/A	93.3	87
Limeum L. sp.	Limeaceae (Caryophyllales)	S. F. Brockington (cultivated from seed, s.n.)	2 402 594/3 316 313	72.4	N/A	98.7	1515
Limonium limbatum Small	Plumbaginaceae (Caryophyllales)	M. Moore 694 (OC)	47 113/81 348	58	N/A	97.7	32.4
Mentzelia perennis Wooton	Loasaceae (Cornales)	M. Moore 917 (OC)	654 939/767 318	85.4	14.3	100	467
Microtea debilis Sw.	Phytolaccaceae (Caryophyllales)	M. Rimachi 11128 (TEX/LL)	580 514/1 146 486	50.6	N/A	95	375
Monococcus echinophorus F. Muell.	Phytolaccaceae (Caryophyllales)	S. F. Brockington (s.n. Burringbar Botanic Gardens Nursery)	652 299/1 014 800	64.3	5.2	97.41	477
Nama carnosum (Wooton) C. L. Hitchc.	Boraginaceae (unplaced lamiid)	M. Moore 678 (OC)	1 069 755/1 606 440	66.6	N/A	99.96	693
Nepenthes alata Blanco	Nepenthaceae (Caryophyllales)	M. Moore 1145 (OC)	528 035/1 106 057	47.7	N/A	82.6	378
Nerisyrenia linearifolia (S. Watson) Greene	Brassicaceae (Brassicales)	M. Moore 671 (OC)	2 462 357/3 247 079	75.8	N/A	99.99	1573
Nolina brittoniana Nash	Asparagaceae (Asparagales)	J. M. Heaney (FLAS)	106 808/333 995	32	N/A	96	64
Oenothera hartwegii Benth.	Onagraceae (Myrtales)	M. Moore 628 (OC)	985 316/1 816 515	54.2	N/A	99.6	566
Petiveria alliacea L.	Phytolaccaceae (Caryophyllales)	L. Majure 4132 (FLAS)	12 249/27 892	43.9	2.5	84.9	9
Phaulothamnus spinescens A. Gray	Achatocarpaceae (Caryophyllales)	M. Moore 976 (OC)	3 445 475/6 452 382	53.4	N/A	99.99	2321
Physena madagascariensis Thouars ex Tul.	Physenaceae (Caryophyllales)	2007-895 (Kew Living Collection)	442 864/601 799	73.6	N/A	82	332
Sarcobatus vermiculatus (Hook.) Torr.	Sarcobataceae (Caryophyllales)	M. Moore 813 (OC)	314 733/652 540	48.2	N/A	94.01	236
Simmondsia chinensis (Link) C. K. Schneid.	Simmondsiaceae (Caryophyllales)	1972-3169 (Kew Living Collection)	861 126/1 432 458	60.1	N/A	93.8	577
Sporobolus nealleyi Vasey	Poaceae (Poales)	M. Moore 659 (OC)	378 858/824 691	45.9	4.5	99.44	312
Stegnosperma Benth. sp.	Stegnospermataceae (Caryophyllales)	S. F. Brockington (s.n.)	1 115 117/1 827 385	61	N/A	96.7	894.1
Tamarix L. sp.	Tamaricaceae (Caryophyllales)	M. Moore 320 (FLAS)	485 054/1 063 674	45.6	N/A	88.3	292.2

Note: N/A = not applicable.

*The data under "% Plastid reads (unenriched)" were taken from separate GAIIx or HiSeq runs without enrichment for the plastome (A. C. Crowl, unpublished *Campanula erinus* data; M. J. Moore, unpublished data for the rest).

congeneric with another species in the bait array (*Oenothera biennis*), species that are in different genera but the same family as species in the bait array (e.g., *Dicranocarpus parviflorus* vs. *Helianthus annuus*; both are Asteraceae), and species that are phylogenetically distant from all other taxa in the bait design (e.g., *Mentzelia perennis* [Loasaceae], *Acleisanthes lanceolata* [Nyctaginaceae]). We also included two monocots—*Nolina brittoniana* (Asparagaceae) and *Sporobolus nealleyi* (Poaceae)—to test whether the probes were effective beyond eudicots.

Some of the species sampled here—and in some cases, the same genomic libraries—were also sequenced in separate Illumina GAIIx or HiSeq (Illumina Inc.) runs (100-bp, single-end or paired-end reads) without enrichment for the plastid genome. Specifically, the following species were sequenced using both enriched and unenriched libraries: *Acleisanthes lanceolata*, *Campanula erinus* (same library), *Dicranocarpus parviflorus*, *Mentzelia perennis*, *Monococcus echinophorus* (same library), *Petiveria alliacea* (same library), *Sarcobatus vermiculatus* (same library), and *Sporobolus nealleyi.* This overlap presents an excellent opportunity to compare both depth and evenness of plastome coverage obtained using enriched vs. unenriched samples.

***Library construction*—**Genomic DNA (1–15 µg) was fragmented using a Covaris E220 sonicator (Covaris, Woburn, Massachusetts, USA) with the following parameters to produce fragmented DNA with a target peak of 500 bp: duty cycle = 5%; intensity = 3; cycles per burst = 200; time = 80 s. The NEBNext DNA Library Prep Master Mix Set for Illumina kit (Cat no.: E6040L, New England BioLabs, Ipswich, Massachusetts, USA) was then used to construct Illumina libraries with the sonicated DNAs and 24 different 5-bp barcodes from Craig et al. (2008). We followed the manufacturer's protocol for library construction, except that half reactions were used for most libraries to reduce per-sample

preparation costs. Following adapter ligation, 300–400-bp fragments (insert size ~200–300 bp) were excised and purified from agarose gels using the Freeze 'N Squeeze kit (Bio-Rad, Hercules, California, USA). The size-selected libraries were then enriched using the Phusion High-Fidelity PCR Master Mix (New England BioLabs) with the following PCR program: one cycle of 98°C for 30 s; 14–18 cycles of 98°C for 10 s, 65°C for 30 s, and 72°C for 30 s; and one cycle of 72°C for 5 min, followed by a hold at 4°C. Adapter dimers were removed from enriched libraries using 0.85 volume per sample of Agencourt AMPure XP beads (Beckman Coulter, Brea, California, USA). After AMPure purification, samples were quantified using a 2100 Bioanalyzer (Agilent, Santa Clara, California, USA) and pooled into a single, equimolar mix in preparation for plastid genome capture using a SureSelect Target Enrichment Kit (Agilent) with the custom RNA baits described above.

***Plastid genome enrichment and sequencing*—**We stress that the methods described here deviate substantially from the manufacturer's protocols (see http://www.genomics.agilent.com/GenericB.aspx?PageType=Custom&SubPageType=Custom&PageID=3120). Additionally, the kit we used has been updated as Agilent has continued to refine its enrichment products (see http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductDetail&PageID=3033). We provide the information not only as a record of our methods, but also to illustrate the robustness of the kit and to encourage further experimentation among other users.

Specifically, three significant deviations were made from the manufacturer's recommendations. First, for many reasons beyond our control, the kit was nine months past the manufacturer's expiration date when it was used—clearly we would not recommend using an expired kit, but our success should reassure others who may find themselves with similarly outdated kits. Second, the kit contains blockers for the adapters that prevent nonspecific capture via adapter-adapter annealing. We used an older kit with blockers for single-end adapters, while our libraries had barcoded paired-end adapters—thus, we did not have the correct blockers in the mix. Lastly, all 24 barcoded libraries were pooled for a single capture, although the SureSelect protocol recommends selecting individual barcoded libraries followed by pooling of samples. Agilent now offers preselection pooling of barcoded libraries, although this is currently limited to 10 libraries, and the cost, while somewhat lower than 10 individual samples, is still significantly higher than one sample. Hence, performing a single selection on pooled barcoded samples is a significant and previously unsupported deviation from the manufacturer's protocol. However, again we think that our results indicate that this method will work in many situations, and this approach is the only cost-effective option for enrichment of a small region such as the plastid genome.

Other than the three significant changes discussed above, we followed the protocol outlined for the SureSelect kit (version 1.2, April 2009), using the custom RNA baits described above. After plastid genome enrichment, the 24-library pool was amplified using the Phusion High-Fidelity Master Mix (New England BioLabs) and the following program: one cycle of 98°C for 30 s; 18 cycles of 98°C for 10 s, 57°C for 30 s, and 72°C for 30 s; and one cycle of 72°C for 7 min, followed by a hold at 4°C. The amplified product was then cleaned using AMPure XP beads and sequenced on a single lane of an Illumina GAIIx at the Interdisciplinary Center for Biotechnology Research (University of Florida) with 100 cycles and single-end reads. The sequencing run generated 47 491 666 reads.

***Plastome assembly*—**Prior to plastome assembly, the reads were barcode-sorted using Novocraft (http://www.novocraft.com/main/index.php) and quality-filtered using Sickle (https://github.com/najoshi/sickle) or the FASTQ Quality Filter (FASTX-Toolkit; http://hannonlab.cshl.edu/fastx_toolkit/). The number of reads obtained for each library is shown in Table 2. De novo assemblies were conducted with the quality-filtered reads using the VelvetOptimizer script provided with Velvet (Zerbino and Birney, 2008; k-mer range: 43–81; http://bioinformatics.net.au/software.velvetoptimiser.shtml) or Geneious (using default settings and medium to high sensitivity; http://www.geneious.com/). The resulting de novo contigs were then assembled against the most closely related available reference plastome (Table 2). Prior to reference-based assembly, we removed one of the inverted repeat regions from each reference. After assembly of the contigs to the reference, we filled in as many gaps as possible by assembling the quality-filtered reads to the reference using Geneious. Any remaining gaps were filled with Ns. Regions with very low coverage in the read-to-reference assembly (below 5× coverage) were also masked with Ns. Following assembly, we used DOGMA (Wyman et al., 2004) to annotate the plastid genomes, allowing an examination of sequence depth distribution in relation to coding vs. noncoding regions of the plastome.

***Assembly statistics*—**The percent completeness of the newly assembled plastomes (vs. the reference genomes used) is presented in Table 2, which also shows the mean coverage of each assembly and the percentage of reads that assembled to the plastome reference. The enrichment efficiency across the 24 samples (i.e., the percentage of reads that assembled to the plastid genome) was on average 59%. The mean plastome coverage, averaged across the 24 species sequenced, was 717×. Examination of the coverage graphs superimposed on the annotated assemblies revealed that the sequence depth is generally nonuniform across the genome, with large spikes in depth clearly present at the coding regions (Figs. 1 and 2). This general pattern, evident across all 24 assemblies, is particularly pronounced in the species more distantly related to those included in the probe design (Figs. 1 and 2). These coverage spikes are also generally accompanied by tails of decreasing depth on either side, usually around 150–400 bp in length, roughly corresponding to the insert sizes of the libraries sequenced.

DISCUSSION

Constructing large data sets of complete (or nearly so) plastid genomes is becoming increasingly feasible due to the ever-increasing sequencing capacities of NGS instruments, particularly the Illumina GAIIx and HiSeq 2000/2500, which currently allow for parallel sequencing of 12–16 (GAIIx) or 36–48 (HiSeq 2000/2500) plastid genomes from pooled, unenriched gDNA samples. Targeted enrichment strategies for the plastid genome offer a promising means of vastly increasing the number of plastid genomes that can be sequenced in parallel, which in turn would dramatically decrease per-sample sequencing costs and increase the accessibility of plastid genome sequencing for routine phylogenetic as well as population and phylogeographic studies. The enrichment approach described in this paper shows considerable promise as a relatively simple and universal means of plastid genome enrichment (across eudicots and monocots, and potentially all angiosperms), making large-scale sequencing of angiosperm plastid genomes a more cost-effective (and therefore broadly accessible) practice.

***Increasing the limits of parallel plastome sequencing*—**A sequencing depth of ~30–50× is recognized as the minimum threshold needed for high-quality assembly of plastid genomes (Straub et al., 2012). Based on the mean coverage obtained across the 24 samples included in this study (717×), it should be theoretically possible to multiplex as many as 344 samples on a single lane of the Illumina GAIIx to obtain ~50× coverage following plastid enrichment using the probe set described here. By coupling this enrichment strategy with the even higher sequencing capacity of the HiSeq 2000 or 2500—which can yield ~187 500 000 reads per lane in a single run (Glenn, 2011)—we estimate that it should be possible, theoretically, to multiplex up to ~1300 samples and still obtain ~50× coverage of the plastid genome (given that the capacity of the HiSeq is roughly four times that of the GAIIx). This method of plastid enrichment therefore substantially increases the number of plastomes that can be sequenced in parallel. However, given the difficulties of pooling numerous samples proportionally, attempts to multiplex ~300 or more samples would probably lead to considerable variation in read numbers obtained per library. Additionally, the number of barcodes available in current adapter sets is limited (e.g., up to 96 in the NEXTflex DNA Barcode kit, Bioo Scientific, Austin, Texas, USA), and designing/purchasing adapter sets with more than 300 barcodes might be prohibitively expensive. Therefore, we suggest 96 samples as a current practical maximum for plastome multiplexing using this targeted enrichment method, but we encourage approaches to expand

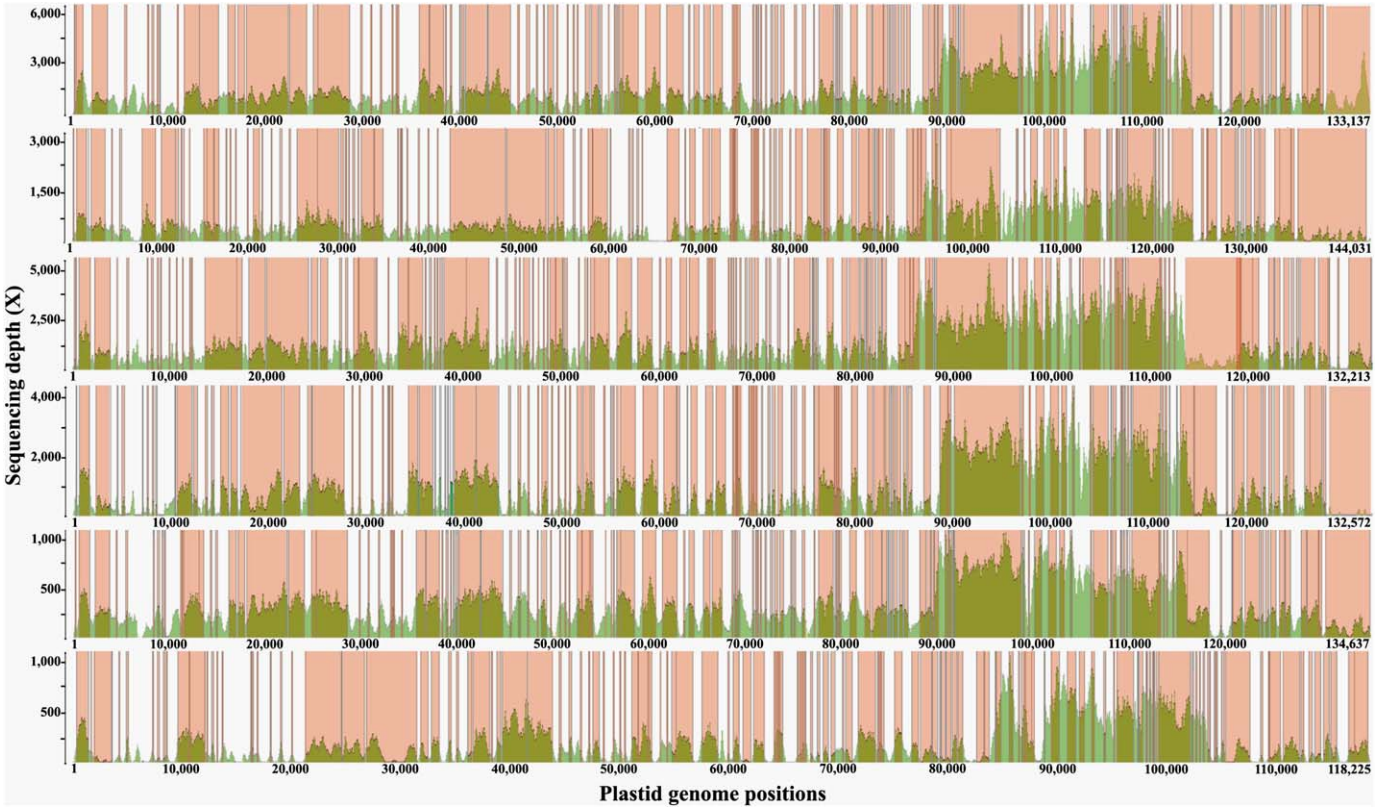

Fig. 1. Coverage graphs for six species included in this study, representing, from top to bottom, increasing phylogenetic distance from the taxa included in the probe design. From top to bottom, the species (and their closest relation to taxa included in the probe design) are: *Cucumis sativus* (same species), *Oenothera hartwegii* (same genus), *Dicranocarpus parviflorus* (same family), *Acleisanthes lanceolata* (same order), *Mentzelia perennis* (same order), *Sporobolus nealleyi* (monocot; outside the probe set's target clade). The coding regions are highlighted in red to show sequence depth obtained for coding vs. noncoding regions.

the number of multiplexed samples beyond 96, particularly to take advantage of the capacity of the HiSeq and other newer instruments that will continue to grow sequencing capacity.

Utility of the probe set—The enrichment strategy described here represents the first attempt to design a plastid probe set across a phylogenetically diverse set of samples (22 eudicot plastomes), making it broadly applicable for angiosperm plastid genome sequencing. This approach proved highly successful in recovering complete to essentially complete plastomes with impressively high coverage across most taxa tested, including monocots (Table 2). However, the depth of coverage was consistently uneven across the genome, with considerable spikes in sequence depth evident at the coding regions (Figs. 1 and 2). Because in many cases we had conspecific references available for plastome assembly, we believe this pattern reflects actual differences in depth of coverage across the genome, rather than an artifact of poor assembly due to a divergent reference. Several studies (Gnirke et al., 2009; Mamanova et al., 2010; Cronn et al., 2012; Lemmon et al., 2012) have demonstrated the importance of relatively long insert lengths for recovering more rapidly evolving (and hence divergent) spacer regions, which are usually flanked by more conserved genes that are more likely to hybridize with baits (Lemmon et al., 2012). Studies requiring more variable portions of the plastome (e.g., shallow phylogenetic and phylogeographic investigations) should therefore consider targeting relatively large insert sizes to increase the sequence coverage of these variable regions when using this plastid enrichment approach. In our study, we targeted 200–300-bp inserts, resulting in tails of decreasing sequence depth ~200–300 bp long on either side of the coverage spikes at the coding regions. Larger inserts would proportionally increase the span of the depth tails flanking the coding regions, thus capturing spacer/intronic regions with greater coverage.

Although the probe set outlined here shows immediate promise for essentially complete plastome sequencing in eudicots and monocots (which collectively represent >95% of angiosperm diversity), we anticipate that its applicability should extend to Magnoliidae, Chloranthaceae, and basal angiosperm lineages (Amborellaceae, Nymphaeales, Austrobaileyales), given that many of the probes target highly conserved coding regions of the plastid genome. However, at increasing phylogenetic distances from eudicots, the probe set will likely recover only the more conserved plastid regions, leaving behind the spacers and rapidly evolving regions useful for species- or population-level investigations (unless relatively large inserts are targeted for enrichment and sequencing). For example, Cronn et al. (2012) showed that probes designed from a single species of *Pinus* (*P. thunbergii*) could be used to enrich conserved plastid regions (i.e., those with >80% pairwise sequence identity) in a very distantly related angiosperm species (*Gossypium raimondii*). These results demonstrate that plastid probes can be successfully used for targeted enrichment (of at least highly conserved regions) across extensive phylogenetic distances.

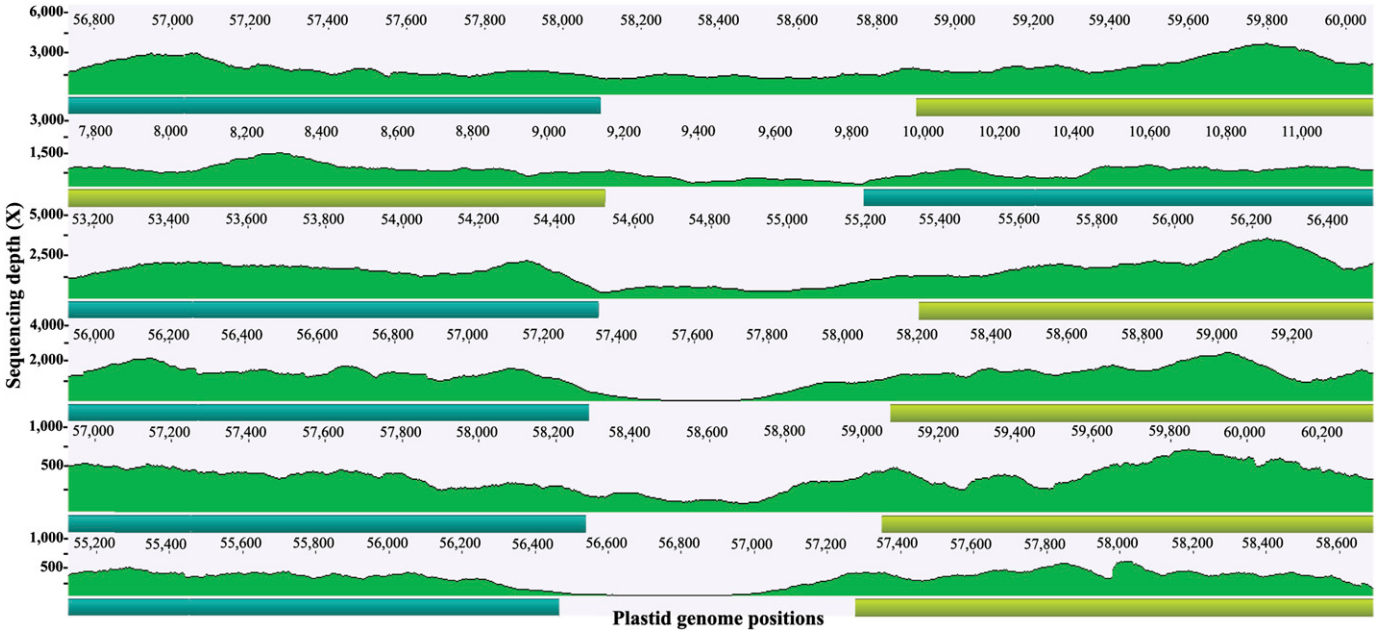

Fig. 2. Close-up of the *atpB-rbcL* spacer, from the same six species shown in Fig. 1, highlighting differences in sequence depth obtained for coding vs. noncoding regions. As in Fig. 1, the phylogenetic distance from taxa included in the probe set increases from top to bottom. The coding regions *atpB* and *rbcL* are indicated by the blue and yellow bars, respectively.

Considerations for multiplex sequencing—The low overall coverage obtained for some of the 24 libraries sequenced for this experiment is probably due to uneven pooling of libraries prior to hybridization enrichment. No phylogenetic pattern is evident in those taxa that had low coverage, and fairly close relatives of these low-coverage samples had much higher coverage. For example, *Sarcobatus* and *Acleisanthes* had extremely high coverage using SureSelect, whereas *Petiveria* had low coverage; all three taxa belong to the clade of Phytolaccaceae + Nyctaginaceae, and all have similar genome structures. When multiplexing large numbers of libraries, even small errors in DNA quantification can lead to significant differences in read numbers that can be compounded by the additional enrichment step after hybridization. Hence, it is crucial to quantify DNA concentration accurately in each library prior to pooling. Multiple methods are possible, including Bioanalyzer (Agilent), the Qubit 2.0 fluorometer (Life Technologies, Grand Island, New York, USA), and quantitative real-time PCR (qPCR). Because qPCR simultaneously amplifies and quantifies DNA samples, it more accurately quantifies the "sequenceable" portion of the library (i.e., the amount of DNA with successfully ligated adapters) and is thus the most accurate method overall; the Bioanalyzer and the Qubit, on the other hand, determine the total quantity of DNA in the sample regardless of adapter ligation. Likewise, fewer cycles should be used to amplify the plastid-enriched library pool. In the experiment outlined here, we used 18 cycles to amplify the 24-plex capture; this might have exacerbated the unequal enrichment of the library pool and consequently led to disparities in the number of reads obtained from each sample in the sequencing run.

Alternative sequencing strategies—Although the method outlined here represents an excellent means of large-scale plastid genome sequencing with great potential for plant phylogenetics and phylogeography, it by no means displaces the importance of alternative NGS strategies in the plant systematics community. For example, genome skimming (also known as genome survey sequencing), which involves low-coverage sequencing of whole-genomic samples, is an effective approach for recovering complete to essentially complete plastid genomes (up to ~48 on a single HiSeq 2000/2500 lane), as well as partially complete mitochondrial genomes and a wealth of nuclear data (Straub et al., 2012; Steele et al., 2012). This method is attractive in that it yields data from all three plant genomes for phylogeny reconstruction without the extra effort/cost associated with targeted enrichment, but it is important to note that, at present, considerably fewer samples can be sequenced in parallel with genome skimming compared to enrichment-based approaches, especially when using the GAIIx instrument. Moreover, only the high-copy nuclear elements (e.g., the rDNA cistron) are usually sequenced with >5× coverage in multiplex genome skimming. The shallow coverage obtained for low-copy nuclear regions may be sufficient for PCR primer design or probe development (for nuclear targeted enrichment) but generally precludes both the determination of orthology/paralogy and the immediate use of these regions in phylogenetic analysis. Targeted nuclear enrichment—employing baits designed to capture hundreds of single/low-copy nuclear loci—represents another promising yet underexplored NGS method for plant systematics. Lemmon et al. (2012) demonstrated how genomic resources could be used to develop a nuclear probe set with utility across vertebrates—a vast phylogenetic distance including ~500 million years of evolutionary history. Using available genomic or transcriptomic resources (e.g., the 1KP dataset: http://www.onekp.com/), similar probe sets could be developed for major plant clades, allowing for the recovery of hundreds of unlinked nuclear loci across hundreds of multiplexed samples.

These three alternative strategies—plastid enrichment/sequencing, genome skimming, and nuclear enrichment/sequencing—all have advantages and disadvantages related to

their cost, time investment, and data output. Although the extra time and effort required for the hybridization-enrichment step is relatively minor compared to the effort required for gDNA library preparation, targeted enrichment kits (e.g., Agilent SureSelect, Roche Nimblegen, MYcroarray) are a somewhat costly investment. Therefore, plastid genome hybridization enrichment will be most efficient in terms of time and money for projects that involve sequencing of hundreds of plastid genomes. For smaller-scale phylogenetic projects, genome skimming remains an excellent and relatively cost-effective means of multiplexing plastid genomes. The increasing availability of nuclear genomic resources makes the development of probe sets for nuclear enrichment a viable and promising NGS strategy, with potential for large-scale sequencing of hundreds of independent nuclear loci. This study and others (Cronn et al., 2012; Lemmon et al., 2012) highlight the general effectiveness of hybridization-based enrichment across relatively large phylogenetic distances, offering promise for the development of nuclear probe sets for major plant clades. Researchers should carefully consider these points and others (Cronn et al., 2012; Steele et al., 2012; Straub et al., 2012; Lemmon et al., 2012) when deciding which sequencing strategy best suits the budget and data requirements of their phylogenetic and phylogeographic studies.

LITERATURE CITED

Arakaki, M., P.-A. Christin, R. Nyffeler, A. Lendel, U. Eggli, R. M. Ogburn, E. Spriggs, et al. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proceedings of the National Academy of Sciences, USA* 108: 8379–8384.

Brockington, S. F., R. Alexandre, J. Ramdial, M. J. Moore, S. Crawley, A. Dhingra, K. Hilu, et al. 2009. Phylogeny of the Caryophyllales sensu lato: Revisiting hypotheses on pollination biology and perianth differentation in the core Caryophyllales. *International Journal of Plant Sciences* 170: 627–643.

Cantino, P. D., J. A. Doyle, S. W. Graham, W. S. Judd, R. G. Olmstead, D. E. Soltis, P. S. Soltis, and M. J. Donoghue. 2007. Towards a phylogenetic nomenclature of *Tracheophyta. Taxon* 56: 822–846.

Craig, D. W., J. V. Pearson, S. Szelinger, A. Sekar, M. Redman, J. J. Corneveaux, T. L. Pawlowski, et al. 2008. Identification of genetic variants using bar-coded multiplexed sequencing. *Nature Methods* 5: 887–893.

Cronn, R., A. Liston, M. Parks, D. S. Gernandt, R. Shen, and T. Mockler. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* 36: e122.

Cronn, R., B. J. Knaus, A. Liston, P. J. Maughan, M. Parks, J. V. Syring, and J. Udall. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99: 291–311.

Drinnan, A. N., P. R. Crane, and S. B. Hoot. 1994. Patterns of floral evolution in the early diversification of non-magnoliid dicotyledons (eudicots). *Plant Systematics and Evolution* 8(Supplement): 93–122.

Glenn, T. C. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11: 759–769.

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Broockman, T. Fennel, et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.

Jansen, R. K., Z. Cai, L. A. Raubeson, H. Daniell, C. W. dePamphilis, J. Leebens-Mack, K. F. Müller, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA* 104: 19369–19374.

Jian, S., P. S. Soltis, M. Gitzendanner, M. Moore, R. Li, T. Hendry, Y. Qiu, et al. 2008. Resolving an ancient, rapid radiation in Saxifragales. *Systematic Biology* 57: 38–57.

Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61: 727–744.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar, E. Howard, et al. 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods* 7: 111–118.

Moore, M. J., A. Dhingra, P. S. Soltis, R. Shaw, W. G. Farmerie, K. M. Folta, and D. E. Soltis. 2006. Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology* 6: 17–30.

Moore, M. J., C. D. Bell, P. S. Soltis, and D. E. Soltis. 2007. Using plastid genomic-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* 104: 19363–19368.

Moore, M. J., P. S. Soltis, C. D. Bell, J. G. Burleigh, and D. E. Soltis. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA* 107: 4623–4628.

Njuguna, W., A. Liston, R. Cronn, T.-L. Ashman, and N. Bassil. 2013. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Molecular Phylogenetics and Evolution* 66: 17–29.

Parks, M., R. Cronn, and A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7: 84.

Parks, M., R. Cronn, and A. Liston. 2012. Separating the wheat from the chaff: Mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). *BMC Evolutionary Biology* 12: 100.

Soltis, D. E., P. S. Soltis, P. K. Endress, and M. W. Chase. 2005. Phylogeny and evolution of the angiosperms. Sinauer, Sunderland, Massachusetts, USA.

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.

Straub, S. C. K., M. Parks, K. Weitemier, M. Fishbein, R. C. Cronn, and A. Liston. 2012. Navigating the tip of the genomics iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Wang, H., M. J. Moore, P. S. Soltis, C. D. Bell, S. Brockington, R. Alexandre, C. C. Davis, et al. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences, USA* 106: 3853–3858.

Whittall, J. B., J. Syring, M. Parks, J. Buenrostro, C. Dick, A. Liston, and R. Cronn. 2010. Finding a (pine) needle in a haystack: Chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology* 19(Supplement): 100–114.

Wyman, S. K., R. K. Jansen, and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

Zerbino, D. R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18: 821–829.

Conserv Genet
DOI 10.1007/s10592-013-0527-2

RESEARCH ARTICLE

Evidence of population bottleneck in *Astragalus michauxii* (Fabaceae), a narrow endemic of the southeastern United States

Wade A. Wall · Norman A. Douglas · William A. Hoffmann · Thomas R. Wentworth · Janet B. Gray · Qiu-Yun Jenny Xiang · Brian K. Knaus · Matthew G. Hohmann

Received: 7 January 2013 / Accepted: 26 August 2013

Abstract Genetic factors such as decreased genetic diversity and increased homozygosity can have detrimental effects on rare species, and may ultimately limit potential adaptation and exacerbate population declines. The Gulf and Atlantic Coastal Plain physiographic region has the second highest level of endemism in the continental USA, but habitat fragmentation and land use changes have resulted in catastrophic population declines for many species. *Astragalus michauxii* (Fabaceae) is an herbaceous plant endemic to the region that is considered vulnerable to extinction, with populations generally consisting of fewer than 20 individuals. We developed eight polymorphic microsatellites and genotyped 355 individuals from 24 populations. We characterized the population genetic diversity and structure, tested for evidence of past bottlenecks, and identified evidence of contemporary gene flow between populations. The mean ratios of the number of alleles to the allelic range (*M* ratio) across loci for *A. michauxii* populations were well below the threshold of 0.68 identified as indicative of a past genetic bottleneck. Genetic diversity estimates were similar across regions and populations, and comparable to other long-lived perennial species. Within-population genetic variation accounted for 92 % of the total genetic variation found in the species. Finally, there is evidence for contemporary gene flow among the populations in North Carolina. Although genetic factors can threaten rare species, maintaining habitats through prescribed burning, in concert with other interventions such as population augmentation or (re)introduction, are likely most critical to the long term survival of *A. michauxii*.

Keywords *Astragalus* · Bottleneck · Endemism · Genetic diversity · Microsatellites · *Pinus palustris*

Electronic supplementary material The online version of this article (doi:10.1007/s10592-013-0527-2) contains supplementary material, which is available to authorized users.

W. A. Wall (✉) · N. A. Douglas · W. A. Hoffmann · T. R. Wentworth · Q.-Y. J. Xiang
Department of Plant Biology, North Carolina State University, Box 7612, Raleigh, NC 27695, USA
e-mail: Wade.A.Wall@usace.army.mil

W. A. Wall · M. G. Hohmann
US Army Corps of Engineers, Engineer Research and Development Center, P. O. Box 9005, Champaign, IL 61826, USA

N. A. Douglas
Department of Biology, Oberlin College & Conservatory, Science Center K107, 119 Woodland St., Oberlin, OH 44074-1097, USA

J. B. Gray
Endangered Species Branch, Fort Bragg, NC 28310, USA

B. K. Knaus
Department of Botany and Plant Pathology, Oregon State University, 2082 Cordley Hall, Corvallis, OR 97331, USA

Introduction

Land use changes over the last 500 years have led to habitat loss and population isolation for many species, leading to a greater concern about the loss of biodiversity and its effects on the biosphere (Balmford and Bond 2005). In the southeastern United States, many terrestrial and aquatic species are threatened with extinction (Dobson et al. 1997). Many of the rare plant species of the coastal plain are part of the longleaf pine (*Pinus palustris* Mill.) ecosystem, an assemblage of fire-dependent communities

Published online: 13 October 2013

dominated by temperate savanna that covered roughly 37 million hectares from east Texas to Florida to southeastern Virginia (Frost 1993). The longleaf pine ecosystem depends on frequent fires to maintain an open understory and facilitate longleaf pine seedling recruitment. Urbanization, fire suppression, and conversion to agriculture have fragmented the original longleaf pine ecosystem and reduced it to 2 % of its former area (Frost 2006), resulting in population declines for many endemic species (Van Lear et al. 2005).

Anthropogenic habitat fragmentation in the southeastern United States has occurred against a background of past, highly dynamic, climatically driven vegetation change during the transition from the Pleistocene to the Holocene (Webb III 1987; Williams et al. 2004). Thus, possible explanations for the current habitat fragmentation of some southeastern plant populations include, in the short term, changes in land use following European settlement and, in the long term, climatic changes that have occurred in the past 20,000 years. Habitat fragmentation results in increased genetic isolation and smaller plant populations (Young et al. 1996), and with increasing time since fragmentation, genetic differentiation between populations generally increases and genetic diversity within populations decreases (Young et al. 1996). If populations have been isolated since the early Holocene, it is assumed that genetic differentiation between populations would be greater relative to that due to population isolation since European settlement.

Extreme reductions in population size can lead to genetic bottlenecks, which are of conservation concern because of the increased risk of extinction (Frankham 2005). Across taxa, it has been observed that genetic diversity is lower in threatened species (Spielman et al. 2004) and in species with restricted ranges (Hamrick and Godt 1989) relative to common, widespread species. This suggests that genetic factors increase the likelihood of extinction of species by reducing fitness of individuals within small populations (Leimu et al. 2006) and restrict evolutionary potential (Franklin 1980). Thus, assessing and maintaining the genetic variation found within rare species is of concern for both their short- and long-term viability and is one of the cornerstones of conservation genetics (Frankham et al. 2010).

Astragalus michauxii (Kuntze) F.J. Herm. is a rare legume endemic to the Fall-line Sandhills region of North Carolina, South Carolina, and Georgia (USA). Only a few species of *Astragalus* are represented east of the Mississippi. Most members of this genus occur in Asia, western North America, and South America, with the North American species belonging to the aneuploid Neo-Astragalus clade, which began to diversify 4.4 Ma (Scherson et al. 2008). The disjunct species in eastern North America may be relicts from past vegetation assemblages of Pliocene or Pleistocene origin (Noss 2013). In South Carolina and Georgia, all known populations of *A. michauxii* exist in a highly fragmented landscape and are separated from the North Carolina populations by over 200 km. The average population size is less than 20 individuals (Wall et al. 2012) and there is substantial risk that these small, isolated populations may be subject to reduced genetic diversity and the negative effects that ensue.

In this study, we investigated genetic diversity and structure of *A. michauxii* populations across the range of the species using eight microsatellite loci. We estimated the level of genetic variation within and among populations and identified the most likely number of genetic clusters within *A. michauxii*. In addition, we searched for evidence of past genetic bottlenecks and contemporary gene flow. We hypothesized that the small, isolated populations would exhibit strong genetic differentiation between populations and low genetic diversity within populations, as well as evidence of genetic bottlenecks because of recent land use changes following European settlement. By examining the genetic variation of *A. michauxii* within the context of past climatic change and land use history, this study provides useful information for any future conservation or restoration efforts of this species and other rare plants of the Fall-line Sandhills.

Methods

Species

Astragalus michauxii is an herbaceous, long-lived legume that is generally found in upland longleaf pine habitat in what has been characterized as the pine/scrub oak sandhill community (Schafale and Weakley 1990). The largest extant populations are found in the loamy soil variant of this community type that generally have higher pH and more nutrients. These areas are known locally as "pea swales" or "bean dips" because of their high diversity of Fabaceae species (James 2000). The species is largely restricted to the Fall-line Sandhills of North Carolina, South Carolina, Georgia, and Alabama (USA) (Sorrie and Weakley 2001; Peet 2006; NatureServe 2012), an extensive ancient dune system located at the boundary between the Coastal Plain and the Piedmont physiographic regions that is characterized by a rolling topography with excessively-well drained, sandy soils in the interfluvial areas.

Astragalus michauxii is one of the earliest flowering legumes in the Fall-line Sandhills, flowering in early May and producing mature fruits by early July (Radford et al. 1968). The species is most likely an obligate outcrosser,

and flowers are pollinated by a variety of flying insects. Fruit set is commonly quite low relative to the number of flowers; however the reasons for this have not been identified (e.g. an overabundance of self pollen, pollinator limitation, florivory, etc.). Like most legumes, *A. michauxii* produces a seed bank, but seed density appears to be rather low (Weeks 2005); possible factors include low seed set and pre- or post-dispersal seed predation. Seed dispersal is probably highly limited, since the seeds lack any obvious dispersal mechanism. Based on four years of observations across 39 populations, only one recruit was observed across all the monitored populations (Wall et al. 2012). Although census sizes are currently stable due to high adult survivorship, the probability of long-term positive population growth rates may be low because of the lack of recruitment (Wall et al. 2012).

NatureServe has given *A. michauxii* a global status of G3 (NatureServe 2012), indicating that the species is considered vulnerable to extinction due to a restricted range with few populations. In Georgia it has an S2 ranking, indicating that it is present in 20 or fewer populations and considered imperiled. *Astragalus michauxii* has an S3 ranking in North Carolina and South Carolina, indicating that the species is considered vulnerable to extinction with typically 21–100 populations and 3,000–5,000 individuals in total. In South Carolina, *A. michauxii* is only known from two populations. In North Carolina most of the populations are found on several public land holdings, mainly on Fort Bragg Military Reservation and the state-managed Sandhills Gamelands, with varying degrees of isolation between populations.

Sampling and population genetic methods

During June 2009–2010, leaves were collected from individuals located in 22 populations (as defined by the North Carolina Natural Heritage Program) on Fort Bragg and Camp Mackall Military Reservations (North Carolina) and from two populations in Georgia (Burke and Candler Counties) (Fig. 1). Leaf samples were stored in a −80 °C freezer until extraction. DNA was extracted from 355 individuals across the 24 populations, using the CTAB method with minor modifications (Doyle and Doyle 1987). This sampling included 80 % of individuals within the populations. Possible polymorphic microsatellite regions were identified using a recently published protocol (Jennings et al. 2011). Briefly, DNA was first sheared using a Bioruptor sonicator (Diagenode Inc., Denville, NJ, USA) and barcoded Illumina DNA libraries were created (Cronn et al. 2008). Libraries were enriched for microsatellites using hybridization with three probes containing repeated dinucleotide motifs. After hybrid capture, the microsatellite-enriched libraries were quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Libraries were then pooled and paired-end sequenced on one lane of an Illumina Genome Analyzer II. The resulting microreads were then sorted by the 4 bp barcodes and searched for dinucleotide motifs, (located near the center of reads to optimize primer development). After filtering redundant reads, sequences were analyzed using BatchPrimer3 to identify PCR primer sites (You et al. 2008).

Microsatellite-containing sequences were screened using agarose gels, and eight polymorphic loci were identified. 6 μL multiplexed PCR reactions were performed using fluorescently-labeled forward primers (6-FAM, HEX, NED) as follows: 3 μL Qiagen multiplex PCR master mix (Qiagen, Hilden, Germany), 0.6 μL Q-Solution, 0.8 μL H_2O, 0.6 μL multiplexed primer pair mix, and 1.0 μL diluted (1:8 DNA:H_2O) DNA. PCR cycling conditions were 95 °C for 15 min; 45 cycles at 94 °C for 30 s, 58 °C for 1 min 30 s., and 72 °C for 1 min; 60 °C for 30 min. PCR products were genotyped on an ABI 3730

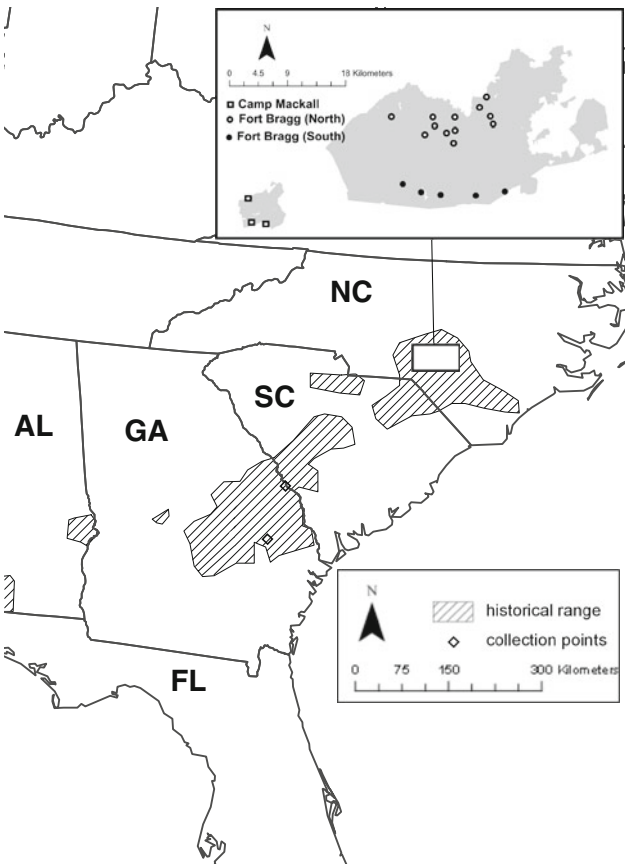

Fig. 1 Historic range and collection sites of *Astragalus michauxii*. Historical range determined based on voucher specimens (UNC Herbarium Flora of the Southeast; http://www.herbarium.unc.edu/seflora). Current range is greatly restricted, with most populations in North Carolina. Survey of Georgia populations only located 13 individuals

sequencer (Applied Biosystems, USA). PCR was performed again for loci that failed to amplify. After a second failed amplification, a locus was marked as missing for that sample. Genotyping was performed using GeneMarker 1.8 (Softgenetics LLC, State College, PA, USA). Exported peak heights were binned using TANDEM (Matschiner and Salzburger 2009). We genotyped 5 % of the individuals twice in order to assess data quality. Consistency across all loci in duplicate samples was 95.5 %.

Genetic structure and diversity

Hardy–Weinberg equilibrium and linkage disequilibrium among loci are assumptions in a number of different genetic analyses. We tested for significant departures from Hardy–Weinberg equilibrium (HWE) in Arlequin 3.5 (Excoffier and Lischer 2010) using a test analogous to Fisher's exact test (Guo and Thompson 1992), whereby the marginal probabilities of the observed allele frequencies are compared to values from simulated data sets explored using a Markov Chain. We performed the Markov Chain with 1,000,000 iterations and a burn-in of 100,000 iterations. We tested for linkage disequilibrium for the eight loci within individual populations in Arlequin 3.5 using a likelihood ratio test (Slatkin and Excoffier 1996), with no assumption of linkage equilibrium. Likelihood was calculated using the expectation–maximization (EM) algorithm (Dempster et al. 1977) to estimate allele frequencies. We performed 10,000 permutations with an initial EM value of 2. Since adjustments for multiple comparisons can make it difficult to identify significance, even when differences exist (Moran 2003), we assessed evidence of linkage disequilibrium using p-values both unadjusted and adjusted by the sequential Bonferroni correction method for multiple comparisons (Holm 1979). We calculated the average number of alleles and absolute number of private alleles (i.e. those found in a single population) for each population. Because populations consisted of varying numbers of individuals, we used a rarefaction method (Kalinowski 2004) for calculating allelic richness and private allelic richness available in HP-Rare 1.1 (Kalinowski 2005). We calculated expected and observed heterozygosity (Nei 1987) using Arlequin 3.5.

We explored the genetic structure of *A. michauxii* populations using the Bayesian clustering algorithm implemented in the software program BAPS 5.2 (Corander et al. 2003, 2004) to infer the number of genetic clusters (K). We used the spatial clustering option, as it has been shown to provide superior results if the genetic data are sparse (Corander et al. 2008). We performed 100 runs in fixed mode for each value of K from one to 24 and assessed admixture within the identified genetic clusters. We also explored the genetic structure of *A. michauxii* using STRUCTURE 2.3.4 under an admixture model with correlated allele frequencies with a burnin period = 100,000, 750,000 MCMC iterations after burnin, and averaged over three runs. STRUCTURE results are available in the supplemental material. We performed an analysis of molecular variance (AMOVA) (Excoffier et al. 1992) as implemented in Arlequin 3.5, to quantify the variance found within the BAPS-identified populations. We also used AMOVA to assess the overall significance of the defined groups. For comparative purposes, we estimated F_{ST} as calculated in Arlequin 3.5; the estimate is identical to the weighted θ (Weir and Cockerham 1984) and is the variance component among groups and the variance component among populations divided by the total variance. Population structure was further visualized graphically by performing a non-metric multi-dimensional scaling (NMDS) ordination using the R (R Development Core Team 2012) package labdsv (Roberts 2010). Pairwise population genetic distances were calculated using Nei's unbiased D (Nei 1978) as implemented in GenAlex 6.5 (Peakall and Smouse 2006). To test for isolation by distance (IBD), we performed a Mantel test using the R package vegan (Oksanen et al. 2009), using the log-transformed geographic distances (to improve normality) and the pairwise population genetic distances (Nei's unbiased D).

Estimating gene flow between populations

To detect possible first generation migrants between populations we used GeneClass2 (Piry et al. 2004), an implementation of the Bayesian method of Rannala and Mountain (1997). For each individual we computed the probability that it was a recent emigrant from another population (i.e. that the identified individual's allele frequencies are more similar to another population than its resident population). We used the simulation algorithm of Paetkau et al. (2004) with the following settings: type I error threshold of 0.01, 1,000 simulated individuals, and all loci included. Gene flow among populations may follow a source-sink metapopulation model, with gene flow from large populations to small populations. To test whether emigration was correlated with population size (i.e. whether larger populations contribute disproportionately to the number of migrants), we performed a logistic regression with emigration as the response variable and population size as the explanatory variable. We tested for significant differences in population size for resident and assigned populations of the identified migrants by simulating 10,000 replicate data sets with replacement and calculating the 95 % confidence interval for mean population size for each group of populations (resident or assigned).

Evidence of genetic bottlenecks across multiple temporal scales

To detect the genetic imprint of recent bottlenecks in *A. michauxii*, we used the software package BOTTLENECK 1.2.02 (Piry et al. 1999), an implementation of the method described by Cornuet and Luikart (1996). We fit two models using the program: the stepwise mutation model (SMM), and a two-phase model (TPM) mutation model with 95 % of the mutations single-step and a variance of 12 (Piry et al. 1999). For both models, 10,000 datasets were simulated with same observed number of alleles and population sample sizes under mutation-drift equilibrium, with the p value calculated as the probability of obtaining the mean expected heterozygosity (H_e) from the observed data based on the distribution of expected mutation-drift heterozygosity (H_{eq}) values. We used a one-tailed Wilcoxon signed rank test to detect significant heterozygote excess in populations because this test is most appropriate when the sample size is less than 30 and the number of loci is less than 10. We only tested for evidence of recent bottlenecks in populations that had >20 gene copies ($N = 10$ for diploid individuals).

To test for bottleneck events that may have occurred over longer time periods (>100 generations), we used the M ratio test (Garza and Williamson 2001) as implemented in Arlequin 3.5. This implementation of the M ratio test excludes monomorphic loci because these can erroneously increase the M ratio. The M ratio is the mean number of alleles in a population divided by the allelic size range. When alleles are lost from a population, the number of alleles decreases at a faster rate than the allelic size range, so small (<0.68) M ratio values are indicative of populations that have gone through a genetic bottleneck at some time in the past. We estimated 95 % confidence intervals by resampling the M ratio estimates for each locus within a population with replacement 10,000 times (Swatdipong et al. 2009). We also calculated a one-sided 95 % confidence interval (M_c) for each population. M_c is a value estimated through 10,000 simulations of a population at equilibrium such that M_c is less than the simulated M values in 95 % of the simulations based on the mean size of non-stepwise mutations and θ. Settings were mean size of non-stepwise mutations = 3.5, θ = 10, and 10,000 iterations, as recommended by Garza and Williamson (2001).

Results

Microsatellite development generated 100 candidate loci, from which we developed eight polymorphic microsatellite loci that amplified consistently and produced no more than two bands in our initial screens (Table 1). Allele frequencies detected significant departures from Hardy–Weinberg expectations in only 19 out of 176 tests at the population level (22 populations * 8 loci); however, our power to detect significant departures was low because of small sample sizes in many of the populations. The northern populations on Fort Bragg had more loci out of HWE compared to populations from other regions. Microsatellite marker AM29 was out of HWE in eight of 22 populations and three of four regions, indicating that this locus may be influenced by null alleles or allelic dropout. After correcting for multiple tests, none of the loci were significantly out of HWE ($p > 0.05$). Linkage disequilibrium (LD) was detected in 91 of 616 tests representing all possible pairwise loci combinations within populations, although only three comparisons were significant after using sequential Bonferroni adjustment. As with testing Hardy–Weinberg equilibrium, our power to detect LD was low because of small sample sizes. Significant evidence of LD was not isolated to pairs of loci, but rather was encountered in all loci compared and was never greater than 30 % (Table S1), suggesting that the observed LD is due to population-level effects such as null alleles, admixture, inbreeding, or genetic drift due to a bottleneck event, rather than physical linkage between loci.

Genetic diversity in *A. michauxii* averaged 10.88 alleles per locus across all populations, with larger populations having a greater number of alleles compared to smaller populations ($R^2 = 0.78$, $p < 0.001$; Table 2). After adjusting for sample size, allelic richness was similar across populations ($R^2 = 0.11$, $p = 0.07$). Allelic richness ranged from 2.43 to 3.80 across populations. As with allele number, the number of private alleles increased with population size ($R^2 = 0.39$, $p = 0.001$), but after rarefaction there was not a significant correlation ($R^2 = 0$, $p > 0.05$).

Genetic clustering indicated the highest posterior probability occurred when the populations were grouped into three clusters of populations. (Table S2 in supplementary material). These clusters combined the Camp Mackall and southern Fort Bragg populations into one cluster, combined the northern and northeastern Fort Bragg populations into another cluster, and separated the two Georgia populations into a distinct cluster (Fig. 2; Fig. S2 in supplemental material includes BAPS and STRUCTURE results for $K = 2$–9). AMOVA results indicated that within population genetic variation accounted for 91.3 % of the total genetic variation (Table 3), with the genetic clusters (K) identified by BAPS 5.2 accounting for 2.9 % of the overall genetic variation. F_{ST} was estimated at 0.08. When populations were grouped according to geographical location, non-metric multidimensional scaling graphically demonstrated the modest population differentiation among *A. michauxii* populations (Fig. 3; Table S3 in supplemental

Table 1 Eight polymorphic loci identified and developed for *Astragalus michauxii*

Primer	Sequence	Repeat	Fragment size (BP)	Alleles
AM_15	F: GTTTCACACTGAGACACAGTTC	GA	24–34 (124–134)	6
	R: AATTCCCAAGTGTAAAAGCTC			
AM_18	F: GAAAACACAAACAAATTCTGG	GA	8–38 (165–195)	13
	R: AGAAAGTCTGTGCTCTCTCATT			
AM_25	F: CAATCCCTAACCTTGAGTTCT	GA	8–36 (107–135)	14
	R: AGCAACGTGGGATAAAAATA			
AM_29	F: AACGGTGTCTGTGTCTATGTC	GT	32–42 (160–170)	6
	R: ATGAAGCGTTTCACATTTTT			
AM_34	F: TGACATACATGCTGAAAGTTG	AG	20–26 (155–161)	4
	R: TTTGGATTCATATAACCACCA			
AM_46	F: GAAAATGGTGGAAAAGGAAT	AG	18–64 (102–148)	22
	R: GTGTAAAAATCGTGCACTTCT			
AM_71	F: AAGATTGTCTAACGATCACCA	GT	187–203 (201 missing)	7
	R:AAAGCCCATGTTTCACTAAAT			
AM_91	F: GGACAAAAGAAGAGGAGAGAG	AG(TACTGG)TG	22–40 (107–125)	10
	R: TAAGTCGAGTTGTTCCAAAGT			

Columns are primer pair name, sequence, repeat, microsatellite range (with fragment size in parentheses), and number of alleles observed over all 22 populations

Table 2 Genetic variation in nineteen *Astragalus michauxii* populations from North Carolina and Georgia, as well as five local and regional geographic groupings, based on eight polymorphic microsatellite loci

Population	Region	N	A	A_R	P	P_R	H_o	H_e	H–W disequilibrium
ASMI053	Camp Mackall	7	3.43	2.61	1.00	0.11	0.55	0.53	None
ASMI054	Camp Mackall	14	4.88	3.23	0.00	0.07	0.59	0.62	AM29
ASMI088	Camp Mackall	5	3.13	2.91	1.00	0.14	0.51	0.55	None
ASMI020	Fort Bragg—North	5	3.13	2.95	0.00	0.12	0.55	0.55	None
ASMI022	Fort Bragg—North	7	4.00	3.42	1.00	0.14	0.60	0.69	AM29, AM46
ASMI023	Fort Bragg—North	27	5.50	3.38	1.00	0.04	0.58	0.66	AM29
ASMI032	Fort Bragg—North	30	6.13	3.48	0.00	0.03	0.62	0.65	None
ASMI034	Fort Bragg—North	24	6.63	3.80	2.00	0.11	0.56	0.71	AM18, AM25, AM29, AM91
ASMI035	Fort Bragg—North	8	4.25	3.36	0.00	0.05	0.58	0.66	AM18, AM25, AM29, AM91
ASMI049	Fort Bragg—North	13	4.50	3.36	0.00	0.00	0.63	0.64	None
ASMI056	Fort Bragg—North	68	7.75	3.63	2.00	0.10	0.62	0.69	AM29, AM34, AM71
ASMI057	Fort Bragg—North	33	5.00	2.93	1.00	0.02	0.47	0.56	AM18, AM46
ASMI061	Fort Bragg—North	6	2.86	2.43	0.00	0.01	0.62	0.55	None
ASMI091	Fort Bragg—North	12	5.00	3.56	0.00	0.07	0.67	0.69	None
ASMI096	Fort Bragg—North	4	3.50	3.50	0.00	0.12	0.69	0.62	None
ASMI030	Fort Bragg—South	9	4.25	3.29	0.00	0.07	0.57	0.62	None
ASMI046	Fort Bragg—South	11	3.88	3.04	0.00	0.02	0.56	0.58	None
ASMI050	Fort Bragg—South	34	5.75	3.19	3.00	0.08	0.58	0.61	AM29
ASMI051	Fort Bragg—South	24	6.00	3.32	1.00	0.10	0.54	0.59	AM29
ASMI097	Fort Bragg—South	5	3.00	2.81	0.00	0.09	0.48	0.53	None
AMBUGA	Georgia	4	3.25	3.25	0.00	0.03	0.56	0.65	None
AMMEGA	Georgia	5	3.63	3.34	0.00	0.12	0.55	0.66	None

Column headings are: *N*, number of individuals; *A*, average number of alleles across loci; A_R, average allelic richness; *P*, number of private alleles; P_R, private allelic richness; H_O, observed heterozygosity; H_e, expected heterozygosity; *H–W disequilibrium*, loci identified as not in Hardy–Weinberg equilibrium

material includes Nei's unbiased *D* pair-wise population distances). The two Georgia populations separated in ordination space from the North Carolina populations, but little separation occurred among the North Carolina populations. Isolation by distance (IBD) results indicated that the genetic distance between populations increased with

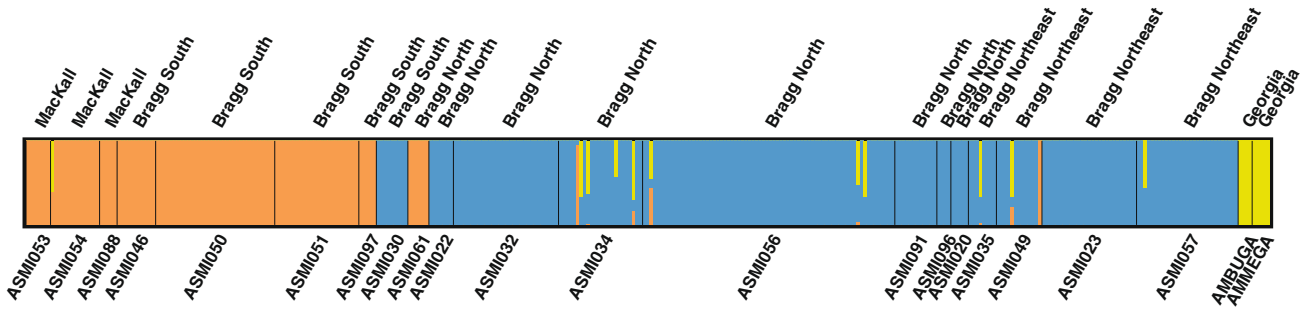

Fig. 2 Population genetic structure for *Astragalus michauxii* as determined by the program BAPS 5.2 using an admixture model with spatial information included and $K = 3$. The highest posterior probability was $K = 3$, with the Camp Mackall and southern Fort Bragg populations clustered and the northern Fort Bragg populations as a second cluster. The Georgia populations form the third cluster

Table 3 Analysis of molecular variance (AMOVA) results for *Astragalus michauxii* populations from North Carolina and Georgia (USA)

Groups (K) correspond to genetic clusters as identified using BAPS 5.2

Source of variation	*d.f.*	Sum of squares	Variance components	Percentage of variation	*p*-value
Among groups (K)	2	39.722	0.07809	2.88	<0.001
Among populations within regions	19	139.067	0.15686	5.78	<0.001
Within populations	688	1704.200	2.47703	91.34	<0.001
Total	709	1882.989	2.71198		

the log of the geographic distance ($r = 0.4553$, $p = 0.003$; Fig. S2 in supplemental material) when all populations were included. However, removal of the Georgia populations indicated that population genetic distance did not increase with the log of the geographic distance for North Carolina populations ($r = 0.05$, $p = 0.36$).

BOTTLENECK results did not indicate evidence of a recent genetic bottleneck in any population under the stepwise mutation model (SMM) or the two-phase mutation model (TPM) based on Wilcoxon's test (Table 4). Contrary to BOTTLENECK results, Critical *M* results indicated a severe bottleneck in *A. michauxii* populations. *M* ratio values across all populations averaged 0.48 and were lower than the 0.68 threshold identified by Garza and Williamson (2001) as indicative of a past genetic bottleneck, with the upper 95 % CI for all estimated *M* ratio values less than the 0.68 threshold in all populations (Fig. 4). The M_c values (90 % SMM) were greater than the observed *M* ratio values in 21 out of 22 populations and the M_c values (80 % SMM) were greater than the observed *M* ratio values in 14 out of 22 populations. Populations with M_c values less than the observed *M* ratio had smaller population sizes (and sample sizes) relative to populations with *M* ratio values less than the M_c values.

For the Fort Bragg and Camp Mackall populations, GeneClass2 results identified 14 putative first-generation migrants out of a total of 346 individuals (Table 5). The average distance between the population where the migrant

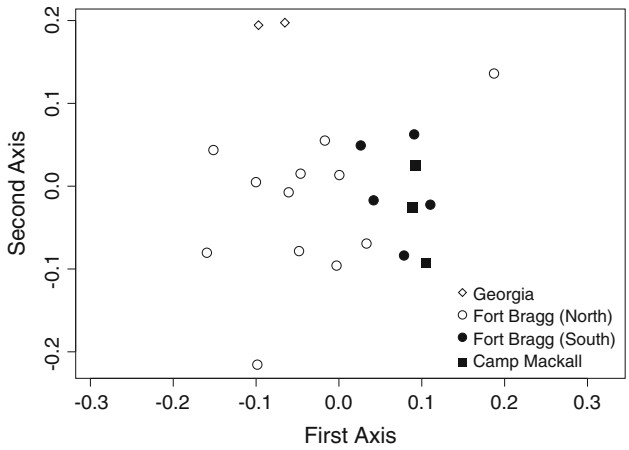

Fig. 3 Non-metric multidimensional scaling ordination of *Astragalus michauxii* population genetic distances based on eight polymorphic microsatellite loci. The Georgia populations appear separate from the North Carolina populations, while the North Carolina populations from Fort Bragg and Camp Mackall are more similar in terms of population genetic distance

was found ("sink") and the source population for individuals identified as migrants was 15.8 km. This distance was not statistically different than the average distance between all sampled North Carolina populations: 15.6 km. There was not a significant correlation between migration and population size, with the population sizes of "sink" populations similar to overall population sizes (18.5 ± 2.6 s.d. vs. 17.3 ± 3.4 respectively; $p > 0.05$).

Table 4 Tests for genetic bottlenecks in *Astragalus michauxii* using BOTTLENECK version 1.2 in populations with >20 gene copies and *M* ratio for all populations as calculated in Arlequin 3.1

Population	Copies	Wilcoxon's test (SMM)	Wilcoxon's test (TPM)	*M* ratio
ASMI020	9.25	–	–	0.43205
ASMI022	13.25	–	–	0.40397
ASMI023	53	0.62891	0.47266	0.49321
ASMI030	18	–	–	0.4812
ASMI032	59.5	0.875	0.76953	0.50957
ASMI034	46	0.84375	0.80859	0.49201
ASMI035	16	–	–	0.49813
ASMI046	22	0.27344	0.27344	0.4777
ASMI049	26	0.15625	0.125	0.48117
ASMI050	67	0.875	0.875	0.456
ASMI051	48	0.99414	0.99414	0.45295
ASMI053	14	–	–	0.50025
ASMI054	27.75	0.76953	0.72656	0.47976
ASMI056	141.5	0.99023	0.98047	0.5195
ASMI057	57.25	0.97266	0.67969	0.46245
ASMI061	12	–	–	0.44602
ASMI088	9.75	–	–	0.43187
ASMI091	24	0.67969	0.47266	0.48862
ASMI096	8	–	–	0.47949
ASMI097	10	–	–	0.50159
AMBUGA	8	–	–	0.5181
AMMEGA	10	–	–	0.49401

Column heading "Copies" is the average number of samples included per locus. Results indicate no recent genetic bottleneck events, but evidence of a severe genetic bottleneck in the more distant past

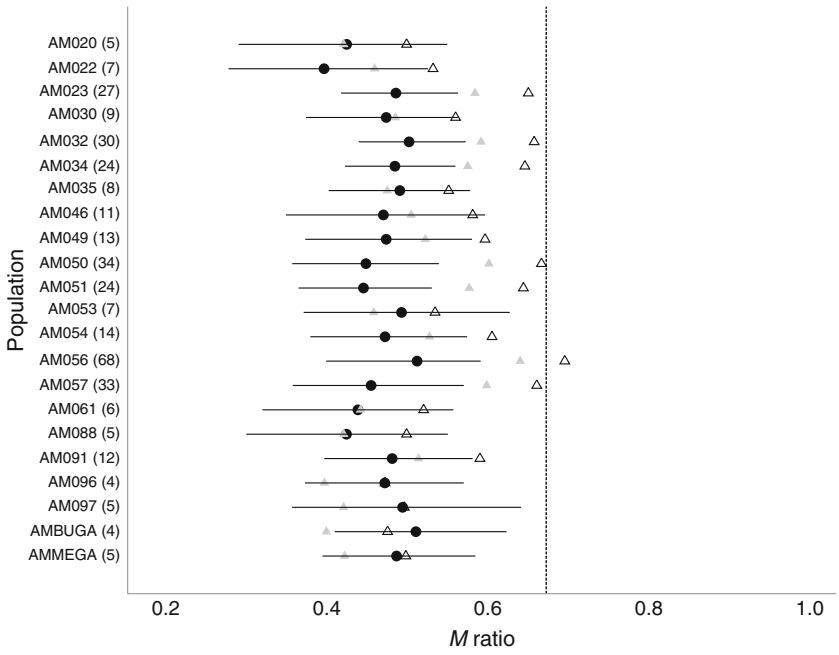

Fig. 4 *M* ratio values (*black circles*) estimated for 22 *Astragalus michauxii* populations across North Carolina and Georgia (USA). *Horizontal lines* represent 95 % CIs, the *vertical line* is the threshold indicative of a past genetic bottleneck, the *open triangles* are the critical M_c (90 % single step mutation model, SSM), and the *gray triangles* are the critical M_c (80 % SMM) (Garza and Williamson 2001). Population sizes are in *parentheses*

Discussion

Habitat fragmentation and degradation has reduced the size and genetic diversity within many plant populations. However, we found that within *A. michauxii* populations, genetic diversity is not lower than other perennial plant species, despite the fact that the longleaf pine ecosystem to which it is endemic has experienced widespread reduction and degradation over the past few centuries. While there are difficulties inherent in comparing microsatellite

diversity across species, mean estimates of expected (H_e) and observed (H_o) heterozygosity for *A. michauxii* populations (0.68 and 0.57 respectively) were comparable to the average H_e (0.61 ± 0.21) and H_o (0.58 ± 0.22) found in a review of plant microsatellite data sets (Nybom 2004).

Populations of *A. michauxii* also exhibited relatively low genetic differentiation, counter to what we expected for a species with varying degrees of population spatial isolation. Within-population genetic variation accounted for 91 % of the total genetic variation; this is not unexpected as other studies have shown that outcrossing, perennial plant species maintain the majority of their genetic diversity within populations (Hamrick and Godt 1996; Nybom 2004). These results are consistent with the genetic structure of three other putative relictual species endemic to the Gulf and Atlantic Coastal Plain Fall-line Sandhills that maintain a large portion of the overall genetic diversity within populations. *Amorpha georgiana* Wilbur (Fabaceae) is a shrub that occurs along river terraces of blackwater rivers in the Fall-line Sandhills and was estimated to contain 89 % of its genetic variation within populations (Straub and Doyle 2009). *Pyxidanthera brevifolia* Wells (Diapensiaceae) occupies xeric upland habitats similar to those occupied by *A. michauxii*, with genetic evidence suggesting within population genetic variation accounted for 90.5 % of the total genetic variation of the species (Wall et al. 2010). Finally, *Lilium pyrophilum* M.W. Skinner and Sorrie (Liliaceae), which occupies relatively more mesic habitats in the Fall-line Sandhills (Douglas et al. 2011) also demonstrates low population differentiation (N. A. Douglas, unpublished data).

The relatively modest genetic differentiation and moderate genetic diversity within *A. michauxii* suggest that population fragmentation occurred following the extensive fragmentation, reduction, and degradation of the longleaf pine ecosystem. Either the populations have not been isolated long enough for genetic drift and mutation to have impacted population differentiation, or gene flow persists despite a fragmented distribution. While the two *A. michauxii* populations in Georgia appear to be distinct from the North Carolina populations based on the NMDS ordination, the genetic differences between the two regions may best be explained by an isolation by distance model (Fig. 3 in supplemental material). If populations had been isolated since the end of Pleistocene, we would expect greater genetic differentiation between populations and/or lower genetic diversity, as has been found in other species (Reisch et al. 2003).

However, these scenarios are not mutually exclusive. If *A. michauxii* is a long-term inhabitant of the Fall-line Sandhills (which has been demonstrated for another endemic taxon, Wall et al. 2010), then it is possible that climatic changes since the Pleistocene, as well as European settlement, have led to the current habitat fragmentation of the Georgia and North Carolina populations. The formation of aeolian river dunes and braided river channels in the Atlantic Coastal Plain during the Late Pleistocene (Ivester et al. 2001; Leigh 2008) suggest an environment with exposed soil and dry, windy climatic conditions (Leigh 2008), not unlike the Great Basin and other ecosystems with a large number of *Astragalus* species (Barneby 1964). The colder, drier conditions of the Pleistocene most likely reduced plant productivity and biomass accumulation, and the region would have been characterized as an open savanna with scattered *Picea* and *Pinus* species and an herbaceous understory (Watts 1980). As climatic conditions became progressively warmer and wetter during the Holocene, it is likely that the biomass and canopy cover increased in many areas. Evidence from extant and historical local populations suggests that *A. michauxii* is sensitive to woody encroachment (North Carolina Natural Heritage Program data). Thus *A. michauxii* may have become restricted to habitats with xeric soil conditions where competition was reduced through increasing fire frequency (upland sites in the Fall-line Sandhills).

These reductions in population size could have led to a genetic bottleneck. Although within population genetic diversity of *A. michauxii* may be comparable to that of other perennial plant species, there is evidence of a past genetic bottleneck based on the ratio of the number of alleles to the allelic range size (*M*-ratio) within the 24 sampled populations. Although the expected mutation-drift heterozygosity (H_{eq}) was not significantly less than H_e in the BOTTLENECK analysis, indicating no evidence of a recent genetic bottleneck, this method may not be as sensitive as evaluating *M* ratio values (Girod et al. 2011). Furthermore, *M* ratio values may be reduced for 100 or more generations, longer than the statistics evaluated in BOTTLENECK. Since *A. michauxii* has an estimated generation time of 9.6 years (Wall, unpublished data), it is not possible to identify whether the bottleneck occurred before or after the anthropogenic fragmentation of the longleaf pine ecosystem using the *M* ratio values alone, as they could be detecting bottlenecks up to 9.6 ka.

For the North Carolina populations, there is evidence of continued interpopulation gene flow (Table 5). Indeed, recent gene flow appears to have occurred broadly across the North Carolina populations, with no evidence of isolation by distance in these populations. However, we must distinguish between statistical migrants, which GeneClass2 identifies based on population gene frequencies, and actual gene flow via pollen or seeds. The results suggest gene flow across an average of 15 km. This is an extreme distance (Greenleaf et al. 2007) for an entomophilous species (Karron 1987; Geer et al. 1995; Crone and Lesica 2004; Becker et al. 2011) with no obvious long distance dispersal

Table 5 *Astragalus michauxii* individuals identified by GeneClass 2 as being the result of possible interpopulation gene flow ($p \geq 0.01$)

Resident population	Assigned population	Individual	Probability	Distance (km)
ASMI022	ASMI032	Ind-36	0.005	5.098
ASMI023	ASMI054	Ind-45	0.006	40.415
ASMI035	ASMI054	Ind-124	0.001	39.628
ASMI049	ASMI034	Ind-137	0.009	6.986
ASMI051	ASMI034	Ind-323	0.003	10.304
ASMI051	ASMI030	Ind-327	0.005	5.379
ASMI053	ASMI050	Ind-1	0.005	26.853
ASMI054	ASMI032	Ind-9	0.01	32.460
ASMI056	ASMI034	Ind-198	0.003	1.956
ASMI056	ASMI046	Ind-188	0.007	13.891
ASMI057	ASMI096	Ind-219	0.001	8.783
ASMI057	ASMI097	Ind-223	0.006	17.144
ASMI057	ASMI035	Ind-238	0.01	2.184
ASMI096	ASMI049	Ind-263	0.01	9159
ASMI097	ASMI057	Ind-342	0	17,144

Resident population refers to the population in which the individual was actually found and assigned population indicates the most likely source of the inferred gene flow (source population). Fourteen individuals across eleven populations were identified

adaptation. Thus, the *A. michauxii* individuals identified as migrants should not necessarily be viewed as actual migrants, but rather as indicative of contemporary gene flow.

Both genetic and demographic factors can affect the long-term viability of plant populations. Our results suggest limited genetic effects of habitat fragmentation and population isolation within *A. michauxii.* However, maintaining future connectivity between populations will be necessary to reduce the negative impacts of future genetic drift and inbreeding. While *A. michauxii* populations may not currently be affected by deleterious genetic processes, demographic analyses suggest that *A. michauxii* may not be maintaining stable populations (Wall et al. 2012), mainly due to limited recruitment of seedlings. This suggests maintaining or increasing population sizes through active management of habitat or more intensive measures, such as augmentation or reintroduction, will likely be critical for the persistence of the species. If active management options are warranted, we recommend that seed source be restricted to the three identified genetic clusters, especially for the Georgia populations. This will maintain the relative genetic distinctiveness of the Georgia populations and the variability identified in the North Carolina populations.

Acknowledgments We thank Fort Bragg Military Reservation for access and support and the Construction Engineering Research Laboratory (US Army Corps of Engineers Cooperative Agreement #W9132T-07-2-0019 to W. Hoffmann) for funding. We also thank Bert Pittman, Sherry Emerine, Jacob Hilton, Kristen Kostelnik, Linda Lee, Xiang Liu, the North Carolina Natural Heritage Program, Tom Patrick, Andrew Walker, the Xiang Lab, and three anonymous reviewers. Assistance with microsatellite development was generously provided by Pacific Northwest Research Station (USDA Forest Service).

References

Balmford A, Bond W (2005) Trends in the state of nature and their implications for human well-being. Ecol Lett 8:1218–1234

Barneby RC (1964) Atlas of North American Astragalus, vol 13. Memoirs of the New York Botanical Garden, NewYork, pp 911–958

Becker T, Voss N, Durka W (2011) Pollen limitation and inbreeding depression in an "old rare" bumblebee-pollinated grassland herb. Plant Biol 13:857–864

Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 163: 367–374

Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics 20:2363–2369

Corander J, Sirén J, Arjas E (2008) Bayesian spatial modeling of genetic population structure. Comput Stat 23:111–129

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Crone EE, Lesica P (2004) Causes of synchronous flowering in *Astragalus scaphoides*, an iteroparous perennial plant. Ecology 85:1944–1954

Cronn R, Liston A, Parks M et al (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res 36:e122

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc: Ser B (Methodol) 39:1–38

Dobson AP, Rodriguez JP, Roberts WM, Wilcove DS (1997) Geographic distribution of endangered species in the United States. Science 275:550–553

Douglas NA, Wall WA, Xiang QY et al (2011) Recent vicariance and the origin of the rare, edaphically specialized Sandhills lily, *Lilium pyrophilum* (Liliaceae): evidence from phylogenetic and coalescent analyses. Mol Ecol 20:2901–2915

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Excoffier L, Lischer HL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Excoffier L, Smouse PE, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Frankham R (2005) Genetics and extinction. Biol Conserv 126: 131–140

Frankham R, Ballou JD, Briscoe DA (2010) Introduction to conservation genetics, 2nd edn. Cambridge University Press, Cambridge

Franklin IR (1980) Evolutionary change in small populations. In: Soule ME, Wilcox BA (eds) Conservation biology: an evolutionary-ecological perspective. Sinauer Associates, Sunderland, pp 135–149

Frost CC (1993) Four centuries of changing landscape patterns in the longleaf pine ecosystem. In: Hermann SM (ed) Proceedings of the tall timbers fire ecology conference. Tall Timbers Research Station, Tallahassee, pp 17–37

Frost CC (2006) History and future of the longleaf pine ecosystem. In: Jose S, Jokela EJ, Miller DL (eds) The longleaf pine ecosystem: ecology, silviculture, and restoration. Springer Science, New York, pp 9–42

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. Mol Ecol 10:305–318

Geer SM, Tepedino VJ, Griswold TL, Bowlin WR (1995) Pollinator sharing by three sympatric milkvetches, including the endangered species *Astragalus montii*. Great Basin Nat 55:19–28

Girod C, Vitalis R, Leblois R, Freville H (2011) Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the msvar method. Genetics 188:165–179

Greenleaf SS, Williams NM, Winfree R, Kremen C (2007) Bee foraging ranges and their relationship to body size. Oecologia 153:589–596

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48:361–372

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer Associates, Sunderland, pp 43–63

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Philos Trans R Soc B: Biol Sci 351: 1291–1298

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Ivester AH, Leigh DS, Godfrey-Smith DI (2001) Chronology of inland eolian dunes on the Coastal Plain of Georgia, USA. Quat Res 55:293–302

James MMR (2000) Legumes in loamy soil communities of the Carolina sandhills: their natural distributions and performance of seeds and seedlings along complex ecological gradients. Master of Science thesis, University of North Carolina

Jennings TN, Knaus BJ, Mullins TD et al (2011) Multiplexed microsatellite recovery using massively parallel sequencing. Mol Ecol Resour 11:1060–1067

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv Genet 5:539–543

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes 5:187–189

Karron JD (1987) The pollination ecology of co-occuring geographically restricted and widespread species of *Astragalus* (Fabaceae). Biol Conserv 39:179–193

Leigh DS (2008) Late quaternary climates and river channels of the Atlantic Coastal Plain, Southeastern USA. Geomorphology 101: 90–108

Leimu R, Mutikainen PIA, Koricheva J, Fischer M (2006) How general are positive relationships between plant population size, fitness and genetic variation? Ecology 94:942–952

Matschiner M, Salzburger W (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. Bioinformatics 25:1982–1983

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. Oikos 100:403–405

NatureServe (2012) NatureServe Explorer: An online encyclopedia of life (web application), Version 7.1. http://www.natureserve.org/explorer. Accessed 28 May 2012

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Noss RF (2013) Forgotten grasslands of the South: natural history and conservation. Island Press, Washington

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13:1143–1155

Oksanen J, Kindt R, Legendre P, et al. (2009) Vegan: community ecology package

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol 13:55–65

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Peet RK (2006) Ecological classification of longleaf pine woodlands. In: Jose S, Jokela EJ, Miller DL (eds) The longleaf pine ecosystem: ecology, silviculture, and restoration. Springer, New York, pp 51–93

Piry S, Luikart G, Cornuet J-M (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503

Piry S, Alapetite A, Cornuet JM et al (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. J Hered 95:536–539

R Development Core Team (2012) R: a Language and environment for statistical computing. Austria, Vienna

Radford AE, Ahles HE, Bell CR (1968) Manual of the vascular flora of the Carolinas. University of North Carolina Press, Chapel Hill

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci 94:9197–9201

Reisch C, Poschlod P, Wingender R (2003) Genetic variation of Saxifraga paniculata Mill. (Saxifragaceae): molecular evidence for glacial relict endemism in central Europe. Biol J Linn Soc 80:11–21

Roberts DW (2010) labdsv: ordination and multivariate analysis for ecology

Schafale MP, Weakley AS (1990) Classification of the natural communities of North Carolina: third approximation, p 321

Scherson RA, Vidal R, Sanderson MJ (2008) Phylogeny, biogeography, and rates of diversification of New World *Astragalus* (Leguminosae) with an emphasis on South American radiations. Am J Bot 95:1030–1039

Slatkin M, Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the expectation–maximization algorithm. Heredity 76:377–383

Sorrie BA, Weakley AS (2001) Coastal plain vascular plant endemics: phytogeographic patterns. Castanea 66:50–82

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci USA 101:15261–15264

Straub SCK, Doyle JJ (2009) Conservation genetics of *Amorpha georgiana* (Fabaceae), an endangered legume of the Southeastern United States. Mol Ecol 18:4349–4365

Swatdipong A, Primmer CR, Vasemägi A (2009) Historical and recent genetic bottlenecks in European grayling, *Thymallus thymallus*. Conserv Genet 11:279–292

Van Lear DH, Carroll W, Kapeluck P, Johnson R (2005) History and restoration of the longleaf pine-grassland ecosystem: implications for species at risk. For Ecol Manag 211:150–165

Wall WA, Douglas NA, Xiang QY et al (2010) Evidence for range stasis during the latter Pleistocene for the Atlantic Coastal Plain endemic genus, *Pyxidanthera Michaux*. Mol Ecol 19:4302–4314

Wall WA, Hoffmann WA, Wentworth TR et al (2012) Demographic effects of fire on two endemic plant species in the longleaf pine-wiregrass ecosystem. Plant Ecol 213:1093–1104

Watts WA (1980) Late-quaternary vegetation history at white pond on the inner Coastal Plain of South Carolina. Quat Res 13:187–199

Webb T III (1987) The appearance and disappearance of major vegetational assemblages: long-term vegetational dynamics in eastern North America. Theory and models in vegetation science. Springer, Berlin, pp 177–187

Weeks S (2005) Factors limiting growth in *Astragalus michauxii* (sandhills milk-vetch). Doctoral dissertation, North Carolina State University

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Williams JW, Shuman BN, Webb T III et al (2004) Late-quaternary vegetation dynamics in North America: scaling from taxa to biomes. Ecol Monogr 74:309–334

You F, Huo N, Gu Y et al (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol 11:413–418

MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 2901–2915 doi: 10.1111/j.1365-294X.2011.05151.x

Recent vicariance and the origin of the rare, edaphically specialized Sandhills lily, *Lilium pyrophilum* (Liliaceae): evidence from phylogenetic and coalescent analyses

NORMAN A. DOUGLAS,* WADE A. WALL,* QIU-YUN (JENNY) XIANG,* WILLIAM A. HOFFMANN,* THOMAS R. WENTWORTH,* JANET B. GRAY† and MATTHEW G. HOHMANN‡

**Department of Plant Biology, PO Box 7612, North Carolina State University, Raleigh, NC 27695, USA, †Directorate of Public Works, Endangered Species Branch, United States Army, Fort Bragg, NC 28310, USA, ‡US Army Corps of Engineers, Engineer Research and Development Center, PO Box 9005, Champaign, IL 618262, USA*

Abstract

Establishing the phylogenetic and demographic history of rare plants improves our understanding of mechanisms that have led to their origin and can lead to valuable insights that inform conservation decisions. The Atlantic coastal plain of eastern North America harbours many rare and endemic species, yet their evolution is poorly understood. We investigate the rare Sandhills lily (*Lilium pyrophilum*), which is endemic to seepage slopes in a restricted area of the Atlantic coastal plain of eastern North America. Using phylogenetic evidence from chloroplast, nuclear internal transcribed spacer and two low-copy nuclear genes, we establish a close relationship between *L. pyrophilum* and the widespread Turk's cap lily, *L. superbum*. Isolation-with-migration and coalescent simulation analyses suggest that (i) the divergence between these two species falls in the late Pleistocene or Holocene and almost certainly post-dates the establishment of the edaphic conditions to which *L. pyrophilum* is presently restricted, (ii) vicariance is responsible for the present range disjunction between the two species, and that subsequent gene flow has been asymmetrical and (iii) *L. pyrophilum* harbours substantial genetic diversity in spite of its present rarity. This system provides an example of the role of edaphic specialization and climate change in promoting diversification in the Atlantic coastal plain.

Keywords: coalescence, divergence, edaphic, *Lilium*, Pleistocene, rarity

Received 18 November 2010; revision received 28 April 2011; accepted 5 May 2011

Introduction

Molecular studies of rare plant taxa usually aim to quantify the level and patterns of genetic diversity in a particular species (Karron 1987; Hamrick & Godt 1990; Ellstrand & Elam 1993; Gitzendanner & Soltis 2000). Phylogeographic studies, on the other hand, often focus on widespread species and try to discern continental-scale patterns (Taberlet *et al.* 1998; Brunsfeld *et al.* 2001; Soltis *et al.* 2006). However, the tools of phylogeography, particularly coalescent-based analyses that provide information about the age and historical demography of species (Knowles 2009), have only rarely been applied to investigate the history of rare species (Raduski *et al.* 2010; Whittall *et al.* 2010).

Of the 'seven forms of rarity' (Rabinowitz 1981), the most extreme describes taxa that have a narrow geographic range, require specific habitats and maintain only small local populations. Many edaphic endemics (plants restricted to soils with unusual physical or chemical properties) belong to this category. While the textbook examples of edaphic endemic plants are restricted to serpentine, various substrates support edaphic endemics, including guano, alkali, salt, and gypsum deposits, limestone, chalk, and granite outcrops, oligotrophic bogs and deep porous sands (Ornduff 1965; Axelrod 1972; Parsons 1976; Kruckeberg &

Correspondence: Norman A. Douglas, Fax: (919) 515 3436; E-mail: norman_douglas@ncsu.edu

Rabinowitz 1985; Kruckeberg 1986; Williamson & Bazeer 1997). Many aspects of the origin of edaphic endemic species are poorly understood (Rajakaruna 2004). For instance, such species often occur in close geographic proximity to their progenitor lineages (e.g. Baldwin 2005), yet it is not usually known whether or how strongly gene flow is interrupted. While taxa displaying edaphic endemic syndromes often show reduced genetic diversity compared with their close relatives (Godt & Hamrick 1993; Baskauf *et al.* 1994; Ayres & Ryan 1999; but see Raduski *et al.* 2010), this may reflect genetic drift due to lower population sizes or the effects of selection. Strong selection imposed by edaphically challenging soils could be sufficient to foster population divergence (Nosil *et al.* 2009; Freeland *et al.* 2010). Some edaphic endemics may represent vicariant populations isolated in narrow parts of formerly wider ranges and niches of their progenitors (e.g. Crawford *et al.* 1985), which may themselves be able to grow on the unusual substrate without being restricted to it.

Edaphic specialists (especially in bog and sand habitats, Sorrie & Weakley 2001) are an important component of the endemic-rich flora of the coastal plain of eastern North America. Few coastal plain endemics have been the subject of molecular analyses. Sand dune habitats in Florida apparently served as Pleistocene refugia for the genera *Dicerandra* and *Conradina* (Edwards *et al.* 2006; Oliveira *et al.* 2007), and in general, Florida has been proposed as a major Pleistocene refugium for many taxa in eastern North America (Soltis *et al.* 2006). Yet, recent phylogeographic work indicates that some coastal plain endemic species likely persisted in northerly latitudes throughout the Pleistocene. For instance, the Atlantic coastal plain endemic Pyxie Moss, *Pyxidanthera* (Diapensiaceae), shows long-term range stasis (Wall *et al.* 2010).

The Fall-Line Sandhills of North and South Carolina (which occur at the western boundary of the coastal plain) provide one of the clearest examples of the edaphic contribution to the botanical diversity of the Atlantic coastal plain. This region is comprised of rolling hills of open, fire-maintained longleaf pine (*Pinus palustris*) savanna dissected by numerous blackwater streams and wetlands, providing a diverse matrix of habitats that support at least eight endemic plants (and numerous near-endemics, Sorrie & Weakley 2001). In the core of the Sandhills region in southern North Carolina, the uppermost deposit is the Pinehurst formation, which is characterized by loose coarse-grained sands found along ridgetops. This formation was deposited in a tidal environment (J. Nickerson, North Carolina Geological Survey, personal communication) and may date to the Eocene (Cabe *et al.* 1992). Below the Pinehurst formation (and exposed along drainages and slopes throughout the region) lies the Cretaceous Middendorf formation, which is of deltaic origin and thus has more abundant clays (Sohl & Owens 1991). At the interface between these (and similar formations in the Carolinas and southeastern Virginia) occur Sandhills seep and streamhead pocosin ecotone communities. When kept open by frequent fires encroaching from the surrounding xeric pine savannas, these wetlands can support extremely high local species richness, among the highest values ever recorded in North America (>102 species per 1/100 ha, Schafale & Weakley 1990). The age of the formations implies that endemic species have potentially had a very long time to adapt to the unusual edaphic conditions.

In this study, we consider the Sandhills lily, *Lilium pyrophilum* (Liliaceae), a striking endemic of the Sandhills in the Carolinas and southeastern Virginia. Formally described only recently (Skinner & Sorrie 2002), specimens of this species were previously identified in herbaria as any of three similar species in the region (*L. superbum*, *L. michauxii* or *L. iridollae*) that share the distinctive 'Turk's cap' morphology, in which flowers are pendent with the tepals reflexed upward. Skinner & Sorrie (2002) identified three specific plant communities (Schafale & Weakley 1990; Sorrie *et al.* 2006) that support *L. pyrophilum*, including Sandhills seep and streamhead pocosin ecotones. The third, small stream swamps are affected by frequent flooding events in addition to seepage and rarely support *L. pyrophilum* (Sorrie *et al.* 2006).

Lilium pyrophilum is a very rare species. There are fewer than 75 historical and extant locations in North and South Carolina, and Virginia (North Carolina Natural Heritage Program 2007), and between 2007 and 2009, a survey of all known populations located <500 stems across 35 populations (W. Wall, unpublished data). Approximately half of the extant populations and a quarter of the individuals occur on Fort Bragg Military Reservation in North Carolina, where prescribed and ordnance-ignited fires maintain appropriate habitat.

In describing *L. pyrophilum* (Skinner & Sorrie 2002), the authors outlined three phylogenetic hypotheses concerning the origin of the species. First, they speculated that *L. pyrophilum* may represent a peripheral isolate of the Turk's cap lily, *L. superbum*, which it most resembles morphologically (albeit with significant differences, Skinner & Sorrie 2002). *Lilium superbum* is distributed throughout much of eastern North America (Fig. 1), and in contrast to the edaphically specialized *L. pyrophilum*, it is a generalist, occurring in rich woods and oligotrophic wetlands from high elevation to sea level. Especially in northern parts of its range (e.g. the Pine Barrens of New Jersey), it can be found in saturated

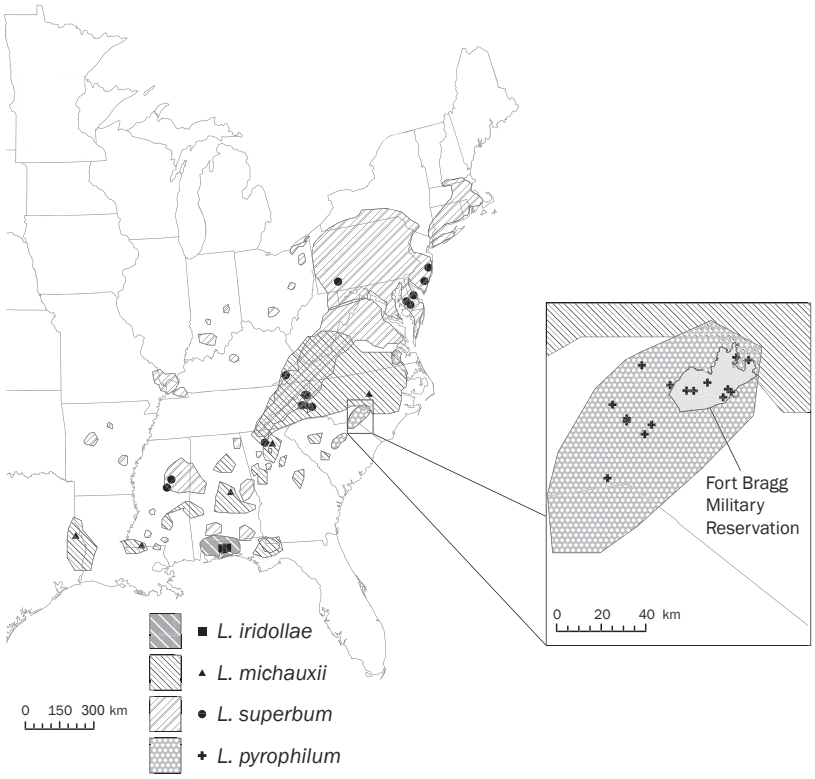

Fig. 1 Distribution of populations included in this study and geographic ranges of the four focal species.

sandy habitats not unlike those preferred by *L. pyrophilum*, but it is not restricted to them. However, it is essentially absent from the Piedmont and Atlantic coastal plain from the Carolinas southward. Thus, it is disjunct from *L. pyrophilum* by at least 150 km everywhere except in southeastern Virginia (Fig. 1) where the coastal plain narrows.

Second, they speculated that *L. pyrophilum* may represent a hybrid species, with the widespread Carolina lily (*L. michauxii*) and *L. superbum* as progenitors. Homoploid hybrid speciation has been implicated in the origin of other edaphic specialists, e.g. *Helianthus paradoxus* (Rieseberg *et al.* 1990) and Hawaiian *Scaevola* (Howarth & Baum 2005). Of the three potentially related species, *L. pyrophilum* resembles *L. michauxii* least, differing in leaf shape and producing fragrant flowers (Skinner 2002). While the range of *L. michauxii* does overlap the range of *L. pyrophilum* (Fig. 1), they occur in contrasting habitats, with *L. michauxii* favouring much drier sites. Notably, *L. michauxii* and *L. superbum* co-occur throughout much of their ranges (Fig. 1), yet natural hybrids are apparently rare (Skinner 2002).

Finally, Skinner and Sorrie suggested the possibility that *L. pyrophilum* may represent a disjunct population of the Pot-o'-gold or Panhandle lily (*L. iridollae*), a narrow endemic of wet pine savannas in northwestern Florida (where it is listed as endangered) and adjacent Alabama. This hypothesis emphasizes similar habitat requirements of the two species, but downplays consistent morphological differences (e.g. details of rhizome structure, Skinner 2002; Skinner & Sorrie 2002) and a range separation of over 700 km (Fig. 1).

In this study, we report the results of a molecular study focused on *L. pyrophilum* and its close relatives. First, we investigated the phylogeny of the eastern pendent species of *Lilium* to address whether *L. pyrophilum* represents a peripheral isolate of *L. superbum*, a hybrid between *L. superbum* and *L. michauxii*, or a disjunct population of *L. iridollae*. Second, we analysed the distribution of genetic variation within and among the taxa thought to be closely related to *L. pyrophilum* and used coalescent-based methods to explicitly evaluate the possible timing of the divergence of *L. pyrophilum*. Our results are interpreted in the context of the evolution of rare, edaphically specialized lineages in the Atlantic coastal plain.

Materials and methods

Sampling and molecular data

Samples were obtained from 50 populations spanning the geographic range of each of the four focal species (Fig. 1). We also sampled two populations of *Lilium*

canadense, another pendant species that lacks the Turk's cap morphology. Sampling information is provided in Table S1 (Supporting Information). Populations were located in the field based on documented occurrences from herbarium specimens, element occurrence records from state Natural Heritage Programs and communication with local botanists. We endeavoured to sample a similar number of populations of *L. superbum* and *L. michauxii* spanning the geographic range of each species. Our sampling of the rare *L. iridollae* was limited to two populations. In general, one individual was taken to represent each population. Genomic DNA was isolated from fresh or frozen leaves, using the CTAB method (Doyle & Doyle 1987). Nuclear ribosomal internal transcribed spacer ('ITS') sequences were obtained with primers ITS4 and ITS5a (White *et al.* 1990; Stanford *et al.* 2000). This locus was sequenced to facilitate comparison with abundant existing data available in GenBank to determine whether the species in this study form a monophyletic group. We screened eight chloroplast markers from Shaw *et al.* (2007); of these, three (the *atpI-atpH*, *psbD-trnT* and *rpl32-trnL* intergenic spacers) consistently amplified and contained variable sites. As the chloroplast behaves as a single nonrecombining locus, sequences of these three regions were concatenated, and this marker is hereafter referred to as 'CP'.

We developed single-copy nuclear markers for *Lilium*. In general, we screened EST or complete CDS sequences from *Lilium* against the *Oryza sativa* genomic sequence at GenBank using SPIDEY (Wheelan *et al.* 2001) with the 'divergent sequences' and 'use large intron sizes' options. Candidate sequences were downloaded and manually aligned in Se-Al (Rambaut 1996) using amino acid translations. Homologous sequences from GenBank were incorporated into the alignments. When we were confident of the positions of the introns in the rice genome, we then designed primers using Primer3 (Rozen & Skaletsky 2000), which were screened against DNA extracted from *L. longiflorum* and an Asiatic hybrid cultivar (which served as positive controls because nearly all of our candidate regions were based on sequences from these cultivated lilies) and the four taxa in our study. We were able to obtain single amplicons for relatively few of these regions even after extensive PCR optimization; it was often the case that primers would amplify nontarget regions or that introns would be small, invariant or missing entirely. The closely related *L. canadense* has a phenomenally large genome (1C = 47.90 pg, 46.9 Gbp; Zonneveld *et al.* 2005; Peruzzi *et al.* 2009), which may have contributed to the difficulty we encountered in obtaining single-copy nuclear sequences. However, we were able to design primers that amplified two novel regions. The first includes two introns between exons 8 and 10 of the *L. longiflorum* alkaline phytase gene, *LlAlp* ('AP', primers: AP8f, 5′-TCTCCTTGGGCTCTTTCTTG and AP10r, 5′-GAAAACCTCAAATGGGCAGAG), which is involved in phytic acid metabolism (Mehta *et al.* 2006). While GenBank contains sequences for two isoforms of this gene, our PCR experiments are consistent with these representing splice variants of a single locus. The second region corresponds to a region between exons 5 and 10 of the *AKT1*-like potassium channel *LilKT1* ('AKT', primers: AKT5f, 5′-AGAGACTCTTGATGCACT TCCTAAA and AKT10r, 5′-AAGAGAACAACA-CAACTTTCATTCC). This locus was more difficult to amplify, and we were unable to generate sequences for *L. iridollae*. Primers and PCR conditions for ITS and the chloroplast loci followed White *et al.* (1990) and Shaw *et al.* (2007). For AP and AKT, PCR contained 2.5 μL 10× PCR buffer, 1% BSA, 200 μM dNTPs, 2.5 mM $MgCl_2$, 4 μM of each primer and 0.5 U Taq DNA polymerase. Cycling conditions were 95 °C for 4 min, followed by 35 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 2.5 min, and a final extension step of 72 °C for 4 min. Amplicons were cleaned with Antarctic Phosphatase and Exonuclease I (New England Biolabs, Ipswich, MA, USA). Sequencing was performed on an Applied Biosystems 3730 capillary sequencer (Foster City, CA, USA) using Big Dye chemistry. Chromatograms were edited in Sequencher 4.1.2 (Gene Codes Corporation, Ann Arbor, MI, USA). Heterozygous bases were easily identified in the chromatograms for the three nuclear regions and coded with standard IUPAC notation. Because of the low levels of divergence among our sequences, alignment was trivial and performed manually in Se-Al. The most likely haplotypic phases of AP and AKT genotype sequences were ascertained with a combination of cloning and the program PHASE 2.1 (Stephens *et al.* 2001; Stephens & Donnelly 2003) called by the 'Open/Unphase genotype' option in DnaSP v. 5 (Librado & Rozas 2009); the inferred alleles form the basis for all further analyses involving these loci. The preferred model of sequence evolution for each locus (ITS: TIM3ef + I + G; CP: K81uf + I; AP: TVM + I; AKT: TVM + I + G) was determined according to Akaike Information Criterion (AIC) in jModelTest (Posada 2008). Sampling details, genotype information and GenBank accession numbers are provided in Tables S1 and S2 (Supporting Information).

Phylogenetic analyses and descriptive population genetics

For the ITS analysis, 44 new sequences were aligned with 49 from GenBank to create a matrix of 93 sequences. Included were the four species in this study, plus 37 other taxa including the pendent eastern North

American species, *L. michiganense*, *L. canadense* and *L. grayi*, and eight others from *Lilium* section *Pseudolirium*, the monophyletic group of North American species (Nishikawa *et al.* 1999) to which all taxa in this study belong. Unweighted parsimony analysis for the ITS locus was accomplished using PAUP* 4.0b10 (Swofford 2002) using 100 random-addition sequence replicates with TBR branch swapping; owing to overall low sequence divergence, parsimony bootstrapping was conducted with 10^6 'fast' stepwise addition sequences (Soltis & Soltis 2003). Maximum-likelihood (ML) analysis for this locus was conducted in GARLI v. 1.0 (Zwickl 2006). Likelihood bootstrap values were obtained with 1000 replicate searches. The statistical parsimony haplotype network was computed for complete sequences of the three chloroplast regions, *atpI-atpH*, *psbD-trnT* and *rpl32-trnL* (38 sequences), using TCS (Clement *et al.* 2000). The nuclear loci (AP: 82 haplotypes; AKT: 62 haplotypes) have a more complicated evolutionary history than chloroplast sequences; thus, network analyses for the two were conducted using the geodesically pruned quasi-median network algorithm (Ayling & Brown 2008) as implemented in SplitsTree4 (Huson & Bryant 2006), which produces pruned networks that connect all sequences (including multistate characters) by at least one shortest path. ML trees (not shown) were inferred for these sequences as well; they were poorly resolved and showed few supported nodes. However, neither nuclear locus showed phylogenetic evidence of paralogy. For *L. michauxii*, *L. superbum* and *L. pyrophilum*, Arlequin v. 3.5 (Excoffier & Lischer 2010) was used to estimate haplotype richness, number of segregating sites, nucleotide diversity π (Nei 1987) and Watterson's (1975) population mutation parameter θ, for the chloroplast and single-copy nuclear loci.

Testing divergence between L. michauxii, L. pyrophilum *and* L. superbum

As our data include a single individual per 'population', we treated species as the main hierarchical level for the purposes of these analyses. Pairwise F_{ST} (Weir & Cockerham 1984) and the exact test of population differentiation (Raymond & Rousset 1995; Goudet *et al.* 1996) between *L. michauxii*, *L. superbum* and *L. pyrophilum* were calculated in Arlequin v. 3.5 (Excoffier & Lischer 2010), with individuals and species used as the hierarchical groupings. Significance was assessed with 10^3 permutations (F_{ST}) or 2×10^6 Markov chain steps (exact test).

The nature of the divergence between *L. superbum* and *L. pyrophilum* was further investigated using the isolation-with-migration model (Nielsen & Wakeley 2001), implemented in IMa2 (Hey & Nielsen 2007). The full model in the two-population case includes six parameters (divergence time, θ for the ancestral and two descendent populations and migration rates between the descendent populations). This model assumes no recombination within loci and free recombination between loci and that markers are selectively neutral. Thus, several recombination detection methods available in the program RDP3 (beta 40; Martin *et al.* 2005) were used to search for recombinant alleles. As selection or demographic changes can cause departures from neutral expectations, DnaSP v. 5 (Librado & Rozas 2009) was used to perform three different tests of neutrality: Tajima's D (Tajima 1989), Fay and Wu's H (Fay & Wu 2000) and R_2 (Ramos-Onsins & Rozas 2002). Critical values for these statistics were obtained using 10^5 coalescent simulations. The chloroplast data set showed no evidence of recombination; the AP and AKT data sets were filtered with IMgc Online (Woerner *et al.* 2007) to create data sets that were free of detectable recombination and infinite sites violations. Maximum priors for the IMa2 analysis were based on recommended starting values given in the program documentation and refined after preliminary exploratory runs. Priors ultimately selected were population mutation rates (for *L. pyrophilum*, *L. superbum* and ancestral population) θ_0, θ_1 and $\theta_2 = 47$, splitting time parameter $t = 3$ and population migration rate m_1 and $m_2 = 10$. Mutation rate priors (CP: 1.5×10^{-9}, AP & AKT: 6.03×10^{-9}) were specified based on values given by Gaut (1998). Seventy geometrically heated chains (using the heating parameters ha = 0.98, hb = 0.50) were run for 750 000 generations beyond a 150 000 generation burn-in and trees were sampled every 75 generations. This process was repeated 10 times using different random number seeds.

Because results from each replicate were similar, 10^5 trees were concatenated into a single run in load-trees mode and the 'test nested models' option was activated. This option evaluates the likelihood of 24 models simpler than the full isolation-with-migration model by constraining parameters (other than divergence time) and rejecting those that are significantly worse than the full model based on a likelihood ratio test. We also compared models using an information-theoretic method (Carstens *et al.* 2009), which allows the relative performance of nested and non-nested models to be compared using AIC. Compared with a hypothesis-testing approach, which simply identifies models that are rejected as significantly worse than the full model, the information-theoretic approach provides model weights that allow the relative performance of each of a given set of models, including the full model, to be compared directly with others given the data (Burnham &

Anderson 2002). We used the full model posterior probability and the 2(log-likelihood ratio) values, which IMa2 estimates for each model under the assumption that the model's posterior probability is proportional to its likelihood, to calculate the AIC for the full model and each nested model. Subsequently, Akaike weights and evidence ratios were calculated (Burnham & Anderson 2002; Carstens *et al.* 2009).

Conversion of the IMa2 parameter estimates from coalescent to demographic units was accomplished assuming a generation time of 20 years. This is arbitrary but conservative, based on what little is known about the natural history of these species. Germination and establishment is slow, taking two seasons, and plants need 7 years to reach flowering size. Year-to-year survivorship is relatively high (>0.95, Wade Wall, unpublished data). Using the equation $T = \alpha + [s/(1 - s)]$, where T = generation time, α = age of first reproduction and s = adult survivorship (Lande *et al.* 2003), we obtain a value of 26 years. Although estimates of survivorship could be too high, the Lande equation does not account for the fact that older plants are typically larger and more fecund than younger ones. In either case, our generation time should be considered a minimum estimate.

Because isolation is implicit in the isolation-with-migration framework, we tested this assumption with a series of coalescent simulations. Briefly, we estimated N_e for each locus using BEAST (Drummond & Rambaut 2007). Because only *L. pyrophilum* and *L. superbum* sequences were included, simpler ML models were utilized (CP: HKY, AP: TnN + I + G, AKT: K81uf + I). We then used Mesquite v. 2.73 (Maddison & Maddison 2010) to simulate 1000 data sets under each of several simple divergence models (using estimated substitution models for each locus). We treated each species as a population such that *L. superbum* had a N_e 3× that of *L. pyrophilum* (the total N_e corresponding to the value from BEAST). The two populations coalesced at times corresponding to 2.58 Ma (earliest Pleistocene), 126 ka (upper Pleistocene) or 18 ka (last glacial maximum). We then conducted parsimony searches using PAUP* 4.10b (Swofford 2002) on each simulated data set saving 1000 consensus trees. Slatkin and Maddison's s (i.e. the number of parsimony steps implied by a given topology treating source population as a character, Slatkin & Maddison 1989) was computed for each tree to create a null distribution for each locus and divergence time. This was compared with the value of s for the empirical data. When minimum empirical values for s were higher than 95% of the simulated values, we rejected the scenario. To evaluate the effect of the level of migration inferred by IMa2, we duplicated these analyses, but allowing migration. Because Mesquite only allows symmetrical migration, we specified a rate of 9.8×10^{-6} migrants per individual per generation, which corresponds to the estimated value of the parameter under the 'equal migration rate' nested model in IMa2. Finally, following Gugger *et al.* (2010), we evaluated the no-divergence scenario by simulating 1000 data sets per locus under a single population scenario. The resulting parsimony consensus trees were contained within the two-population model described previously, and the null distributions of s were calculated. In this case, the scenario was rejected if the maximum empirical values of s were lower than 95% of the simulated values. As coalescent parameter estimates based on single loci are highly sensitive to stochastic error (Edwards & Beerli 2000), these simulations were conducted for both the upper and lower 90% HPD estimates of N_e from BEAST.

Table 1 Genetic diversity and results of neutrality tests

	Lilium michauxii			*Lilium pyrophilum*			*Lilium superbum*		
Species locus	CP	AKT	AP	CP	AKT	AP	CP	AKT	AP
Individuals (haplotypes)	8 (8)	5 (10)	7 (14)	15 (15)	13 (26)	18 (36)	13 (13)	12 (24)	15 (30)
Aligned length (bp)	2361	1428	453	2360	1428	453	2361	1428	453
Segregating sites	7	10	13	7	24	8	9	30	18
Observed haplotypes	5	7	9	4	16	9	7	17	12
Nucleotide diversity π	0.0010	0.0033	0.0098	0.0008	0.0024	0.0016	0.0009	0.0040	0.0053
Watterson's theta θ	0.0011	0.0025	0.0090	0.0009	0.0044	0.0043	0.0012	0.0061	0.0100
Tajima's D	−0.4150	0.0487	0.3349	−0.4468	−1.7637*	−1.8536**	−1.0835	−1.2142	−1.6319*
Fay and Wu's H	1.7857	0.8000	2.2418	1.3429	−8.8862*	−2.8794*	−1.9615	−4.8333	0.6437
R_2	0.1577	0.2091	0.1597	0.1301	0.0625**	0.0495***	0.1105*	0.0828	0.0692*

Sampling represents the number of individuals and the number of haplotypes (for phased nuclear loci). Significance of neutrality tests was assessed with 10^5 coalescent simulations in DnaSP v. 5.1 (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

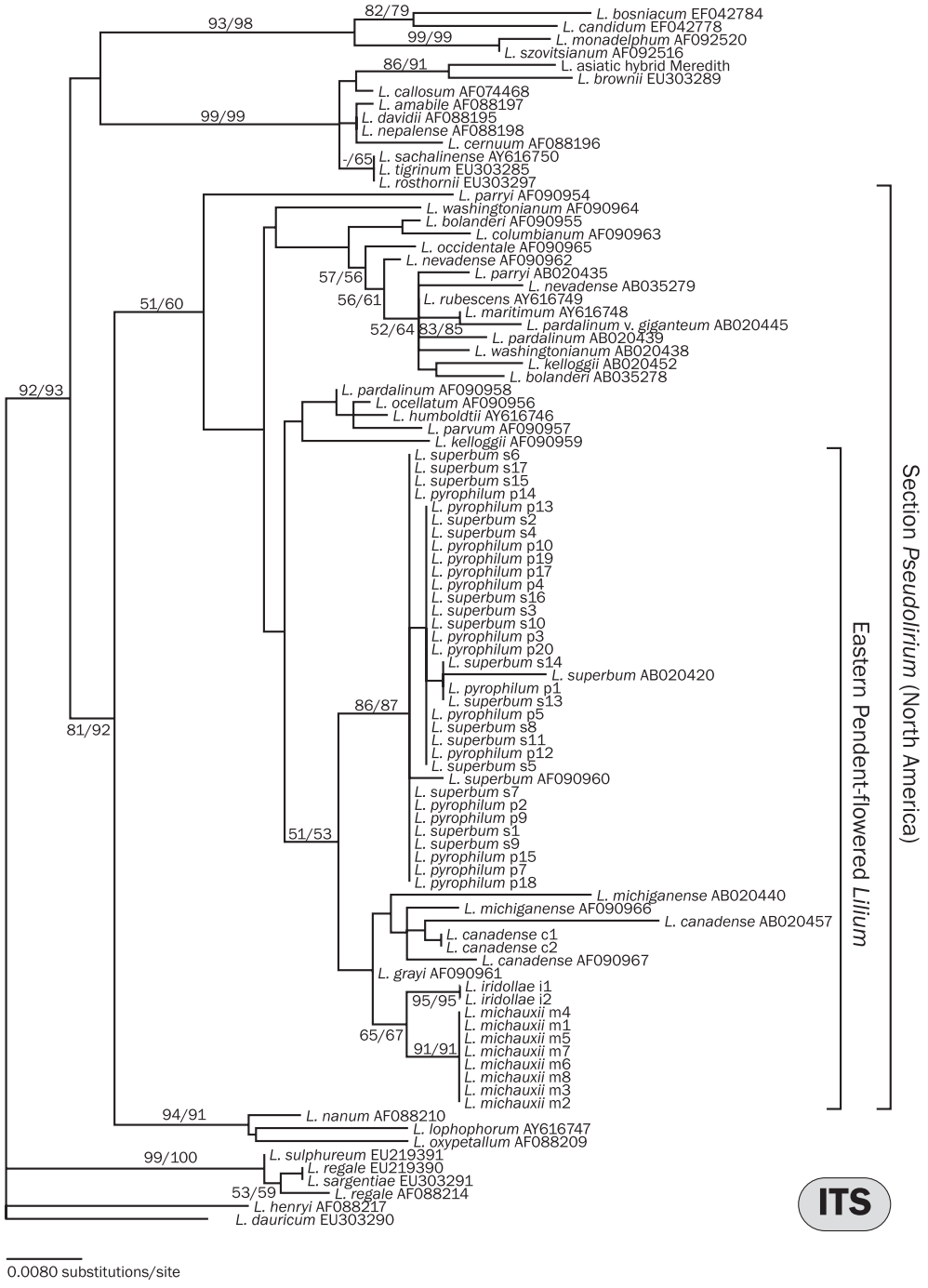

Fig. 2 Maximum-likelihood (ML) Phylogram of internal transcribed spacer sequences. Support values are ML bootstrap/Bayesian posterior probability.

Results

Phylogenetic analyses

In the analysis of ITS data, overall support is quite weak at the level of intra- and interspecific relationships, with no significant (≥70%) bootstrap support for the monophyly of the North American section *Pseudolirium* or the eastern pendent-flowered species (Fig. 2). However, there is a relatively high level of support for the branch uniting two accessions of *Lilium iridollae*, for that uniting the eight samples of *L. michauxii*, and, finally, for the branch leading to the 32 samples of *L. pyrophilum* and *L. superbum*. Little divergence is evident among the

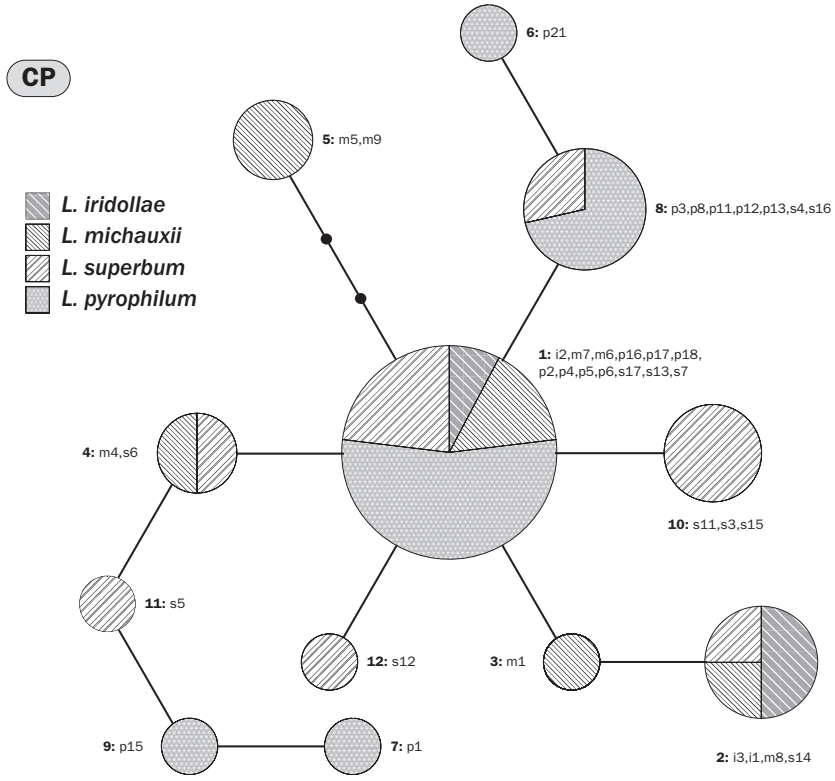

Fig. 3 Chloroplast haplotype network. Statistical parsimony network for CP haplotypes. Chart area reflects the frequency of the haplotype; each slice reflects the frequency at which each haplotype was found in each species. Haplotype numbers (bold) and sample abbreviations correspond to those in Tables S1 and S2 (Supporting Information). Edges represent mutations, black dots unsampled haplotypes.

accessions of each species (with the exception of the GenBank sequences for *L. superbum*, *L. canadense* and *L. michiganense*). The statistical parsimony network (Fig. 3) computed for the chloroplast data revealed a common haplotype (1) that was found in all four species, plus 11 less common types. Overall, four of the six non-singleton haplotypes occur in multiple species. Quasi-median networks produced for the AKT and AP loci (Fig. 4) showed that, while AP haplotype 8 is one mutational step from the nearest *L. michauxii* haplotype (m4a), most *L. michauxii* (and *L. iridollae* in AP) haplotypes are separated from a cloud of *L. pyrophilum* and *L. superbum* haplotypes, which are thoroughly intermixed and frequently shared. No haplotypes were shared between *L. pyrophilum* and *L. michauxii*.

Genetic diversity

Haplotype richness h, segregating sites S, nucleotide diversity π and Watterson's θ are given in Table 1. Nucleotide diversity is relatively low, with values between 0.0008 and 0.00978 substitutions per site, and average values for AP and AKT are nearly five times the value for the chloroplast data set.

Tests of neutrality

Departures from neutrality were detected in the nuclear data sets in *L. pyrophilum* and *L. superbum*, where there were significant negative estimates of Tajima's D and R_2. Fay and Wu's H is significant in *L. pyrophilum* only. Tajima's D is sensitive to both demographic expansion and selection, and R_2 is designed to detect population expansion (Ramos-Onsins & Rozas 2002). While Fay and Wu's H is most sensitive to recent positive selection, it may be sensitive to particular demographic conditions involving structured populations (Fay & Wu 2000). We believe these loci are unlikely to be under positive selection, because there is no obvious reason two loci should deviate from neutrality more strongly in *L. pyrophilum* than in the other two taxa. The chloroplast data also show some demographic expansion in *L. superbum* (weakly significant R_2) without a significantly negative D. Thus, while we cannot eliminate the possibility of some background selection in the nuclear data sets (which does not violate the assumptions of IMa2), it is more likely that demographic factors explain the significant values for these statistics.

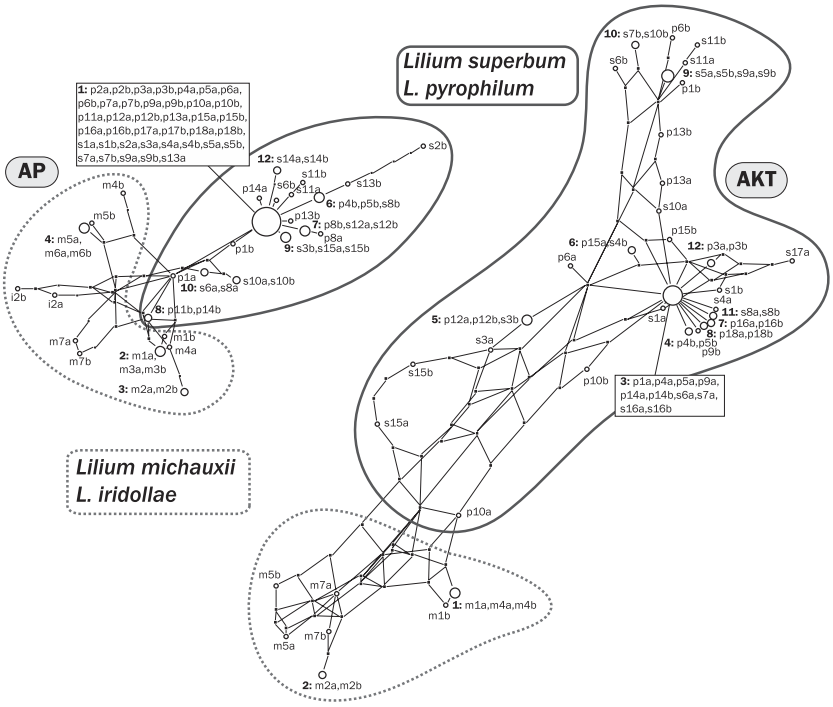

Fig. 4 Quasi-median joining networks for the nuclear loci AP and AKT. Network representations of the relationships between nuclear haplotypes (bold numbers and sample abbreviations correspond to Tables S1 and S2, in Supporting Information). In quasi-median-joining networks, each haplotype is connected to the others by at least one shortest path. Mutational steps are indicated by edges, and black dots represent potential unsampled haplotypes.

Differentiation of L. michauxii

Pairwise F_{ST} values (Table 2) revealed that *L. michauxii* was significantly divergent from *L. pyrophilum* and *L. superbum* for the AKT and AP data sets, whereas differentiation between *L. pyrophilum* and *L. superbum* was minimal and only significant in the AKT data set. No significant differentiation was detected among any of the three species for the CP data set. Conversely, all pairwise exact differentiation tests (Raymond & Rousset 1995) were significant for the two nuclear loci; for the cpDNA, a significant result was only obtained between *L. pyrophilum* and *L. michauxii*.

Divergence between L. pyrophilum *and* L. superbum

Under the isolation-with-migration model, estimates of the mutation parameter theta (θ) were *L. pyrophilum*: 3.736; *L. superbum*: 10.79; and ancestral population: 1.292, corresponding to effective population sizes (95% highest posterior density interval, abbreviated '95% HPD') of 11 400 (2800–29 700), 32 900 (12 800–86 900) and 3900 (0–14 400), respectively (Fig. 5a). The splitting time between *L. pyrophilum* and *L. superbum* was estimated as 0.7725 coalescent units, with the 95% HPD being 0.3435–2.405 (Fig. 5b). This estimate corresponds to a divergence time of 188 ka (95% HPD 84–586 ka) with the assumed mutation rates and generation time. The posterior distribution of splitting time did not reach zero (nor did it for much higher prior values in preliminary runs), so 95% HPD intervals should be interpreted with caution. The coalescent migration rate m from *L. superbum* into *L. pyrophilum* was highest at zero, while the converse was 1.915. Thus, population migration rates (2 NM, Hey & Nielsen 2004) are asymmetrical and quite high from *L. pyrophilum* into *L. superbum* (2 NM = 9.98, Fig. 5c). The model selection procedure (Table 3) preferred a model that holds the two species' population sizes equal and the migration rate from *L. superbum* to *L. pyrophilum* at zero (model weight $w = 0.32$). The next best model ($w = 0.22$) also fixed the *L. superbum* → *L. pyrophilum* migration rate at zero but allowed the population sizes to vary. The full model ($w = 0.19$) had the next highest weight, and the next three models differed in that they fixed the population sizes as above (model 4), held migration rates equal (model 5) and held the *L. pyrophilum* → *L. superbum* migration rate at zero (model 6). The six best models are assigned 95.6% of the total weight. The remaining 19 models had some combination of zero migration, and one or both of the population sizes

Table 2 Pairwise *Fst* and exact test of population differentiation

	Lilium michauxii	*Lilium pyrophilum*	*Lilium superbum*
L. michauxii		0.109/0.393***/0.625***	0.046/0.328***/0.567***
L. pyrophilum	*/***/***		0.007/0.021/0.057*
L. superbum	–/***/**	–/**/*	

Loci: CP/AP/AKT. Above diagonal, pairwise F_{ST}; below diagonal, exact test of differentiation (Goudet *et al*. 1996; Raymond & Rousset 1995). Significance assessed in Arlequin by either 10^3 permutations (F_{ST}) or 2×10^6 Markov chain steps (exact test); *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

equal to the ancestral population size. For the sake of comparison, likelihood ratio tests comparing each nested model to the full model rejected 20 of 24 nested models. The four that were not rejected, combined with the full model, represent 94.2% of the cumulative model weight from the information-theoretic analysis. Coalescent simulations under both the earliest Pleistocene (129 000 generations, 2.58 Ma) and upper Pleistocene (6300 generations, 126 ka) divergence scenarios were rejected (Table 4). However, divergence during the last glacial maximum (900 generations, 18 ka) was not rejected, and neither was the single population scenario under either the highest or lowest credible estimates for N_e. Inclusion of migration in these simulations did not qualitatively change the results.

Discussion

Three hypotheses

Our results do not favour two of the three hypotheses concerning the relationships of *Lilium pyrophilum* advanced by Skinner & Sorrie (2002). First, it is unlikely that *L. pyrophilum* simply represents a disjunct population of *L. iridollae*: the ITS phylogeny unambiguously allies *L. pyrophilum* with *L. superbum*, whereas *L. iridollae* is most closely related to *L. michauxii*. That *L. pyrophilum* and *L. iridollae* are independent only heightens the conservation concern of each of these rare species.

Second, the hypothesis that the species originated as a hybrid between *L. michauxii* and *L. superbum* is not supported by network analyses (Fig. 4). If *L. pyrophilum* represented a recent hybrid, single-copy nuclear loci should be related to both parental species. Instead, most *L. pyrophilum* and *L. superbum* haplotypes are closely related to each other (and many are shared), while they show less similarity to *L. michauxii*. The phylogenetic analysis of ITS sequences placed the *L. pyrophilum* samples with *L. superbum* sequences only, to the exclusion of the *L. michauxii* sequences.

Lilium pyrophilum appears to be a peripheral isolate of *L. superbum*. Our results indicate that the overall magnitude of divergence between the two lily species is very low and that the origin of *L. pyrophilum* is likely to have been very recent, i.e. during the latter Pleistocene or Holocene. Our estimated divergence date from the IMa2 analysis of 188 ka (Fig. 5b) would fall within the Illinoian glacial period. The minimum credible divergence time of 84 ka would seem to indicate that *L. pyrophilum* is in fact isolated from *L. superbum*. In spite of low F_{ST} values (Table 2), zero probability is assigned to the most recent divergence times in this analysis. The results of the simulation

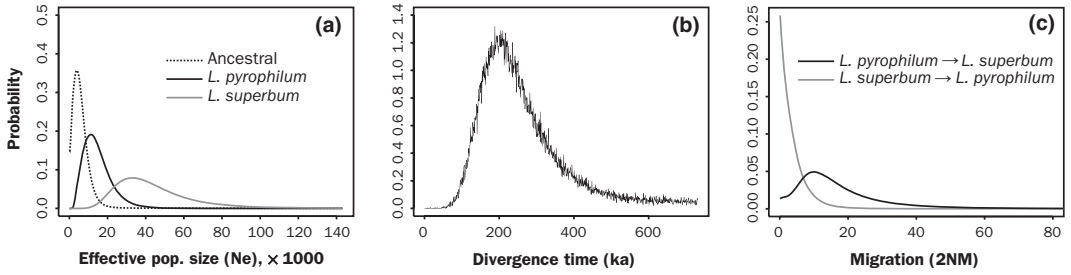

Fig. 5 Posterior probability distributions for IMa2 model parameters under the full model. (a) Effective population size for both species and the ancestral population. Both descendent taxa are inferred to have larger effective population size in this analysis. Estimated values for N_e are *Lilium pyrophilum*, 11 400 (95% HPD 2800–29 700); *L. superbum*, 32 900 (12 800–86 900); and ancestral, 3900 (0–14 400). (b) Divergence time. No probability is found for divergence times near zero; however, the distribution fails to reach zero at the upper end. The peak corresponds to a value of 188 (84–586) ka. (c) Migration rate. Highest probability for migration from *L. superbum* into *L. pyrophilum* is zero; there is, however, a higher probability of migration in the opposite direction (2 NM = 9.98).

Table 3 IMa2 analysis of nested models

Model description	log(*P*)	Terms	AIC	ΔAIC	Model weight	Cum. weight	d.f.	2LLR	*P*-value, LRT
θ (*pyrophilum*) = θ (*superbum*), *m* zero from *superbum* to *pyrophilum*	−4.442	3	14.884	0	0.301	0.301	2	2.986	0.2247
m zero from *superbum* to *pyrophilum*	−3.825	4	15.65	0.766	0.2052	0.5062	1	1.752	0.1856
Full IM model	−2.949	5	15.898	1.014	0.1813	0.6875	—	—	—
θ (*pyrophilum*) = θ (*superbum*)	−3.972	4	15.944	1.06	0.1772	0.8647	1	2.045	0.1527
Symmetrical migration	−4.803	4	17.606	2.722	0.0772	0.9419	1	3.707	0.0542
m zero from *pyrophilum* to *superbum*	−6.29	4	20.58	5.696	0.0174	0.9593	1	6.681	0.0097
θ (*pyrophilum*) = θ (ancestral), *m* zero from *superbum* to *pyrophilum*	−7.985	3	21.97	7.086	0.0087	0.968	2	10.07	0.0065
θ (*pyrophilum*) = θ (ancestral), *m* zero from *pyrophilum* to *superbum*	−8.116	3	22.232	7.348	0.0076	0.9757	2	10.33	0.0057
θ (*pyrophilum*) = θ (*superbum*), symmetrical migration	−8.408	3	22.816	7.932	0.0057	0.9814	2	10.92	0.0043
θ (*pyrophilum*) = θ (ancestral), symmetrical migration	−8.424	3	22.848	7.964	0.0056	0.987	2	10.95	0.0042
All θ equal, *m* zero from *superbum* to *pyrophilum*	−9.858	2	23.716	8.832	0.0036	0.9906	3	13.82	0.0032
θ (*pyrophilum*) = θ (ancestral)	−7.899	4	23.798	8.914	0.0035	0.9941	1	9.9	0.0017
θ (*superbum*) = θ (ancestral), *m* zero from *superbum* to *pyrophilum*	−9.192	3	24.384	9.5	0.0026	0.9967	2	12.49	0.0019
All θ equal	−9.858	3	25.716	10.832	0.0013	0.9981	2	13.82	0.001
θ (*superbum*) = θ (ancestral)	−9.192	4	26.384	11.5	0.001	0.999	1	12.49	0.0004
θ (*pyrophilum*) = θ (*superbum*), *m* zero from *pyrophilum* to *superbum*	−10.63	3	27.26	12.376	0.0006	0.9997	2	15.36	0.0005
θ (*superbum*) = θ (ancestral), symmetrical migration	−12.1	3	30.2	15.316	0.0001	0.9998	2	18.3	0.0001
All θ equal, symmetrical migration	−13.4	2	30.8	15.916	0.0001	0.9999	3	20.9	0.0001
θ (*superbum*) = θ (ancestral), zero migration	−14.26	2	32.52	17.636	0	0.9999	3	22.63	0
Zero migration	−13.35	3	32.7	17.816	0	1	2	20.8	0
θ (*superbum*) = θ (ancestral), *m* zero from *pyrophilum* to *superbum*	−14.26	3	34.52	19.636	0	1	2	22.63	0
All θ equal, *m* zero from *pyrophilum* to *superbum*	−18.52	2	41.04	26.156	0	1	3	31.13	0
θ (*pyrophilum*) = θ (*superbum*), zero migration	−24.86	2	53.72	38.836	0	1	3	43.83	0
θ (*pyrophilum*) = θ (ancestral), zero migration	−29.23	2	62.46	47.576	0	1	3	52.57	0
All θ equal, zero migration	−30.93	1	63.86	48.976	0	1	4	55.97	0

Models include the full IM model and 24 simpler nested models for the two-population case. Information-theoretic statistics, based on log(*P*), follow Burnham & Anderson (2002) and have been sorted by model weight. Models not rejected under traditional-likelihood ratio tests (LRT) are included in the 95% confidence set of models selected by AIC.

analysis lead us to interpret the IMa2 results with caution, however, because they reject divergence >6300 generations (126 ka) ago for each locus and fail to reject the scenarios with divergence at 900 generations (18 ka) and with no divergence (Table 4). The models tested in this approach, however, were simplified with respect to the full IMa2 model and treat each locus separately rather than jointly. Regardless of whether the IMa2 results or the coalescent simulation results are preferred, the isolation between the two taxa is not ancient. Mid- to late Pleistocene divergence times have been found in surprisingly few studies of plants (e.g. Strasburg & Rieseberg 2008; Bittkau & Comes 2009; Cooper *et al.* 2010).

Our results provide insight into the demographic patterns that have affected the two species. Deviations from neutral expectation indicate population expansion in both taxa (e.g. the average value for Tajima's *D* across three loci: *L. pyrophilum* = −1.35, *L. superbum* = −1.31, Table 1). This result is corroborated by the IMa2 analysis, which demonstrates modern effective population sizes higher than the ancestral, with the widespread *L. superbum* having a larger value (N_e 2.7 times that of the endemic *L. pyrophilum*, Fig. 5a). It is worth noting that the effective population size of *L. pyrophilum* (11 000 individuals) is surprisingly high considering the very small range of the species; in fact, our estimate of N_e is well in excess of the current census population size estimated by a recent inventory. Two factors may explain this discrepancy. First, our estimated generation time may be too low, which would cause us to overestimate effective population size (and underestimate divergence time). Second, agriculture, timber harvesting and fire suppression have

Table 4 Results of coalescent simulation study

	Marker		
Simulation model	AKT	AP	CP
Divergence time (in generations) without migration			
129 000	**0.000**	**0.000**	**0.000**
6300	**0.023**	**0.008**	**0.036**
900	0.992	0.379	0.394
Divergence time (in generations) with migration			
129 000	**0.001**	**0.001**	**0.006**
6300	**0.028**	**0.013**	**0.034**
900	0.839	0.438	0.422
No divergence			
High N_e	0.122	0.148	0.546
Low N_e	0.065	0.181	0.235

P-value for each model was obtained by comparison of either minimum (divergence) or maximum (no divergence) empirical *s* value (Slatkin & Maddison 1989) with simulated distributions of *s* under coalescent scenarios to test whether observed data were consistent with divergence times discussed in text. Simulations were based on assumed 20-year generation time.

dramatically transformed much of the landscape of the Sandhills over the past few hundred years, which may well have extirpated many populations. As these plants are long-lived outcrossers, too few generations may have elapsed for the impact of the current bottleneck to be fully reflected in the estimated N_e (Lande & Barrowclough 1987). Although our results suggest that the existing population has apparently been greatly reduced recently, much of the original genetic diversity remains and could be conserved, minimizing the impact of the present-day population bottleneck.

Gene flow is inferred from *L. pyrophilum* to *L. superbum*. Models including symmetrical migration are not weighted heavily compared with models that have zero or nearly zero gene flow from *L. superbum* to *L. pyrophilum* (Table 3). Presently, the two species are disjunct. However, the plants are visited by strong-flying pollinators, such as swallowtail butterflies and hummingbirds (Skinner 2002), and the seeds are adapted for wind dispersal. Why migration would be asymmetrical is unknown, but this could be explained by pollinator behaviour, dispersal or intrinsic barriers to gene flow.

Edaphic endemism in the Sandhills

The Sandhills pre-date the Pleistocene and may be substantially older, raising the possibility that some endemic taxa may have originated in the Pliocene or earlier and maintained populations in the region continuously. How might Pleistocene climate changes have affected the distribution of *Lilium* spp. in the coastal plain and effected the isolation of *L. pyrophilum*? While periods of severe climate change may eliminate edaphic endemics that are unable to migrate to areas with a suitable climate and substrate, edaphic endemics may in fact be likely to endure climate change in their geographic ranges. As their niches are defined more by soils than climate, they are likely to remain the best competitors on restrictive soils under a wide range of conditions. In fact, the degree of edaphic restriction exhibited by a species often varies with climate: populations may be widespread in environments with low competition and edaphically restricted in more favourable climates (Brooks 1987; Harrison *et al.* 2009).

The edaphic conditions that currently support populations of *L. pyrophilum* have probably been relatively stable, because the erosional process has no doubt continually exposed the interface between permeable and impermeable soils, creating seeps. Yet, the divergence between *L. pyrophilum* and *L. superbum* is comparatively recent. Genetic diversity of *L. pyrophilum*, while lower than that of *L. superbum*, is still high, making a vicariant scenario likely. Thus, the phenotypic divergence described by Skinner & Sorrie (2002) probably occurred in the context of large populations and substantial gene flow.

The combination of long-term persistence and recent divergence of *L. pyrophilum* indicates that this species descends from locally adapted populations that were stranded in the Sandhills as *L. superbum* retreated to higher elevations. It is not clear why the intervening Piedmont region supports neither taxon; however, many groups show a similar disjunction (Braun 1955; Sorrie & Weakley 2001). This study indicates that for these lilies, at least, the disjunction coincided with Pleistocene climate oscillations; this may apply to other taxa that share similar distributions. More in-depth studies of the *L. pyrophilum/L. superbum* system, using microsatellite markers, will quantify genetic structure within *L. pyrophilum*, and gene flow within and between *L. pyrophilum* and *L. superbum*. These more detailed analyses will improve estimates of divergence time and gene flow and identify populations of high conservation priority. Better understanding of this group will provide further insight into the role of edaphic specialization, possibly brought on by climate change, in promoting diversification.

Acknowledgements

We thank Fort Bragg Military Reservation and the Endangered Species Branch for logistic support and the Construction Engineering Research Laboratory (US Army Corps of Engineers Agreement #W9132T-07-2-0019) for funding. We also thank Xiang Liu, David Thomas, Patrick Zhou, Esther Ichugo, Matt

Cleary and Jacob Hilton for assistance with laboratory work and Marshall Wilson, Tom Phillips, Mac Alford, Rob Naczi, Gary Shurette, Viola Walker, Emil Devito, Heather Sullivan, Paul Manos, John Pogacnik, James Smith, Wayne Longbottom, Misty Buchanan and others for assistance locating *Lilium* populations.

References

Axelrod DI (1972) Edaphic aridity as a factor in angiosperm evolution. *The American Naturalist*, **106**, 311–320.

Ayling SC, Brown TA (2008) Novel methodology for construction and pruning of quasi-median networks. *BMC Bioinformatics*, **9**, 115.

Ayres DR, Ryan FJ (1999) Genetic diversity and structure of the narrow endemic *Wyethia reticulata* and its congener *W. bolanderi* (Asteraceae) using RAPD and allozyme techniques. *American Journal of Botany*, **86**, 344–353.

Baldwin BG (2005) Origin of the serpentine-endemic herb *Layia discoidea* from the widespread *L. glandulosa* (Compositae). *Evolution*, **59**, 2473–2479.

Baskauf CJ, McCauley DE, Eickmeier WG (1994) Genetic analysis of a rare and a widespread species of *Echinacea* (Asteraceae). *Evolution*, **48**, 180–188.

Bittkau C, Comes HP (2009) Molecular inference of a Late Pleistocene diversification shift in *Nigella* s. lat. (Ranunculaceae) resulting from increased speciation in the Aegean archipelago. *Journal of Biogeography*, **36**, 1346–1360.

Braun EL (1955) The phytogeography of unglaciated Eastern United States and its interpretation. *The Botanical Review*, **21**, 297–375.

Brooks RR (1987) *Serpentine and its Vegetation: A Multidisciplinary Approach*. Dioscorides Press, Portland, OR.

Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS (2001) Comparative phylogeography of northwestern North America: a synthesis. In: *Integrating Ecology and Evolution in a Spatial Context* (eds Silvertown J, Antonovics J), pp. 319–340. Blackwell Science, Oxford.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York.

Cabe S, Nickerson JG, Hoffman CW, Farrell K (1992) Cretaceous and Tertiary stratigraphy of the northern Sandhills area, North Carolina. In: *Geologic field guides to North Carolina and vicinity* (eds Dennison J, Stewart KG), Annual Meeting of the Southeastern Section of the Geological Society of America. Winston-Salem, NC: Geological Society of America, Boulder, CO.

Carstens BC, Stoute HN, Reid NM (2009) An information-theoretical approach to phylogeography. *Molecular Ecology*, **18**, 4270–4282.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cooper EA, Whittall JB, Hodges SA, Nordborg M (2010) Genetic variation at nuclear loci fails to distinguish two morphologically distinct species of *Aquilegia*. *PLoS ONE*, **5**, e8655.

Crawford DJ, Ornduff R, Vasey MC (1985) Allozyme variation within and between *Lasthenia minor* and its derivative species, *L. maritima* (Asteraceae). *American Journal of Botany*, **72**, 1177–1184.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

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

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Edwards CE, Soltis DE, Soltis PS (2006) Molecular phylogeny of *Conradina* and other scrub mints (Lamiaceae) from the southeastern USA: evidence for hybridization in Pleistocene refugia? *Systematic Botany*, **31**, 193–207.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics*, **24**, 217–242.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405.

Freeland JR, Biss P, Conrad KF, Silvertown J (2010) Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. *Journal of Evolutionary Biology*, **23**, 776–782.

Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. *Evolutionary Biology*, **30**, 93–120.

Gitzendanner MA, Soltis PS (2000) Patterns of genetic variation in rare and widespread plant congeners. *American Journal of Botany*, **87**, 783–792.

Godt MJW, Hamrick JL (1993) Genetic diversity and population structure in *Tradescantia hirsuticaulis* (Commelinaceae). *American Journal of Botany*, **80**, 959–966.

Goudet J, Raymond M, de-Meeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Gugger PF, Sugita S, Cavender-Bares J (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses from the fossil record. *Molecular Ecology*, **19**, 1877–1897.

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources* (eds Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer Associates, Sunderland, MA.

Harrison S, Damschen E, Going BM (2009) Climate gradients, climate change, and special edaphic floras. *Northeastern Naturalist*, **16**, 121–130.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, **104**, 2785–2790.

Howarth DG, Baum DA (2005) Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of

Scaevola (Goodeniaceae) in the Hawaiian Islands. *Evolution*, **59**, 948–961.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Karron JD (1987) A comparison of levels of genetic polymorphism and self-compatibility in geographically restricted and widespread plant congeners. *Evolutionary Ecology*, **1**, 47–58.

Knowles LL (2009) Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 593–612.

Kruckeberg AR (1986) An essay: the stimulus of unusual geologies for plant speciation. *Systematic Botany*, **11**, 455–463.

Kruckeberg AR, Rabinowitz D (1985) Biological aspects of endemism in higher plants. *Annual Review of Ecology and Systematics*, **16**, 447–479.

Lande R, Barrowclough GF (1987) Effective population size, genetic variation, and their use in population management. In: *Viable Populations for Conservation* (ed. Soulé ME), pp. 87–124. Cambridge University Press, Cambridge.

Lande R, Engen S, Sæther B-E (2003) *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, Oxford.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Maddison WP, Maddison DR (2010) *Mesquite: A Modular System for Evolutionary Analysis*. http://mesquiteproject.org.

Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics*, **21**, 260–262.

Mehta BD, Jog SP, Johnson SC, Murthy PPN (2006) Lily pollen alkaline phytase is a histidine phosphatase similar to mammalian multiple inositol polyphosphate phosphatase (MINPP). *Phytochemistry*, **67**, 1874–1886.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Nishikawa T, Okazaki K, Uchino T, Arakawa K, Nagamine T (1999) A molecular phylogeny of *Lilium* in the internal transcribed spacer region of Nuclear ribosomal DNA. *Journal of Molecular Evolution*, **49**, 238–249.

North Carolina Natural Heritage Program (2007) *North Carolina Natural Heritage Program Element Occurrence Database*. NC Department of Environment and Natural Resources, Raleigh, NC, 27699–1610.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Oliveira LO, Huck RB, Gitzendanner MA, Judd WS, Soltis DE, Soltis PS (2007) Molecular phylogeny, biogeography, and systematics of *Dicerandra* (Lamiaceae), a genus endemic to the southeastern United States. *American Journal of Botany*, **94**, 1017–1027.

Ornduff R (1965) Ornithocoprophilous endemism in Pacific basin angiosperms. *Ecology*, **46**, 864–867.

Parsons RF (1976) Gypsophily in plants-a review. *American Midland Naturalist*, **96**, 1–20.

Peruzzi L, Leitch IJ, Caparelli KF (2009) Chromosome diversity and evolution in Liliaceae. *Annals of Botany*, **103**, 459–475.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Rabinowitz D (1981) Seven forms of rarity. In: *The Biological Aspects of Rare Plant Conservation* (ed. Synge H), pp. 205–217. Wiley, New York.

Raduski AR, Rieseberg LH, Strasburg JL (2010) Effective population size, gene flow, and species status in a narrow endemic sunflower, *Helianthus neglectus*, compared to its widespread sister species, *H. petiolaris*. *International Journal of Molecular Sciences*, **11**, 492–506.

Rajakaruna N (2004) The edaphic factor in the origin of plant species. *International Geology Review*, **46**, 471–478.

Rambaut A (1996) *Se-Al: Sequence Alignment Editor*. Available at http://evolve.zoo.ox.ac.uk/.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution*, **49**, 1280–1283.

Rieseberg LH, Carter R, Zona S (1990) Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution*, **44**, 1498–1511.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, **132**, 365–386.

Schafale MP, Weakley AS (1990) *Classification of the Natural Communities of North Carolina, Third Approximation*. North Carolina Department of Environment, Health, and Natural Resources, Division of Parks and Recreation, Natural Heritage Program, Raleigh.

Shaw J, Lickey EB, Schilling EE, Small RL (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.

Skinner MW (2002) *Lilium*. In: *Flora of North America North of Mexico. Magnoliophyta: Liliidae: Liliales and Orchidales* (ed. Flora of North America Editorial Committee), pp. 172–197. Oxford University Press, Oxford.

Skinner MW, Sorrie BA (2002) Conservation and ecology of *Lilium pyrophilum*, a new species of Liliaceae from the Sandhills region of the Carolinas and Virginia, USA. *Novon*, **12**, 94–105.

Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, **123**, 603–613.

Sohl NF, Owens JP (1991) Cretaceous stratigraphy of the Carolina Coastal Plain. In: *The Geology of the Carolinas: Carolina Geological Society Fiftieth Anniversary Volume* (eds Horton JW, Zullo VA, Carolina Geological Society), pp. 191–220. Univ. of Tennessee Press, Knoxville, TN.

Soltis PS, Soltis DE (2003) Applying the bootstrap in phylogeny reconstruction. *Statistical Science*, **18**, 256–267.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Sorrie BA, Weakley AS (2001) Coastal plain vascular plant endemics: phytogeographic patterns. *Castanea*, **66**, 50–82.

Sorrie BA, Gray JB, Crutchfield PJ (2006) The vascular flora of the longleaf pine ecosystem of Fort Bragg and Weymouth Woods, North Carolina. *Castanea*, **71**, 129–161.

Stanford AM, Harden R, Parks CR (2000) Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *American Journal of Botany*, **87**, 872–882.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Strasburg JL, Rieseberg LH (2008) Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*—large effective population sizes and rates of long-term gene flow. *Evolution*, **62**, 1936.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer, Sunderland, MA.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Wall WA, Douglas NA, Xiang QY, Hoffmann WA, Wentworth TR, Hohmann MG (2010) Evidence for range stasis during the latter Pleistocene for the Atlantic Coastal Plain endemic genus, *Pyxidanthera* Michaux. *Molecular Ecology*, **19**, 4302–4314.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wheelan SJ, Church DM, Ostell JM (2001) Spidey: a tool for mRNA-to-genomic alignments. *Genome Research*, **11**, 1952–1957.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplifications and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: A Guide to Methods and Applications* (eds Innis M, Gelfand D, Sninsky J, White T), pp. 315–322. Academic Press, San Diego, CA.

Whittall JB, Syring J, Parks M *et al.* (2010) Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology*, **19**, 100–114.

Williamson PS, Bazeer SK (1997) Self-incompatibility in *Abronia macrocarpa* (Nyctaginaceae). *Southwestern Naturalist*, **42**, 409–415.

Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics*, **23**, 1851–1853.

Zonneveld BJM, Leitch IJ, Bennett MD (2005) First nuclear DNA amounts in more than 300 angiosperms. *Annals of Botany*, **96**, 229–244.

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. http://code.google.com/p/garli/.

N.A.D. is an evolutionary biologist who focuses on phylogenetics and phylogeography of arid-adapted and edaphic endemic plants. W.A.W. is a plant ecologist interested in the interplay between population genetics and population dynamics of rare and endemic species. Q.Y.X.'s research explores the mechanisms underlying biodiversity patterns and morphological variation, with emphasis on the Cornales. W.A.H. is interested in ecology and conservation of savanna ecosystems. T.R.W. is a vegetation scientist with research interests in vegetation/environment relationships and biological diversity. J.B.G. is a botanist working on conservation of rare plants in Coastal Plain habitats. M.G.H. uses diverse approaches to inform rare species conservation.

Data accessibility

Sample and haplotype information is found in Table S1 (Supporting Information). DNA sequences: GenBank accessions JF829316–JF829423 (Table S2, in Supporting Information). ITS data and phylogenetic tree available at http://purl.org/phylo/treebase/phylows/study/TB2:S11519.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Sampling. Haplotype numbers correspond to sample labels in Figs 3 and 4, and to names accessioned in GenBank (Table S2, in Supporting Information). For AP and AKT sequences, integers identify haplotypes recovered more than once in this study and other identifiers refer to unique haplotypes

Table S2 GenBank accession numbers. Haplotype names correspond to samples in Table S1 (Supporting Information)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 4302–4314 doi: 10.1111/j.1365-294X.2010.04793.x

Evidence for range stasis during the latter Pleistocene for the Atlantic Coastal Plain endemic genus, *Pyxidanthera* Michaux

WADE A. WALL,* NORMAN A. DOUGLAS,* QIU-YUN (JENNY) XIANG, * WILLIAM A. HOFFMANN,* THOMAS R. WENTWORTH* and MATTHEW G. HOHMANN†

**Department of Plant Biology, Box 7612, North Carolina State University, Raleigh, NC 27695, USA, †US Army Corps of Engineers, Engineer Research and Development Center, PO Box 9005, Champaign, IL 61826, USA*

Abstract

The general phylogeographical paradigm for eastern North America (ENA) is that many plant and animal species retreated into southern refugia during the last glacial period, then expanded northward after the last glacial maximum (LGM). However, some taxa of the Gulf and Atlantic Coastal Plain (GACP) demonstrate complex yet recurrent distributional patterns that cannot be explained by this model. For example, eight co-occurring endemic plant taxa with ranges from New York to South Carolina exhibit a large disjunction separating northern and southern populations by >300 km. *Pyxidanthera* (Diapensiaceae), a plant genus that exhibits this pattern, consists of two taxa recognized as either species or varieties. We investigated the taxonomy and phylogeography of *Pyxidanthera* using morphological data, cpDNA sequences, and amplified fragment length polymorphism markers. Morphological characters thought to be important in distinguishing *Pyxidanthera barbulata* and *P. brevifolia* demonstrate substantial overlap with no clear discontinuities. Genetic differentiation is minimal and diversity estimates for northern and southern populations of *Pxyidanthera* are similar, with no decrease in rare alleles in northern populations. In addition, the northern populations harbour several unique cpDNA haplotypes. *Pyxidanthera* appears to consist of one morphologically variable species that persisted in or near its present range at least through the latter Pleistocene, while the vicariance of the northern and southern populations may be comparatively recent. This work demonstrates that the refugial paradigm is not always appropriate and GACP endemic plants, in particular, may exhibit phylogeographical patterns qualitatively different from those of other ENA plant species.

Keywords: amplified fragment length polymorphism, cpDNA, Diapensiaceae, phylogeography, Pleistocene, refugium

Received 6 April 2010; revision received 28 June 2010; accepted 7 July 2010

Introduction

The alternating glacial and interglacial periods that characterized the Pleistocene had major impacts on the biogeography and genetic diversity of plant species in the Northern Hemisphere (Comes & Kadereit 1998; Hewitt 2000). The LGM, approximately 18 000 years BP, saw the Laurentide ice sheet reach its southern extent in eastern North America (ENA) (Ehlers & Gibbard 2004). The primary scenario describing plant species' ranges during and following the LGM in ENA includes (i) range contraction to southern refugia (Delcourt & Delcourt 1981) and (ii) subsequent recolonization of northern habitats after the retreat of the glaciers (Dorken & Barrett 2004). Previous studies have identified the resulting phylogeographical patterns of plant species in ENA and made inferences about possible refugia during the glacial maxima (Soltis *et al.* 2006). Several such patterns have been identified in the ranges

Correspondence: Wade A. Wall, Fax: (919) 515 3436;
E-mail: wade.wall@gmail.com

of ENA plant and animal species. These include east–west divisions between Gulf Coast and Atlantic Coast populations (Mylecraine *et al.* 2004), phylogeographical separation by river drainage systems (Church *et al.* 2003), and the identification of refugia closer to the glacial front, either in the southern Appalachians (Gonzales *et al.* 2008) or farther north (Magni *et al.* 2005; McLachlan *et al.* 2005).

Eastern North America has generally been divided into four physiogeographical regions: Interior Lowlands, Appalachian Highlands, Piedmont, and Gulf and Atlantic Coastal Plain (GACP) (Fenneman 1938). Compared to the other regions, the GACP is well defined geologically and floristically (Takhtajan 1986), but little is known about the phylogeography of the many widespread species that are endemic to the GACP. Previous phylogeographical studies of GACP plant species have generally focused on narrow endemics with small latitudinal ranges (Evans *et al.* 2000; Oliveira *et al.* 2007) or species with ranges that cross multiple physiographical regions (Morris *et al.* 2008; however, see Mylecraine *et al.* 2004). This is unfortunate; over 1300 species and 47 genera are endemic to the region (Sorrie & Weakley 2001), the second-highest concentration in the United States and only exceeded by the California Floristic Province (Flora of North America Editorial Committee 1993). Without taking into account the endemic species of the GACP and their distributional patterns, any attempt to understand the postglacial phylogeography of ENA is limited.

Sorrie & Weakley (2001) documented 27 different recurrent distributional patterns for endemic plant species of the GACP. One of the most interesting of these patterns is the disjunct distribution of eight taxa (*Calamovilfa brevipilis*, *Dichanthelium hirstii*, *Eupatorium resinosum*, *Gentiana autumnalis*, *Lobelia canbyi*, *Narthecium americanum*, *Pyxidanthera barbulata*, and *Rhynchospora pallida*) that occur in New York and New Jersey and eastern North and South Carolina, but not in the intervening areas of Maryland, Delaware, and most of Virginia. In addition to these eight taxa, numerous species exhibit the same disjunction between New Jersey and the southern GACP, but are more widespread in the southern part of the GACP. Common distributional patterns may be the result common biogeographical processes, but there is always the possibility of pseudocongruence (Hafner & Nadler 1990; Cunningham & Collins 1994).

We focus here on the genus *Pyxidanthera* Michaux as a case study to investigate the refugial paradigm in the GACP. *Pyxidanthera* is in Diapensiaceae, a small family with a circumboreal distribution, with some taxa extending southward into eastern Asia and ENA. The genus includes two recognized species; both are woody, winter-flowering, evergreen cushion plants. Populations of the more widespread *P. barbulata* occur on Long Island in New York, the Pine Barrens of New Jersey, several locations in southeastern Virginia, and the coastal plain of North Carolina and South Carolina. *P. brevifolia* has a more limited range; it has only been documented in six counties in the Sandhills region of North Carolina and South Carolina. *P. brevifolia*, currently under intensive study as a species at risk by the US Department of Defense, is considered vulnerable to extinction in North Carolina, with over 80% of the North Carolina populations confined to Fort Bragg Military Reservation, NC (Buchanan & Finnegan 2008). *P. brevifolia* is nearly restricted to xeric sandhill scrub communities within the long-leaf pine ecosystem (Schafale & Weakley 1990; Sorrie *et al*. 2006), one of the most imperiled ecosystems in North America, with approximately 2% of the historical area currently extant (Frost 2006). In addition to containing most of the remaining *P. brevifolia* populations, Fort Bragg Military Reservation is also one of the few places where the two species of *Pyxidanthera* co-occur. When sympatric, *P. barbulata* and *P. brevifolia* occupy nonoverlapping ecological habitats, with *P. barbulata* occupying wetter sites such as pocosin ecotones and *P. brevifolia* occurring on extremely xeric sand ridges.

Pyxidanthera was monotypic until 1929, at which time *P. brevifolia* was separated from sympatric populations of *P. barbulata* in the Sandhills region of North Carolina and upper South Carolina based on habitat differences, shorter leaves, and dense pubescence relative to the more widespread *P. barbulata* (Wells 1929). Differing ecological niches and morphological characters of *P. barbulata* and *P. brevifolia* led to a debate regarding the proper taxonomic status of the two taxa. In 1964, *P. brevifolia* was reduced to a variety of *P. barbulata* without comment (Ahles 1964). Afterwards, several studies investigated the appropriate taxonomic status of *P. brevifolia*. An embryological study (Reynolds 1966) concluded that both notable developmental similarities and differences existed between the two species and ultimately relied on the ecological and morphological differences to support the continued recognition of two species. Primack & Wyatt (1975) found correlation between leaf length and soil moisture of *P. brevifolia and P. barbulata* at a single site in South Carolina and concluded that the difference in leaf length between the two species is clinal, suggesting that *P. brevifolia* is simply a morphological variant of *P. barbulata*. More recently, an allozyme study – restricted to the populations from the southern range of the genus – found that the two species share similar levels of genetic diversity, with very little intertaxa genetic differentiation (Godt & Hamrick 1995). However, recent floras for the region

(Weakley 2008; Sorrie *et al.* 2009) have continued to recognize two species, emphasizing the morphological, ecological, and embryological differences between them.

In this study, we use cpDNA sequences, amplified fragment length polymorphism (AFLP) markers, and morphological measurements to investigate the taxonomy and phylogeography of both *P. barbulata* and *P. brevifolia* across the entire range of the genus. We attempt to determine whether clear morphological and genetic differences exist between the two species and whether the morphological, ecological, and embryological variation previously observed in the southern populations of *Pyxidanthera* (in the text, *Pyxidanthera* will refer to both *P. barbulata* and *P. brevifolia*) correlate with greater genetic diversity in the south. Using genetic data, we attempt to distinguish between two plausible phylogeographical scenarios. The first scenario represents typical refugial patterns described for numerous species in ENA; the genus *Pxyidanthera* was isolated in one or more southern refugia during the Pleistocene and subsequently recolonized northern areas after the LGM. Genetic patterns supporting this scenario would include reduced genetic diversity in northern populations (Hewitt 2000), recolonized areas containing only a subset of refugial population alleles (Broyles 1998), and putative refugia having a greater number of rare alleles, which may reflect historical processes better than genetic diversity estimates (Comps *et al.* 2001; Paun *et al.* 2008) Alternatively, the two species in *Pyxidanthera* could have persisted in their present ranges through the later Pleistocene rather than retreating into one or more glacial refugia. Genetic patterns that would suggest this second scenario include no reduction in genetic diversity or rare alleles in northern populations and the presence of alleles restricted to northern populations.

Materials and methods

Sampling and morphological measurements

We collected leaf tissue samples of 423 individuals from 29 *Pyxidanthera brevifolia* populations (defined as all *P. brevifolia* individuals that occurred within 0.75 km of each other) and 178 individuals from 14 *P. barbulata* populations, across the ranges of both species. A priori taxonomic identity was determined based on habitat differences, State Natural Heritage Program records, and geographical region (*P. brevifolia* is restricted to the Sandhills region of North and South Carolina). For each sample, we measured the longest leaf length and width and categorized the leaf pubescence into one of two categories: pubescence covering more than half of the leaf, and pubescence covering half or less of the leaf. We evaluated differences in leaf length and leaf width means between *P. barbulata* and *P. brevifolia* using *t*-tests and evaluated differences in leaf pubescence categories using a chi-square test.

Molecular methods

DNA was extracted from 319 *P. brevifolia* individuals across 17 populations and 157 *P. barbulata* individuals across 14 populations using the CTAB method with minor modifications (Doyle & Doyle 1987). After an initial screening of 16 cpDNA regions known to be highly polymorphic (Shaw *et al.* 2007), we amplified two polymorphic regions – *atpI-atpH* and *psbD-trnT*(GUU) – of 63 and 42 samples from 14 and 10 populations of *P. barbulata* and *P. brevifolia*, respectively, using universal primer pairs (Shaw *et al.* 2007). PCR conditions followed Shaw *et al.* (2005) in 12.5 μL solutions using the following protocol: 1 hold (5 min per 80 °C), 30 cycles [(1 min per 95 °C), (1 min per 50 °C), (4 min per 65 °C)], 1 hold (5 min per 65 °C). PCR products were cleaned prior to sequencing using Antarctic Phosphatase (0.5 Units), Exonuclease I (0.2 Units), and 1 μL 10× Antarctic Phosphatase buffer (New England BioLabs, Ipswich, MA, USA) at 37 °C (15 min) and 80 °C (15 min). We sequenced in the forward direction for the *atpH* and *psbD* regions using the Big Dye 3.1 kit (Applied Biosystems, USA) and analysed the products using an ABI 3730 DNA sequencer (Applied Biosystems, USA). We edited and aligned sequences using Sequencher 4.2.2 (Gene Codes Corporation, Ann Arbor, MI, USA) and MEGA version 4 (Tamura *et al.* 2007). Sequences were submitted to GenBank under accession numbers nos. HM564379-HM56491 (Table S1).

Amplified fragment length polymorphism markers are appropriate for examining low levels of genetic divergence within and between closely related taxa (Coart *et al.* 2002; McKinnon *et al.* 2008) and have been successfully used in phylogeographical studies (Meudt & Bayly 2008; Perez-Collazos *et al.* 2009). AFLP genotyping followed the multiplexing protocol described by Trybush *et al.* (2006), with the minor modification that the restriction and ligation steps were combined in a single reaction at a total volume of 10 μL. For the preamplification reaction, we used EcoRI+A and *Mse*I+C primers. Three selective primer pairs were chosen after a trial based on the number of reproducible polymorphic markers produced: Eco-ACC/*Mse*I-CAT (Hex), Eco-ATG/*Mse*I-CAT (FAM), and Eco-AGG/*Mse*I-CAT (NED). Selective amplification products were separated and analysed using an ABI 3730 DNA sequencer (Applied Biosystems, USA) and automatically scored with Genemarker version 1.8 (Softgenetics LCC, State College, PA, USA) using the default settings, with the

exception that we normalized the FAM-dyed markers and set the allele evaluation peak score to 'pass' if it was ≥1 (Holland *et al*. 2008). The reproducibility of the AFLP profiles was evaluated by running eight duplicate samples for each 96-well plate. Error rates between duplicates were calculated using a Euclidean distance measure (Bonin *et al*. 2004). To reduce the error rate, we removed bands with 10 or more errors when comparing duplicate samples (Zhang *et al*. 2010).

cpDNA data analysis

Because the chloroplast represents a single nonrecombining locus, sequences of the two sampled regions were concatenated. We recoded insertions or deletions (indels) that did not violate the assumptions of the infinite sites model (Kimura 1969) as identified by SNAP Workbench (Price & Carbone 2005). We performed three separate tests of neutrality to test for evidence of population expansion or selection in the cpDNA – Fu and Li's D* and F*(Fu & Li 1993) and Fu's F_S (Fu 1997) – using SNAP Workbench. Fu and Li's D* and F* neutrality tests are more powerful for detecting background selection, while Fu's F_s is more powerful for detecting population growth (Ramos-Onsins & Rozas 2002). We estimated an unrooted haplotype network using the haploNet function as implemented in the pegas package (Paradis 2009) in R (R Development Core Team 2009). This package implements the statistical parsimony method for network reconstruction (Templeton *et al*. 1992). We performed two analyses of molecular variance (AMOVA) using Arlequin version 3.01 (Excoffier *et al*. 2005) with the data set hierarchically partitioned by region and individual populations within regions (with two regions defined as New Jersey and New York, hereafter referred to as northern populations, and Virginia, North Carolina, and South Carolina, hereafter referred to as southern populations) and by species (*P. barbulata* and *P. brevifolia*) and populations within species. We estimated the nucleotide genetic diversity (π) (Nei 1987) for each population using DnaSP version 5 (Librado & Rozas 2009).

To test for phylogeographical structure in the data set, we compared two measures of genetic differentiation between populations – G_{ST}, based on haplotype frequency, and N_{ST}, by similarities between haplotype sequences – using PERMUT 2.0 (Pons & Petit 1996) with 1000 permutations. If N_{ST} is significantly greater than G_{ST}, it is taken as evidence of a phylogeographical signal in the data set. To test for isolation by distance (IBD), we performed Mantel tests using the R package vegan (Oksanen *et al*. 2009) between the log-transformed geographical distance matrix and the pairwise population N_{ST} matrix as calculated in DnaSP version 5 (Librado & Rozas 2009) for all populations and for the southern and northern populations separately.

We reconstructed the gene genealogy for the sampled chloroplast regions using Genetree version 9.0 (Bahlo & Griffiths 2000) as implemented in SNAP Workbench (Price & Carbone 2005). We estimated the population mutation rate (θ), using Watterson's method (1975) as calculated in Genetree for both geographical regions and used the average between the two regions as the starting θ. Because of the larger geographical area covered by the two species of *Pyxidanthera* in the southern populations, we assumed a model of unequal population sizes, with the southern population twice as large as the northern population, and nonexponential population growth. We performed ten independent simulations with different starting values of 10^6 iterations, selecting the rooted genealogy and mutation age estimates with the highest probability.

To simultaneously analyse the effects of incomplete lineage sorting and gene flow on the genetic structure of the northern and southern *Pyxidanthera*, we employed an isolation with migration model of population divergence (Nielsen & Wakeley 2001) implemented in the program IMa2 (Hey & Nielsen 2007). IMa2 estimates the following parameters based on the genetic data: θ for all populations (extant and ancestral), migration parameters (m) for gene flow between populations, and t, time in coalescent units since divergence of the extant populations. We performed three independent runs with ten chains each under an infinite sites model with a burn-in period of 150 000 steps. We sampled 500 000 genealogies , saving one genealogy every 100 steps. We evaluated proper mixing based on the absence of trends in plotted parameter estimates and congruence of parameter estimates between runs 100 000 of the 500 000 saved genealogies were combined to evaluate 24 models that were either nested within the full model or that constrained select parameters by setting them equal to each other (e.g. equal migration between populations). We compared the different model posterior probabilities using an information-theoretic approach recently extended to phylogeographical data (Carstens *et al*. 2009). Information theory statistics were calculated according to Burnham and Anderson (2002).

AFLP data analysis

We calculated the percentage of polymorphic loci (P%) and Nei's expected heterozygosity (Nei 1987) using AFLPsurv version 1.0 (Vekemans *et al*. 2002) and the 'frequency down weighted marker score' (DW) (Schönswetter & Tribsch 2005) using the R script AFLPdat (Ehrich 2006); several population genetic diversity

measures were included to ensure consistency between methods. DW is calculated by summing each occurrence of a particular marker in a population and dividing that value by the sum of the marker across all populations. For each population, these values are then averaged across all markers. Populations that have been isolated are expected to accumulate rare markers and thus their DW scores should be higher. We first removed populations that contained fewer than seven samples to minimize effects of low sample size (Bonin *et al.* 2007), leaving a total of 437 samples from 25 populations. We tested for effects of sample size on all the calculated genetic diversity estimates by regressing estimated diversity on sample size, and we also tested for correlation between all possible pairings of the included diversity measures. Populations were grouped according to taxonomic identity and region, and we tested for significant differences between diversity estimates using *t*-tests in R.

Population differentiation and structure were explored by first running an ordination using nonmetric multidimensional scaling (NMDS) to graphically display population pairwise genetic distances (D) (Nei 1972) in a reduced dimensional space using the R package labdsv (Roberts 2010). We included all populations regardless of sample size for the analysis. In addition, we explored population genetic structure using STRUCTURE 2.3.2.1 (Pritchard *et al.* 2000; Falush *et al.* 2007). For K1 through 9, we performed three runs with a burn-in length of 10 000 and post burn-in length of 25 000, assuming admixture and correlated allele frequencies. We determined the most likely number of populations by graphically analysing the model log likelihoods for each K. Because Ln *P*(D) did not increase monotonically to the optimal K (Herrera & Bazaga 2008), we did not use the methods of Evanno *et al.* (2005). Three analyses of molecular variance (AMOVA) were performed using Arlequin version 3.01 (Excoffier *et al.* 2005), with partitioning of the data following the chloroplast AMOVAs. To test for IBD, we performed a Mantel test between the population genetic distance matrix and the log-transformed geographical distance matrix using the R package vegan (Oksanen *et al.* 2009).

Results

Morphology

Pyxidanthera barbulata has significantly longer leaf lengths (6.3 mm vs. 4.5 mm, respectively, $P < 0.001$) and widths (1.9 mm vs. 1.3 mm, $P < 0.001$) compared to *P. brevifolia*, but there is considerable overlap between the two species in both traits (Fig. 1). The variation in leaf length is continuous between *P. barbulata* and *P. brevifolia* with no obvious break, certainly not at the 3.5–4 mm size suggested in taxonomic keys (Sorrie *et al.* 2009). There is a significant difference in leaf pubescence between the two species ($P < 0.001$). All *P. barbulata* had pubescence covering less than half of their leaves, but 49% of *P. brevifolia* also had pubescence covering more than half of their leaves. As with leaf length and width, there is considerable variation within taxa.

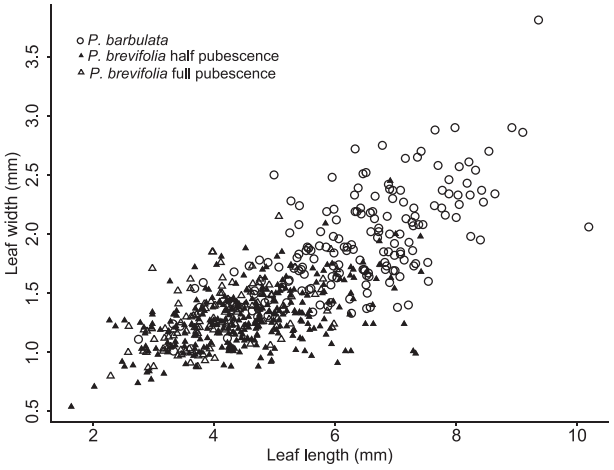

Fig. 1 Morphological variation in leaf length, leaf width, and pubescence of *Pyxidanthera barbulata* (circles) and *P. brevifolia* (triangles). Solid triangles represent *P. brevifolia* specimens that had pubescence for half or less than half of the leaf; open triangles represent *P. brevifolia* specimens with pubescence greater than half of the leaf. Although there are statistically significant differences between the two species for mean leaf length, leaf width, and pubescence, there is considerable overlap in the ranges of these traits between the two species.

cpDNA

The two sampled cpDNA regions for 105 individuals yielded 975 characters, of which 14 were polymorphic (Table S2). The data set included 12 substitutions and two indels that did not violate the infinite sites model. None of the three neutrality tests (Fu and Li's D* and F* and Fu's Fs) were significant ($P > 0.05$), indicating that there is no evidence of either population growth or background selection. A statistical parsimony haplotype tree identified 12 haplotypes (Fig. 2). The interior haplotypes of the network (H7, H11) are geographically widespread compared to the derived haplotypes, which tend to be both less frequent and geographically restricted (Fig. 2; Table S2). Results from Genetree indicate that H11 is the haplotype with the highest probability of being ancestral (average relative likelihood 69.5%); however, both of the interior haplotypes were

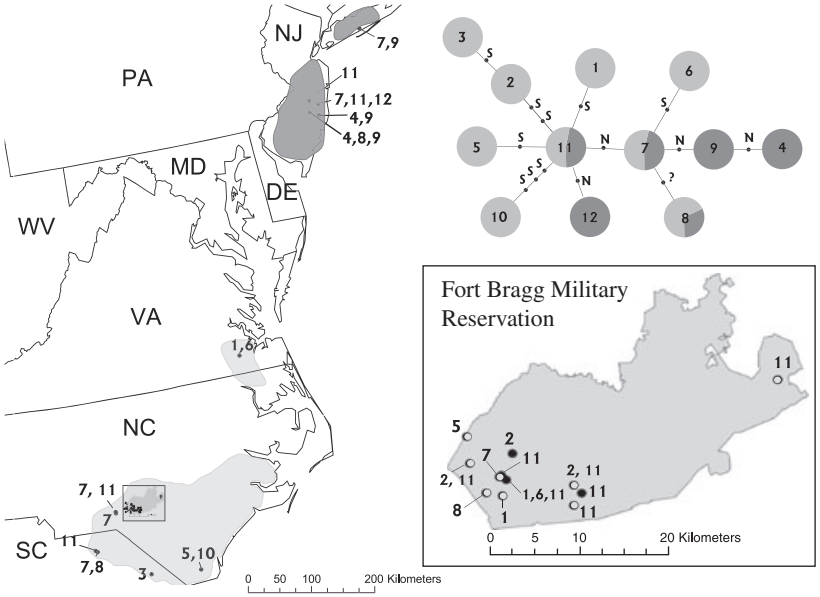

Fig. 2 Geographical distribution (shaded in grey) and statistical parsimony network for 12 haplotypes from 2 cpDNA regions of *Pyxidanthera*. State names are in bold abbreviations and numbers represent haplotypes from Table S2. Black dots in the haplotype network represent mutational steps; associated letters (S for South, N for North) represent the most likely (>95% probability) geographical origin of mutations inferred using Genetree 9.0. Light grey shading of haplotype network represents proportion of the associated haplotype comprised of southern individuals and darker grey shading represents proportion comprised of northern individuals. Inset map: Sampling of *Pyxidanthera* populations on Fort Bragg Military Reservation. *Pyxidanthera barbulata* populations are represented by closed circles and *P. brevifolia* populations are represented by open circles.

almost equally common in the northern and southern populations (Fig. 2). Two of the four haplotypes derived from H7 (H4 and H9) only occur in New Jersey and New York and had the highest probability of a northern origin, while the other two derived haplotypes (H6 and H8) occur in both northern and southern population. Five of the six haplotypes derived from the second interior haplotype, H11, only occur in southern populations and most likely are of southern origin. Only H12 has a higher probability of a northern origin; it is a private haplotype restricted to one northern population.

Region explains a small but statistically significant percentage of the genetic variation when used as the highest grouping variable in a hierarchical AMOVA (17.27%, $P < 0.05$), revealing that genetic variation is not evenly spread across the northern and southern populations (Table 1). Populations within regions explain most of the variation (56.58%, $P < 0.001$); within-population genetic differences and region

Table 1 Analyses of molecular variance (AMOVA) results for *Pyxidanthera barbulata* and *P. brevifolia* using cpDNA sequences and amplified fragment length polymorphism (AFLP) markers

	AFLP			cpDNA		
Source of variation	d.f.	Variance	% of variation	d.f.	Variance	% of variation
Grouped by species						
Between species	1	0.37	1.60**	1	−0.01	−1.71 NS
Among populations within species	23	1.85	7.87***	22	0.63	72.22***
Within populations	412	21.30	90.53***	81	0.26	29.49***
Grouped by region (North vs. South)						
Between regions	1	0.76	3.20**	1	0.17	17.27*
Among populations within regions	23	1.86	7.79***	22	0.56	56.58***
Within populations	412	21.30	89.01***	81	0.26	26.15***

***Indicates *P*-value <0.001, ** *P*-value <0.01, * *P*-value <0.05, and NS indicates nonsignificance of variation.

account for a smaller but still significant percentage of the variation (26.15%, $P < 0.001$). When species is used as the highest grouping variable, AMOVA results demonstrate significant genetic differences among populations (72.22%, $P < 0.001$), but not significant differences between the two species (0%, $P > 0.05$). The nucleotide genetic diversity (π) averages 0.0004 across all specimens with no significant differences between means for either regions or species ($P > 0.05$) (Table 2).

Geographically distant populations are not more differentiated from each other than populations in closer geographical proximity (Fig. 3). N_{ST} is significantly greater than G_{ST} (0.788 vs. 0.695, $P < 0.01$), indicating that there is a phylogeographical signal in the chloroplast data; in other words, haplotypes within populations are more similar to each other than expected. However, Mantel tests for IBD find no significant signal across the range of *Pyxidanthera* populations (Fig. 3, $R = 0.01$, $P = 0.39$). This pattern generally arises when genetic drift exerts more influence than gene flow at the regional scale (Hutchison & Templeton 1999). When northern and southern populations are analysed separately, there is no significant IBD in the northern populations ($R = -0.05$, $P = 0.47$) but there is marginally significant IBD in the southern populations, although the effect is weak ($R = 0.13$, $P = 0.049$).

Table 2 Genetic diversity indices for *Pyxidanthera barbulata* and *P. brevifolia* based on cpDNA sequences and AFLP markers

Population	Species	State	*N*	%P	DW	He	Π	Haplotypes
NC_1	*barbulata*	NC	14 (5)	41.0	12.99	0.15	0.0019	H1,H6,H11
NC_2	*barbulata*	NC	0 (5)	–	–	–	0.0000	H11
NC_3	*barbulata*	NC	0 (4)	–	–	–	0.0000	H2
NC_4	*barbulata*	NC	9 (4)	50.0	7.89	0.14	0.0000	H11
NC_6	*barbulata*	NC	19 (1)	26.5	7.40	0.08	0.0000	H7
NC_8	*barbulata*	NC	12 (4)	23.2	4.09	0.09	0.0005	H7,H11
NC_9	*barbulata*	NC	19 (6)	44.2	18.91	0.15	0.0013	H5,H10
NJ_CB	*barbulata*	NJ	0 (5)	–	–	–	0.0010	H7,H11,H12
NJ_CW	*barbulata*	NJ	17 (5)	34.2	14.41	0.11	0.0012	H4,H8,H9
NJ_WB	*barbulata*	NJ	15 (6)	27.1	9.48	0.10	0.0000	H11
NJ_WG	*barbulata*	NJ	18 (5)	38.7	18.69	0.12	0.0004	H4,H9
NY_1	*barbulata*	NY	0 (5)	–	–	–	0.0004	H7,H9
SC_1	*barbulata*	SC	8 (3)	50.3	12.67	0.16	0.0000	H3
VA_1	*barbulata*	VA	13 (5)	32.6	10.08	0.12	0.0019	H1,H6
SC_HP	*brevifolia*	SC	8 (5)	25.2	1.83	0.08	0.0004	H7,H8
SC_SL	*brevifolia*	SC	26 (5)	26.1	8.33	0.08	0.0000	H11
002A	*brevifolia*	NC	19 (5)	41.0	16.25	0.12	0.0000	H11
10	*brevifolia*	NC	17 (0)	47.4	23.21	0.15	–	–
20	*brevifolia*	NC	21 (5)	51.3	23.17	0.14	0.0008	H2,H11
24	*brevifolia*	NC	19 (0)	46.1	17.26	0.14	–	–
026D	*brevifolia*	NC	19 (1)	39.0	9.66	0.12	0.0000	H1
03_25	*brevifolia*	NC	30 (5)	36.1	15.18	0.11	0.0008	H2,H11
028E	*brevifolia*	NC	0 (4)	–	–	–	0.0000	H7
33	*brevifolia*	NC	33 (0)	31.3	22.40	0.11	–	–
038D	*brevifolia*	NC	15 (0)	41.9	13.65	0.14	–	–
057Y	*brevifolia*	NC	24 (3)	41.9	25.43	0.13	0.0000	H8
058B	*brevifolia*	NC	22 (0)	40.6	18.61	0.14	–	–
065N	*brevifolia*	NC	15 (5)	31.6	7.24	0.11	0.0000	H11
066A	*brevifolia*	NC	0 (4)	–	–	–	0.0000	H11
092B	*brevifolia*	NC	8 (0)	34.2	3.26	0.11	–	–
93_115	*brevifolia*	NC	17 (0)	43.2	20.41	0.14	–	–
Overall mean				37.8	13.70	0.12	0.0004	
barbulata				36.8	11.66	0.12	0.0006	
brevifolia				38.5	15.06	0.12	0.0002	
Northern				33.3	14.20	0.11	0.0006	
Southern				38.5	13.63	0.12	0.0004	

Amplified fragment length polymorphism (AFLP) genetic diversity indices were only calculated for populations with more than seven genotyped individuals (437 total specimens). %P represents the number of polymorphic loci, DW is a measure of rare alleles per population, and H*e* is a measure of expected heterozygosity based on the AFLP markers. π is a measure of cpDNA nucleotide diversity. *N* represents the number of specimens for each population for AFLP markers and cpDNA sequences (in parentheses).

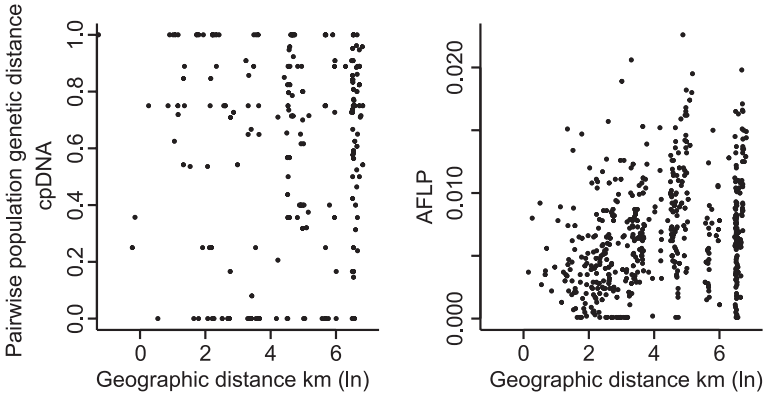

Fig. 3 Isolation by distance for cpDNA (left side) and amplified fragment length polymorphism (AFLP) (right side) markers for *Pyxidanthera barbulata* and *P. brevifolia* populations across the range of the genus *Pyxidanthera*. cpDNA data demonstrates no genetic isolation by geographical distance ($R = 0.01$, P-value = 0.39), while AFLP markers demonstrate weak but significant ($R = 0.27$, P-value = 0.02) isolation by distance at shorter distances with effects of genetic drift more evident at greater distances.

The highest posterior density for θ_{South} was higher than both $\theta_{Ancestral}$ and θ_{North}, although there is significant overlap between the 95% confidence intervals (Fig. S1, Supporting Information). The highest posterior density parameter estimate for migration from south to north is 2.09, while the estimate for migration from northern populations into southern populations is 0.01, indicating there has been gene flow between the two regions, with possibly greater migration from the southern populations into the northern. Past gene flow between the two populations is also supported by the model selection exercise; the worst-performing models constrained both migration parameters to 0 (Table S3). Time since divergence, t, was poorly estimated and failed to converge; this typically reflects a lack of a signal available in analyses that incorporate only a single locus with limited informative characters (J. Hey, personal communication).

AFLP

Three hundred and ten polymorphic bands were scored based on the three primer pairs. Each individual produced a unique AFLP profile, and the Euclidean error rate (based on 47 replicate pairs) was 4.2%, within the margin of acceptable error rates (Bonin *et al.* 2004). All genetic diversity indices were highly correlated, and sample size was not significantly correlated with any of the genetic diversity values (all $P > 0.05$). The population genetic diversity estimates for *P. barbulata* and *P. brevifolia* populations do not differ significantly for %P, DW, or H_e (all $P > 0.05$). In addition, there are no significant differences between regional genetic diversity estimates for percentage %P, H_e, or DW (all $P > 0.05$) (Table 1). The percentage of polymorphic loci (%P) ranges from 23.2% to 51.3%, with a mean of 37.8%, while Nei's population genetic diversity (H_e) ranges from 0.08 to 0.16 with a mean of 0.12 (Table 1).

Nonmetric multidimensional scaling ordination based on the population genetic distances (D) reveals no discrete grouping of populations based on either region or species (Fig. 4). Results from STRUCTURE also demonstrate little population genetic structure based on either geographical location or taxonomic identification (Fig. S2), and there was no graphical evidence for an optimal number of K distinct genetic groups (Fig. S3). The hierarchical AMOVAS grouped according to species (*P. barbulata* vs. *P. brevifolia*) and geographical region (North vs. South) found small but significant variation was explained by species (1.60%, $P < 0.01$) and region (3.20%, $P < 0.01$)(Table 1), while within-population variation remained high (90.53% and 89.01%, respectively, $P < 0.001$). There is evidence for a weak but significant effect of IBD in the AFLP data (Fig. 3; $R = 0.27$, $P = 0.02$). Genetic differentiation between populations increases with geographical distance, indicating low to moderate levels of short distance gene flow but little evidence of long-distance

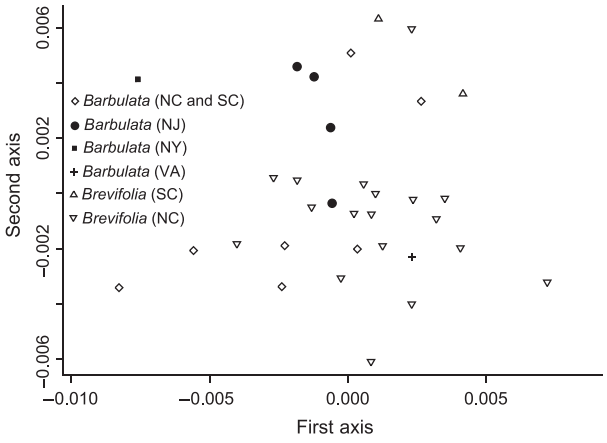

Fig. 4 Nonmetric multidimensional scaling ordination of *Pyxidanthera barbulata* and *P. brevifolia* population genetic distances (Nei's D) based on amplified fragment length polymorphism markers. In the legend, letters in parentheses represent US states. Little separation is evident among populations defined according to either taxonomic identity or geographical location.

gene flow; both gene flow and genetic drift influence the pattern depending on the geographical scale (Hutchison & Templeton 1999). At shorter distances, gene flow is dominant, increasing the correlation between genetic and geographical distances, while at greater distances genetic drift predominates.

Discussion

Taxonomy

There does not appear to be clear separation between *Pyxidanthera barbulata* and *P. brevifolia* based on either the morphological or genetic data (Figs 1 and 4). Although *P. brevifolia* in general has shorter and narrower leaves than the more widespread *P. barbulata*, there is significant overlap between the two species for both leaf length and width. Previously published work on the leaf morphology indicated that the differentiation between the two species was because of hydrological differences between the habitats that *P. barbulata* and *P. brevifolia* occupy (Primack & Wyatt 1975), with leaf length increasing continuously with increasing soil moisture. Although *P. brevifolia* individuals tend to be more pubescent than *P. barbulata* individuals (Fig. 1), there is significant variation in the pubescence of *P. brevifolia* both at the taxonomic level and within-populations (data not shown), with both glabrous and pubescent individuals represented in most populations. Interestingly, there are herbarium specimens of *P. barbulata* from xeric habitats of the Outer Coastal Plain of North Carolina that exhibit the shorter leaves of *P. brevifolia* specimens, but that are not pubescent; the extreme pubescence appears to be restricted to *P. brevifolia*.

Although several authors have suggested that *P. brevifolia* may represent a preadapted *P. barbulata* ecotype that moved into the Sandhills region from the Outer Coastal Plain (Wells & Shunk 1931; Primack & Wyatt 1975), the current study using cpDNA sequences and AFLP markers and a previous study using allozymes (Godt & Hamrick 1995) do not support this hypothesis. *P. barbulata* and *P. brevifolia* populations in the Sandhills are not genetically distinct from each other, with *P. barbulata* populations on Fort Bragg sharing cpDNA haplotypes with nearby *P. brevifolia* populations (Fig. 2 inset). In addition, there is no separation between *P. barbulata* and *P. brevifolia* populations in their AFLP profiles (Fig. 4). We cannot rule out the possibility that *P. brevifolia* is a recently derived ecotype of *P. barbulata*, restricted to the Sandhills, and that a few mutations have led to local adaptation, but this would need to have been recent enough that genetic differentiation is not apparent in AFLP profiles. Even though *P. brevifolia* appears to be an extreme morphological variant of *P. barbulata* associated with sandy, xeric sites, in our estimation, it warrants continued active management – specifically the regular prescribed fire schedule that Fort Bragg Military Reservation maintains – and further study because of its potentially critical role as an early season pollen and nectar provider and as a system for studying physiological adaptation to drought stress and phenotypic plasticity.

Phylogeography of the genus Pyxidanthera

Contrary to the well-documented trends of range contraction observed in many temperate plant species during the last glacial period in ENA, we found little evidence for either a southern refugium or range expansion following the LGM in the genus *Pyxidanthera*. Genetic diversity estimates for both the AFLP and cpDNA markers were not significantly different for northern and southern *P. barbulata* populations (Table 1), and northern populations contained several cpDNA haplotypes that did not occur in the southern populations (Fig. 2). More pointedly, estimates of the number of rare AFLP markers (DW), which may be more helpful in identifying refugial phylogeographical patterns (Paun *et al.* 2008), did not demonstrate significant differences between northern and southern populations. Finally, the two interior haplotypes – H7 and H11 – were widespread in both northern and southern populations with comparable frequencies (Fig. 2). These genetic patterns are contrary to what would be expected if there was a southern refugium for *Pyxidanthera* (Comps *et al.* 2001; Ikeda *et al.* 2008; Paun *et al.* 2008). Thus, it appears that the most likely scenario includes range stasis through the later Pleistocene. Furthermore, evidence of gene flow between geographically close populations suggests a possible explanation for low levels of genetic differentiation between northern and southern populations; these populations may not have been as geographically isolated in the recent past, with populations in the intervening area facilitating gene flow.

Several studies of tree species have also demonstrated the absence of typical refugial patterns (Palme *et al.* 2003; Maliouchenko *et al.* 2007), indicating that some species may have persisted closer to the ice sheet than previously thought. There is increasing evidence for 'cryptic refugia' in more northern latitudes for a number of mammal and plant species (Stewart & Lister 2001). Although mid-latitude refugia are possible, several alternatives have also been put forth. In the case of *Salix caprea*, which demonstrates little phylogeographical patterning, Palme *et al.* (2003) posit

high rates of dispersal, hybridization with other *Salix* species, and high mutation rates as possible reasons. These explanations are not very probable in the case of *Pyxidanthera*. *Pyxidanthera* seeds lack obvious morphological adaptations for dispersal, although ants have been observed transporting seeds (W. Wall, personal observation). Hybridization with other species is implausible, because *Pyxidanthera* is well differentiated from all other taxa within Diapensiaceae (Ronblom & Anderberg 2002). Although we have not estimated mutation rates, this alone would not generate the observed patterns.

That *P. barbulata* would persist, rather than retreat, during the climatic oscillations of the Pleistocene is consistent not only with the genetic data but also with our knowledge of Pleistocene habitats and the species' natural history. The GACP physiographical region, relative to more interior physiographical regions, may have been climatically buffered during the Pleistocene because of the moderating influence of the Atlantic Ocean (Rahmstorf 2002); moderation of climatic extremes could have allowed persistence closer to the ice sheet during glacial periods for some GACP species. Still, *P. barbulata* populations in New Jersey and New York would have experienced much colder conditions through much of the last glacial period (Jacobson *et al.* 1987; French *et al*. 2003, 2007). The vegetation community of the late Pleistocene in some of the areas of ENA does not have a modern analogue; most likely, it would have been a relatively open spruce (*Picea* spp.) forest with an herbaceous understory dominated by *Carex* spp. (Overpeck *et al*. 1992). The most important factors in determining the ecological niche of *P. barbulata* may be high light levels and an absence of competition, rather than temperature or moisture. The frequently burned habitats of the Sandhills of North and South Carolina and the Pine Barrens of New Jersey provide this habitat; it is conceivable that environments near the glacial boundary that lacked a dominant canopy cover during the last glacial period did as well. Finally, lower sea levels during glacial periods may have periodically increased available habitats for Atlantic Coastal Plain species such as *P. barbulata* on the exposed continental shelf (Hobbs 2004).

The present-day disjunction in the range of *P. barbulata* may be related to regional geomorphology. The Atlantic Coastal Plain is characterized by a series of alternating arches and embayments (Ward 1992); *Pyxidanthera* populations occur on the Cape Fear, Norfolk, and South New Jersey Arches, but are absent in the intervening Salisbury Embayment. The current disjunction in the range of the genus *Pyxidanthera* may be the result of oscillating sea levels that inundate embayment areas while arches remain above sea level (Bloom 1983; Sorrie & Weakley 2001). It is unlikely that long-distance gene flow between the northern and southern populations without intermediate populations would be high enough to prevent genetic differentiation. This suggests that the current vicariance between northern and southern populations may be recent and that during periods of relatively low sea levels, suitable habitat was exposed on the continental shelf, connecting northern and southern populations and allowing gene flow to minimize genetic differentiation.

The GACP floristic province contains the second-highest level of endemism in North America north of Mexico, yet the endemic plant species have been relatively understudied. Despite subtle topographic variation across the region, complex vegetation patterns exist and the biogeographical processes involved elude simple characterization. Although more phylogeographical studies of GACP endemic plant species are needed to determine whether the recent phylogeographical history of the genus *Pyxidanthera* is representative of multiple taxa or is simply an isolated case, it is apparent that the simple refugial model cannot account for the phylogeographical pattern in the genus *Pyxidanthera*. If similar phylogeographical patterns are found in similarly distributed GACP endemics, it would suggest a common mechanism was responsible and the remaining challenge would be to explain why only these taxa were thusly affected. Refugia are generally thought of as existing in the past; it could be the case that contemporary distributional patterns represent modern-day refugia for many Atlantic Coastal Plain endemic plant species.

Acknowledgments

We thank Fort Bragg Military Reservation and the Endangered Species Branch for logistic support and the Construction Engineering Research Laboratory (US Army Corps of Engineers Agreement #W9132T-07-2-0019) and the Native Plant Society for funding. We also thank the following individuals and laboratories for assistance: Consuelo Arellano, Ignazio Carbone, Matthew Cleary, Emile Devito, Janet Gray, Jacob Hilton, Xiang Liu, Bert Pittman, Bruce Sorrie, Andrew Walker, Alan Weakley, Steve Young and the Xiang and Hoffmann labs.

References

Ahles HE (1964) New combinations for some vascular plants of southeastern United States. *Journal of the Elisha Mitchell Scientific Society*, **80**, 172–173.

Bahlo M, Griffiths RC (2000) Inference from gene trees in a subdivided population. *Theoretical Population Biology*, **57**, 79–95.

Bloom AL (1983) Sea level and coastal changes. *Late-Quaternary Environments of the United States*, **2**, 42–51.

Bonin A, Bellemain E, Bronken Eidsen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.

Bonin A, Ehrich D, Manel S (2007) Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, **16**, 3737–3758.

Broyles SB (1998) Postglacial migration and the loss of allozyme variation in northern populations of Asclepias exaltata (Asclepiadaceae). *American Journal of Botany*, **85**, 1091–1097.

Buchanan M, Finnegan JT (2008) *Natural Heritage Program List of Rare Plant Species of North Carolina*, North Carolina Natural Heritage Program, Raleigh, NC.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, Verlag.

Carstens BC, Stoute HN, Reid NM (2009) An information-theoretical approach to phylogeography. *Molecular Ecology*, **18**, 4270–4282.

Church SA, Kraus JM, Mitchell JC, Church DR, Taylor DR (2003) Evidence for multiple Pleistocene refugia in the postglacial expansion of the eastern tiger salamander, Ambystoma tigrinum tigrinum. *Evolution*, **57**, 372–383.

Coart E, Lamote V, De Loose M, Van Bockstaele E, Lootens P, Roldán-Ruiz I (2002) AFLP markers demonstrate local genetic differentiation between two indigenous oak species [Quercus robur L. and Quercus petraea (Matt.) Liebl.] in Flemish populations. *Theoretical and Applied Genetics*, **105**, 431–439.

Comes HP, Kadereit JW (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.

Comps B, Gomory D, Letouzey J, Thiébaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, **157**, 389–397.

Cunningham CW, Collins TM (1994) Developing model systems for molecular biogeography: vicariance and interchange in marine invertebrates. In: *Molecular Ecology and Evolution: Approaches and Applications* (eds B Schierwater, B Streit, P Wagner, R DeSalle). pp. 405–433, Birkhauser Verlag, Basel, Switzerland.

Delcourt PA, Delcourt HR (1981) Vegetation maps for eastern North America: 40,000 yr BP to the present. In: *Geobotany II* (ed RC Romans). pp. 123–165, Plenum, New York.

Dorken ME, Barrett SCH (2004) Chloroplast haplotype variation among monoecious and dioecious populations of Sagittaria latifolia (Alismataceae) in eastern North America. *Molecular Ecology*, **13**, 2699–2707.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin*, **19**, 11–15.

Ehlers J, Gibbard PL (Eds.) (2004) *Quaternary Glaciations : Extent and Chronology*. Elsevier, Amsterdam.

Ehrich D (2006) AFLPdat: a collection of R functions for convenient handling of AFLP data. *Molecular Ecology Notes*, **6**, 603–604.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Evans MEK, Dolan RW, Menges ES, Gordan DR (2000) Genetic diversity and reproductive biology in Warea carteri (Brassicaceae), a narrowly endemic Florida scrub annual. *American Journal of Botany*, **87**, 372–381.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Fenneman NM (1938) *Physiography of Eastern United States*. McGraw-Hill, New York.

Flora of North America Editorial Committee (1993) *Flora of North America North of Mexico*. Oxford University Press, New York.

French HM, Demitroff M, Forman SL (2003) Evidence for late-pleistocene permafrost in the New Jersey Pine Barrens (Latitude 39 degrees N), Eastern USA. *Permafrost and Periglacial Processes*, **14**, 259–274.

French HM, Demitroff M, Forman SL, Newell WL (2007) A chronology of Late-Pleistocene permafrost events in southern New Jersey, eastern USA. *Permafrost and Periglacial Processes*, **18**, 49–59.

Frost C (2006) History and future of the longleaf pine ecosystem. In: *The Longleaf Pine Ecosystem: Ecology, Silvaculture, and Restoration* (eds S Jose, EJ Jokela, DL Miller). pp. 9–42, Springer Science, New York.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Godt MJW, Hamrick JL (1995) Low levels of allozyme differentiation between Pyxidanthera (pyxie-moss) taxa (Diapensiaceae). *Plant Systematics and Evolution*, **195**, 159–168.

Gonzales E, Hamrick JL, Chang S (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.

Hafner MS, Nadler SA (1990) Cospeciation in host-parasite assemblages: comparative analysis of rates of evolution and timing of cospeciation events. *Systematic Biology*, **39**, 192–204.

Herrera CM, Bazaga P (2008) Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. *Molecular Ecology*, **17**, 5378–5390.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, **104**, 2785–2790.

Hobbs III CH (2004) Geological history of Chesapeake Bay, USA. *Quaternary Science Reviews*, **23**, 641–661.

Holland BR, Clarke AC, Meudt HM (2008) Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Systematic Biology*, **57**, 347–366.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the

relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**, 1898–1914.

Ikeda H, Senni K, Fujii N, Setoguchi H (2008) Survival and genetic divergence of an arctic-alpine plant, Diapensia lapponica subsp obovata (Fr. Schm.) Hulten (Diapensiaceae), in the high mountains of central Japan during climatic oscillations. *Plant Systematics and Evolution*, **272**, 197–210.

Jacobson GL, Webb III T, Grimm EC (1987) Patterns and rates of vegetation change during the deglaciation of eastern North America. In: *North America and Adjacent Oceans during the Last Deglaciation. The Geology of North America* (eds WF Ruddiman, HE Wright Jr). pp. 277–288, Geological Society of America, Boulder, Colorado.

Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**, 893–903.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Magni CR, Ducousso A, Caron H, Petit RJ, Kremer A (2005) Chloroplast DNA variation of Quercus rubra L in North America and comparison with other Fagaceae. North America and comparison with other Fagaceae. *Molecular ecology*, **14**, 513–524.

Maliouchenko O, Palme AE, Buonamici A, Vendramin GG, Lascoux M (2007) Comparative phylogeography and population structure of European Betula species, with particular focus on B. pendula and B. pubescens. *Journal of Biogeography*, **34**, 1601–1610.

McKinnon GE, Vaillancourt RE, Steane DA, Potts BM (2008) An AFLP marker approach to lower-level systematics in Eucalyptus (Myrtaceae). *American Journal of Botany*, **95**, 368–380.

McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, **86**, 2088–2098.

Meudt HM, Bayly MJ (2008) Phylogeographic patterns in the Australasian genus Chionohebe (Veronica s.l., Plantaginaceae) based on AFLP and chloroplast DNA sequences. *Molecular Phylogenetics and Evolution*, **47**, 319–338.

Morris AB, Ickert-Bond SM, Brunson DB, Soltis DE, Soltis PS (2008) Phylogeographical structure and temporal complexity in American sweetgum (Liquidambar styraciflua; Altingiaceae). *Molecular Ecology*, **17**, 3889–3900.

Mylecraine KA, Kuser JE, Smouse PE, Zimmermann GL (2004) Geographic allozyme variation in Atlantic white-cedar, Chamaecyparis thyoides (Cupressaceae). *Canadian Journal of Forest Research*, **34**, 2443–2454.

Nei M (1972) Genetic distance between populations. *American Naturalist*, **106**, 283–292.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Oksanen J, Kindt R, Legendre P *et al.* (2009) *vegan: Community Ecology Package*.

Oliveira LO, Huck RB, Gitzendanner MA, Judd WS, Soltis DE, Soltis PS (2007) Molecular phylogeny, biogeography, and systematics of Dicerandra (Lamiaceae), a genus endemic to the southeastern United States. *American Journal of Botany*, **94**, 1017–1027.

Overpeck JT, Webb RS, Webb III T (1992) Mapping eastern North American vegetation change of the past 18 ka: No-analogs and the future. *Geology*, **20**, 1071.

Palme AE, Semerikov V, Lascoux M (2003) Absence of geographical structure of chloroplast DNA variation in sallow, Salix caprea L. *Heredity*, **91**, 465–474.

Paradis E (2009) pegas: Population and Evolutionary Genetics Analysis System. *R package version 0.3*.

Paun O, Schönswetter P, Winkler M, Intrabiodiv C, Tribsch A (2008) Historical divergence vs. contemporary gene flow: evolutionary history of the calcicole Ranunculus alpestris group (Ranunculaceae) in the European Alps and the Carpathians. *Molecular Ecology*, **17**, 4263–4275.

Perez-Collazos E, Sanchez-Gomez P, Jimenez JF, Catalan P (2009) The phylogeographical history of the Iberian steppe plant Ferula loscosii (Apiaceae): a test of the abundant-centre hypothesis. *Molecular Ecology*, **18**, 848–861.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237–1245.

Price EW, Carbone I (2005) SNAP: workbench management tool for evolutionary population genetic analysis. *Bioinformatics*, **21**, 402–404.

Primack RB, Wyatt R (1975) Variation and taxonomy of Pyxidanthera (Diapensiaceae). *Brittonia*, **27**, 115–118.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Development Core Team (2009) R: a Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 (available from http://www.R-project.org).

Rahmstorf S (2002) Ocean circulation and climate during the past 120,000 years. *Nature*, **419**, 207–214.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Reynolds JD (1966) *Morphological* Studies in Diapensiaceae. The Embryology of Pyxidanthera Mich, PhD thesis, University of South Carolina.

Roberts DW (2010) *labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.4-1*.

Ronblom K, Anderberg AA (2002) Phylogeny of Diapensiaceae based on molecular data and morphology. *Systematic Botany*, **27**, 383–395.

Schafale MP, Weakley AS (1990) *Classification of the Natural Communities of North Carolina: Third Approximation*. North Carolina Natural Heritage Program, Raleigh, NC.

Schönswetter P, Tribsch A (2005) Vicariance and dispersal in the alpine perennial Bupleurum stellatum L. (Apiaceae). *Taxon*, **54**, 725–732.

Shaw J, Lickey EB, Beck JT *et al.* (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, Lickey EB, Schill EE, Small RL (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.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Sorrie BA, Weakley AS (2001) Coastal plain vascular plant endemics: phytogeographic patterns. *Castanea*, **66**, 50–82.

Sorrie BA, Gray JB, Crutchfield PJ (2006) The vascular flora of the longleaf pine ecosystem of Fort Bragg and Weymouth Woods, North Carolina. *Castanea*, **71**, 129–161.

Sorrie BA, Weakley AS, Neesom GL (2009) *Pyxidanthera*. In: *Flora of North America North of Mexico Volume 3* (eds Flora of North America Editorial Committee). pp. 336–337, Oxford University Press, New York.

Stewart JR, Lister AM (2001) Crypitc refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, **16**, 608–613.

Takhtajan A (1986) *Floristic Regions of the World*. University of California Press, Berkeley, CA.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Trybush S, Hanley S, Cho K *et al.* (2006) Getting the most out of fluorescent amplified fragment length polymorphism. *Canadian Journal of Botany*, **84**, 1347–1354.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**, 139–151.

Ward LW (1992) Tertiary molluscan assemblages from the Salisbury embayment of Virginia. *Virginia Journal of Science*, **43**, 85–100.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256.

Weakley AS (2008) *Flora of the Carolinas, Virginia, Georgia, northern Florida, and surrounding areas* (*Working Draft of 7 April 2008*).

Wells BW (1929) A new pixie from North Carolina. *Journal of the Elisha Mitchell Scientific Society*, **44**, 238–239.

Wells BW, Shunk IV (1931) The vegetation and habitat factors of the coarser sands of the North Carolina coastal plain: an ecological study. *Ecological Monographs*, **1**, 465–520.

Zhang YY, Zhang DY, Barrett SCH (2010) Genetic uniformity characterizes the invasive spread of water hyacinth (*Eichhornia crassipes*), a clonal aquatic plant. *Molecular Ecology*, **19**, 1774–1786.

W.A.W. is a plant ecologist interested in the interplay between population genetics and population dynamics of rare and endemic species. N.A.D. conducts research on phylogenetics and phylogeography of arid adapted and edaphic endemic plants. Q.Y.X.'s research explores the mechanisms underlying biodiversity patterns and morphological variation, with emphases on the Cornales. W.A.H. is interested in ecology and conservation of savanna ecosystems. T.R.W. is a vegetation scientist with research interests in vegetation/environment relationships and biological diversity. M.G.H. uses genetic approaches to inform rare species conservation.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Parameter estimates for θ (southern, northern, and ancestral populations), time since divergence, and migration (gene flow) between northern and southern populations of the genus *Pyxidanthera* based on results from IMa2.

Fig. S2 Population genetic structure for the genus *Pyxidanthera* as inferred from the program STRUCTURE for K2 through 9.

Fig. S3 Log likelihood [Ln *P*(D)] and standard deviation results from program STRUCTURE for K1 through 9.

Table S1 Chloroplast haplotype accession numbers as archived in Genbank for the *atpI-atpH* intergenic spacer region (partial sequence) and the *psbD-trnT* intergenic spacer region (partial sequence)

Table S2 Polymorphisms of the 12 cpDNA haplotypes based on the cpDNA regions *atpI-atpH* and *psbD-trnT* in the genus *Pyxidanthera*

Table S3 Summary of model statistics for the 24 IMa2 models.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

A new tribal classification of Nyctaginaceae

Norman Douglas[1] & Richard Spellenberg[2]

1 *Department of Plant Biology, North Carolina State University, Campus Box 7612, Raleigh, North Carolina 27695-7612 U.S.A.*
2 *Department of Biology, MSC 3AF, New Mexico State University, Las Cruces, New Mexico 88003-8001, U.S.A.*
Author for correspondence: *Norman Douglas, norman_douglas@ncsu.edu*

Abstract Recent phylogenetic work shows that existing tribal concepts within Nyctaginaceae are incompatible with the principle of recognizing monophyletic taxa. We review the history of supergeneric classification in Nyctaginaceae, clarify issues pertaining to priority of certain generic names, and discuss the application of the conserved family name to the tribe Nyctagineae. *Pisoniella* and *Phaeoptilum* are moved from tribe Nyctagineae to Pisonieae and Bougainvilleeae, respectively, while tribe Abronieae, containing *Abronia* and *Tripterocalyx*, is submerged into Nyctagineae. Two distinctive genera, *Caribea* and *Colignonia*, are assigned to their own tribes, recognizing both their uniqueness and the uncertainty of their phylogenetic relationships. Finally, subtribes are not recognized in tribe Nyctagineae. Updated descriptions are provided for each tribe and one new tribe, Caribeaeae Douglas & Spellenberg, is recognized. Plant habit and general geographic distribution seem to be at least as pertinent as the often-homoplasious morphological details which had been emphasized in previous classifications.

Keywords conserved names; nomenclature; Nyctaginaceae; phylogenetics; supergeneric taxonomy

■ INTRODUCTION

Diversity, distribution, and morphology. — The family Nyctaginaceae comprises 300–400 species of trees, shrubs, and herbs classified in approximately 30 genera (Mabberley, 1987; Bittrich & Kühn, 1993). Species in Nyctaginaceae are found in all warmer areas of the world, but the bulk of the diversity at the generic and specific levels occurs in two regions: the Neotropics, and arid western North America. In the Neotropics, the majority of species are shrubs or small trees found in the three large genera *Neea*, *Guapira*, and *Pisonia*. In addition, the familiar genus *Bougainvillea*, known primarily for the horticulturally important *B. glabra*, *B. spectabilis* and hybrids, is endemic to South America, and is especially diverse in Bolivia. Species of *Colignonia* are restricted to the Andean region. Diversity at the generic level is greater in the arid regions of North America where nearly half of the recognized genera in the family are present, the most diverse being the herbaceous and suffrutescent *Abronia*, *Boerhavia*, *Mirabilis*, and *Acleisanthes*. *Commicarpus*, with roughly 30 species in Africa and western Asia, is also found in North and South America with five endemic species.

The number of species in genera of Nyctaginaceae follows a classic "hollow curve" pattern of diversity (Willis, 1922). The few large genera differ in geographic distribution and are morphologically variable. There are many genera of low diversity, 14 of those being monotypic. While in some cases monotypic genera, or genera with very few species, represent minor offshoots of dubious distinction (for instance, three genera for four species in tribe Boldoeae), it is also apparent that many small genera are well differentiated from other members of the family. For example, the monotypic genus *Phaeoptilum* is a unique xeromorphic shrub with winged fruits. This endemic of southwestern Africa is the only genus not occurring in the Americas. Likewise, the monotypic genus *Grajalesia*, a poorly known forest species from Central America, possesses winged fruits quite unlike those of other trees in the family. Its fruits are more similar in general appearance to those of *Phaeoptilum* and the xerophytic herbs found in North America, *Tripterocalyx* and some *Acleisanthes*. *Okenia*, with one or two species, has unique geocarpous fruits.

Nyctaginaceae have provided considerable fodder at the suprageneric and infrageneric levels for several generations of taxonomists. At the species level, perceived intergradation across morphospecies, in combination with a proliferation of names, has led to much taxonomic confusion, e.g., *Mirabilis*, (Reed, 1969; Spellenberg, 2003a), and spicate species of *Boerhavia* (Kearney & Peebles, 1964; Reed, 1970; Spellenberg, 2002). Some of the differences in species concepts result from actual incomplete differentiation of populations, perhaps in response to recently changing environments as in geologically diverse western North America (*Boerhavia*, *Abronia*, *Mirabilis*), and in response to changing taxonomic fashion with regard to species concepts (i.e., "lumping" and "splitting") over the past century (Weakley, 2005). In the arborescent tropical genera, comparatively poor preservation in herbarium specimens of many characters, combined with incomplete collections of these often-dioecious plants, has meant that species in the genera *Neea*, *Guapira*, and *Pisonia* remain poorly understood.

In contrast, differences in classification at the generic level and above more likely result from the paucity of characters provided by the Nyctaginaceae flower with its simple and often reduced perianth, and the simple, single-carpel, uni-ovulate ovary. Given the apparent ease with which the perianth (and thus the anthocarp) is modified in response to pollination and dispersal pressures, characters of these organs are perhaps less conserved than are characters of the gynoecium wall in other families. We draw this inference from the fact that the accessory fruit characters (e.g., wings, ribs, glands) previously used in classifications display substantial homoplasy (Douglas &

Manos, 2007). Thus, while ovary and fruit characteristics carry morphological information that is phylogenetically important in many families (e.g., Lawrence, 1951; Stebbins, 1974), artificial groupings have resulted when similar confidence has been extended to the characteristics of the anthocarps in Nyctaginaceae. Such artificial grouping was clearly demonstrated at the generic level, when three previously recognized genera, *Acleisanthes* s.str. (long perianth, vespertine, ribbed anthocarp), *Ammocodon* (= *Selinocarpus*) *chenopodoides* (short perianth, matutinal, winged anthocarp), and the remaining species of *Selinocarpus* (long perianth, vespertine, winged anthocarp) (see Fowler & Turner, 1977) each were found to be non-monophyletic separately, but instead to comprise multiple lineages embedded within a single clade (Levin, 2000). They are now all considered to belong within *Acleisanthes* (Levin, 2002). Similarly, changing concepts of the boundaries between groups at the generic level have resulted in very different taxonomies, e.g., *Mirabilis* (Standley, 1918; Spellenberg, 2003a), *Boerhavia* (Fosberg, 1978; Spellenberg, 2003b).

Supergeneric classifications. — The family was first recognized by Jussieu (1789). Early treatments of the family by Meisner (1841), Choisy (1849) and Bentham & Hooker (1880) established the outlines of a tribal classification by establishing Mirabileae Meisner, Pisonieae Meisner, Bougainvilleae Choisy, and Leucastereae Benth & Hook. Heimerl (1889) redrew the three tribes recognized by Bentham & Hooker. Standley (1918) recognized Heimerl's subtribes of tribe Mirabileae as tribes Abronieae, Bougainvilleae and Colignonieae. His publication of Abronieae and Colignonieae at tribal rank is valid by virtue of the inclusion of a dichotomous key to the tribes. Heimerl (1934) updated the tribal classification of Nyctaginaceae, recognizing five tribes. The tribe Pisonieae included genera that represented the majority of the tropical arborescent taxa, except the three genera in Leucastereae. The largest of Heimerl's tribes, Mirabileae, was further divided into four subtribes, one of which, Boerhaaviinae, contained most of the herbaceous and suffrutescent desert taxa. Bittrich & Kühn (1993) updated the treatment of the family and made several adjustments to Heimerl's 1934 treatment, including segregating from subtribe Boerhaaviinae a new subtribe, Nyctagininae, which represented genera with substantial bracts forming an involucre (Fig. 1).

The proper name for the tribe Boldoeae (Heimerl, 1889) has been confused in the literature for over 90 years. The tribe contains three genera that are similar to a degree that Standley included in *Salpianthus* Humb. & Bonpl. (1807) the genera *Boldoa* (commonly cited as Cav. ex Lag. 1816), and *Cryptocarpus* Kunth (1817) (Standley, 1918, 1931). When he combined these genera, he renamed the tribe Boldoeae as Salpiantheae. However, he apparently erred in determining the priority to be with *Salpianthus*. The original publication of the name *Boldoa* is as follows: Cavanilles, *Hortus regius matritensis* (1803: 8–9, tab. 7). This was later cited in Lagasca, *Nova genera et species* (1816: 10) (Heimerl, 1889). Both works include Latin descriptions of *Boldoa* and *Boldoa purpurascens* (likely explaining Standley's error) but the former also has an illustration of the species and clearly represents valid publication. Thus *Boldoa* has priority over *Salpianthus* if the four species in the group were to be placed into a single genus, as Standley preferred. In his last treatment of the family, Heimerl (1934) maintained the three genera; naturally, he maintained Boldoeae. However, many treatments (e.g., Spellenberg, 2001) have followed Standley in treating *Boldoa* as a synonym of *Salpianthus*.

Another issue that has never been satisfactorily addressed concerns the proper name for the tribe containing the type of the family name Nyctaginaceae. *Nyctago*, the generic name upon which the family name is based (Art. 10.6 of the *ICBN*, McNeill & al., 2006) is a superfluous and illegitimate name to be typified by the type of *Mirabilis*. Nowicke (1970) noted that Article 19 of the *Code* required that Heimerl's (1934) tribe Mirabileae, subtribe Boerhaaviinae be changed to tribe Nyctagineae, subtribe Nyctagininae, establishing these names by reference to Heimerl (1934). However, she referred to Mirabileae in two subsequent publications on pollen morphology (Nowicke, 1975; Nowicke & Skvarla, 1979). Bogle (1974) discussed a conflict that existed in the 1966 and 1972 *Codes* that pertained to families with conserved names, noting that both Heimerl's and Nowicke's names could be considered correct, depending on whether one emphasized the requirement that subfamilial taxa conform to the "correct" name of an included genus, or the requirement that the names of such groups be based on the same stem as the next higher taxon. Article 19.3 was revised in the Sydney *Code* (Voss & al., 1983) so that such a taxon was to be based on the "type of the adopted, legitimate name of the family to which it is assigned, but without citation of an author's name". This wording is maintained in the *Vienna Code* Article 19.4 except that the proscription of author citations in tribal and subtribal names disappeared with the Tokyo *Code* (Greuter & al., 1988). In this case, *Nyctago* nom. illeg. is the type genus. Article 19.5 further clarifies this issue by stating that subfamilial names based on illegitimate generic names are legitimate if they are also the base of a conserved family name, meaning that the proper name of the tribe that includes *Mirabilis* is Nyctagineae. Bittrich & Kühn (1993), like Nowicke (1970), recognized the bulk of the genera in Heimerl's (1934) tribe Mirabileae as tribe Nyctagineae, citing no author as per the final clause in Art. 19.3 then in effect. In fact, the first use of this tribal name was by Horaninov (1847: 105–107), whose tribe Nyctagineae was equivalent to the modern Nyctaginaceae and was one of four tribes in a broadly conceived "Allioniaceae", along with Plumbagineae, Staticeae, and Plantagineae. If a subtribe including *Mirabilis* were recognized (which we will not do) its proper name would be Nyctagininae Nowicke.

■ TAXONOMIC IMPLICATIONS OF MOLECULAR STUDIES

The first molecular phylogenetic study of Nyctaginaceae (Levin, 2000) explicitly focused on some members of tribe Nyctagineae. Due to sampling limitations and poor backbone resolution, the only result pertinent to higher-level classification was the suggestion that *Boerhavia* and *Allionia* were relatively closely related, calling into question the separation of subtribes Boerhaviinae and Nyctagininae. A recent phylogenetic study

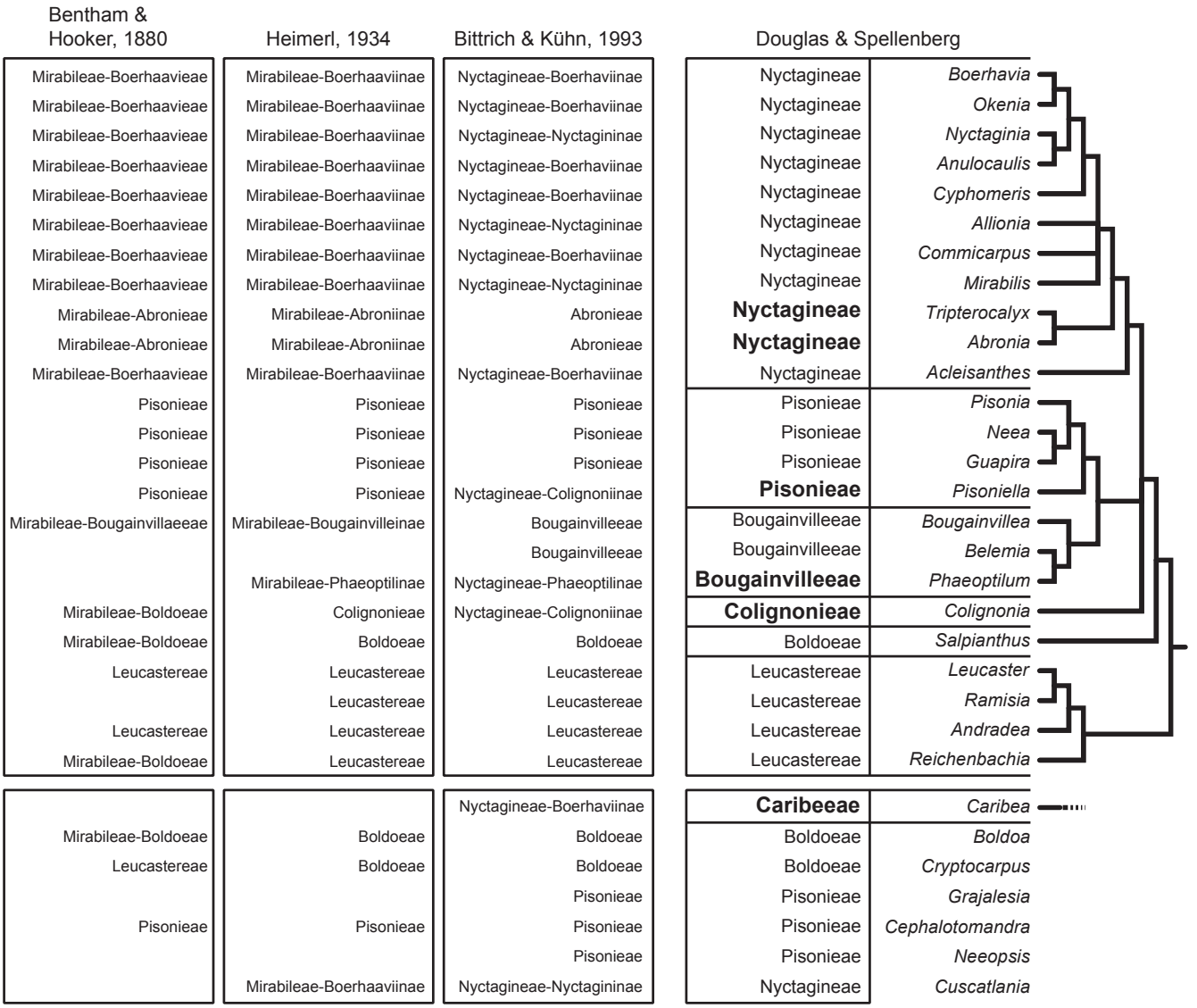

Fig. 1. Incongruence of historically influential classification schemes for Nyctaginaceae with respect to the phylogenetic hypothesis of Douglas & Manos (2007) and the new classification proposed here. All branches in the phylogenetic tree were supported at least >70% in parsimony or likelihood bootstrap or >95 bayesian posterior probability; otherwise they were collapsed. Hyphenated names correspond to tribe-subtribe; names in bold represent changes from Bittrich & Kühn (1993); genera unconnected to the tree were not included in the 2007 analysis.

of Nyctaginaceae that included nearly all currently accepted genera (Douglas & Manos, 2007) found that all previous classifications of Nyctaginaceae at the tribal level are incongruent with the evolutionary relationships among the genera demonstrated by molecular evidence (Fig. 1). With respect to the most recent (Bittrich & Kühn, 1993), the tribe Nyctagineae is not monophyletic due the basal position of the genus *Acleisanthes* (tribe Nyctagineae subtribe Boerhaviinae, now including *Selinocarpus* and *Ammocodon*; Levin, 2002) with respect to a clade that contains *Abronia* and *Tripterocalyx* (tribe Abronieae) and the remaining members of tribe Nyctagineae subtribes Nyctagininae and Boerhaviinae. The inclusion in tribe Nyctagineae of the distantly related *Phaeoptilum* and *Colignonia* (as the monogeneric subtribes Phaeoptilinae and Colignoniinae, respectively) is incompatible with a monophyletic tribe. The distinction between the subtribes Boerhaviinae and Nyctagininae is artificial due to the highly homoplasious distribution of involucral bracts in the North American xerophytic clade (Douglas & Manos, 2007). Finally, *Caribea* was included in subtribe Boerhaviinae, but a preliminary result based on a fragmentary *ndhF* sequence indicated that this poorly known taxon is probably more closely related to *Bougainvillea* or *Pisoniella* than to any members of the Nyctagineae. Now that the relationships among the genera of the family are better understood, a formal reclassification of the family into monophyletic tribes is warranted. We propose the following classification, which accomplishes the goal of recognizing monophyletic groups, is conservative with respect to previous classifications, and which accommodates remaining phylogenetic uncertainty with the recognition of two monogeneric tribes:

(1) The tribe Colignonieae Standl. (1918) is recognized, containing the genus *Colignonia*. The ambiguity in the exact placement of this distinctive genus in the molecular phylogeny of the family (Douglas & Manos, 2007) means that our recognition of the tribe Colignonieae is likely to remain stable even if phylogenetic resolution is improved. Furthermore, the genus *Pisoniella* will be recognized in Pisonieae, as it had been in every classification until Bittrich & Kühn (1993) moved it into tribe Nyctagineae, subtribe Colignoniinae. The molecular results place *Pisoniella* sister, with high support, to the other genera in Pisonieae.

(2) Similarly, the genus *Caribea* is removed from tribe Nyctagineae and placed in a distinct tribe, Caribeaeae. This distinctive Cuban endemic is known only from the type collection (*Alain & Lopez P. 7013*, Cuba, Oriente, 1959. Holotype at LS, isotype at NY!). The present phylogenetic uncertainty (Douglas & Manos, 2007) refutes the present classification without clearly suggesting a justifiable alternative placement. This, in combination with the unique morphology of the genus (Alain, 1960), leads us to conclude that tribal status will be stable in the face of new evidence if and when this genus is rediscovered and can be studied in more detail.

(3) The monotypic genus *Phaeoptilum* is transferred to the Bougainvilleeae. This transfer, in combination with proposed changes 1 and 2, will remove from tribe Nyctagineae those genera that are demonstrably not closely related to the clade, which comprises the bulk of tribe Nyctagineae sensu Bittrich & Kühn (1993). We note that Heimerl (1901) mentioned several similarities between *Phaeoptilum spinosum* and *Bougainvillea patagonica* (= *B. spinosa*), especially in a particular short hair type, and leaf form and arrangement, and for a time placed the two genera in the same subtribe (Heimerl, 1897). Like Leucastereae, this tribe is essentially native to the southern hemisphere.

(4) The tribe Abronieae Standl. (1918) is no longer recognized. The two genera within it, *Abronia* and *Triperocalyx,* are now in the tribe Nyctagineae. These two genera are clearly related based on morphological similarity and molecular evidence, and they are distinctive within the family based on characters of pollen morphology, stigma shape, inflorescence architecture, fruit morphology, and embryo shape. This concentration of unique apomorphic characters makes this clade a poor fit with any other tribe, which is why it has often been recognized as distinct. The molecular evidence clearly indicates that it is derived from within the North American xerophytic clade.

(5) We do not recognize any subtribes within tribe Nyctagineae. The close relationship of *Nyctaginia* to *Anulocaulis* (previously in separate subtribes) and low support values for the relationships between *Allionia, Commicarpus*, and *Mirablis*, preclude any justifiable grouping of these genera into subtribes. Nevertheless, the Nyctagineae now represents a coherent group of mostly herbaceous genera based the North American xerophytic clade.

Four additional recognized genera were not sampled for the phylogenetic study; without evidence to justify alternative assignments, we propose no changes in the tribal assignment of these genera. It should be noted that the winged fruits of *Grajalesia*, an arborescent species from Mexico and Central America, bear at least superficial similarity to those of *Phaeoptilum*, but at present we lack any convincing evidence to suggest its placement elsewhere, or its removal from, the Pisonieae. Based on general morphological similarity, the other unsampled genera, *Neeopsis*, *Cephalotomandra*, and *Cuscatlania*, seem likely to remain within the tribes where they are currently placed.

Descriptions of the tribes follow below. Characters of the tribes are gleaned from the literature, and where possible from specimens when such were available. Genera included, their size, and general geographic distribution are noted.

Tribe 1. Leucastereae Benth. & Hook. f., Gen. Pl. 3: 3. 1880 ('*Leucasterae*').

Trees or scandent shrubs, unarmed. Leaves alternate, petiolate, elliptic to lanceolate, with scurfy stellate hairs or scales; base symmetric to slightly asymmetric; margin entire. Inflorescences of axillary paniculate cymes or racemes; bracts minute, triangular, at base of pedicel or absent. Flowers perfect, rotate; perianth contracted in the middle or tubular throughout, tomentose or not, accrescent, 3–5-lobed, the lobes spreading or reflexed. Stamens 2 or 3 (12–20), connate at base, included. Pollen 3-colpate, 20–35 µm, exine reticulate or spinulose. Style linear or thickened, or essentially absent; stigma lateral, crested, or sulcate. Anthocarp with 12 ribs or anthocarp usually absent. Embryo hooked. *Andradea* (1 sp., E Brazil); *Ramisia* (1 sp., SE Brazil); *Reichenbachia* (2 spp., trop. South Amer.); *Leucaster* (1 sp., SE Brazil).

Tribe 2. Boldoeae Heimerl in Engler & Prantl, Nat. Pflanzenfam. 3(1b): 21, 31. 1889.

Plants suffrutescent or herbaceous, often subshrubs, unarmed. Roots unknown. Leaves alternate, petiolate, thickish, elliptic to rhomboid, tomentose or not, viscid, hairs straight or hooked; bases symmetric or nearly so; margin entire. Inflorescences congested axillary or terminal paniculate cymes of glomerules or racemes, bostryches (Bittrich & Kühn, 1993), bracts absent. Flowers perfect, rotate, 3–5-lobed, 2–5 mm, tubular to campanulate, not contracted above the ovary, pubescent. Stamens 3–5, exserted, filaments free. Pollen tricolpate, 20–46 µm, exine spinulose. Style short, linear-filiform, stigma delicate, fimbriate. Anthocarp globose or subglobose, coriaceous. Embryo curved. *Boldoa* (2–3 spp., Mexico); *Cryptocarpus* (1 sp., S Mexico, Centr. Amer., NW South Amer.); *Salpianthus* (1 sp., Mexico, Cent. Amer., N South Amer.).

Tribe 3. Colignonieae Standl. in Britton, N. Amer. Fl. 21: 195. 1918.

Lianas or scandent shrubs, unarmed. Roots tuberous or fibrous, also adventitious. Leaves opposite or whorled, short-petiolate, deltoid, ovate, or elliptic, glabrous, puberulous, or rufo-pilose, trichomes 3–4-celled; base truncate; margin entire. Inflorescence a cymose, umbel-like condensed dichasium; bracts white, green, or reddish, broadly lanceolate to obovate. Flowers perfect, rotate; perianth lobes 3(–4)–5, campanulate or

spreading, connate only at base. Stamens 5, exserted, episepalous, basally connate; filaments flattened, nectariferous; anthers subglobose. Pollen 12-pantoporate, 17–30 µm, exine spinulose. Ovary stalked, style clavate, stigma penicillate. Anthocarp present, winged or angled. Embryo curved. *Colignonia* (6 spp., Andean South Amer.).

Tribe 4. Bougainvilleeae Choisy in Candolle, Prodr. 13: 427, 436. 1849. ('*Bougainvilleae*')

Trees, or shrubs, sometimes scandent, perennial, occasionally with spines. Leaves alternate, opposite, or fasciculate on short branches, petiolate or nearly sessile, ovate, orbicular, to linear-lanceolate, succulent to membranous, glabrous or pubescent; base symmetric or nearly so, margin entire. Inflorescence cymose or racemose or flowers solitary, bracts absent or 3, often showy, the flowers often borne on costae of bracts. Flowers perfect or imperfect (plants then possibly polygamous, Stannard, 1988), rotate, tubular or salverform. Stamens 5–12, often connate at base, included or didymous with longer stamens exserted. Pollen tricolpate with reticulate exine or pantocolpate. Style short, filiform or stout, stigma linear to penicilliate or multifid. Anthocarp fusiform and 5-ribbed or with 4 translucent, scarious wings. Embryo curved. *Belemia* (1 sp., E Brazil); *Bougainvillea* (14–18 spp., Centr. & trop. South Amer.); *Phaeoptilum* (1 sp., SW Africa).

Tribe 5. Pisonieae Meisn., Pl. Vasc. Gen. 10: Tab. Diagn. 318, Comm. 230. 1841.

Trees, shrubs, or scandent shrubs, unarmed or with paired axillary spines. Leaves alternate, opposite, whorled, or verticillate, sometimes anisophyllous, petiolate, lanceolate, elliptic to (ob)ovate, membranous to coriaceous or fleshy, glabrous to glandular puberulent; base symmetric; margins entire. Inflorescences axillary and terminal paniculate cymes, corymbs, or glomerules; bracts 2–3 beneath each flower, persistent or cauducous. Flowers perfect or imperfect, or polygamous, campanulate, urceolate, rotate-salverform, or tubular, the limb 5-lobed. Stamens (2–)5–10(–many), exserted or included, connate at base, adnate to base of pistil in perfect flowers, filaments unequal, Pollen generally tricolpate-spinulose. Style exserted, stigma penicillate or pappilose. Anthocarps oblong, clavate, or ellipsoid, 5-ribbed, coriaceous and glandular-sticky, or globose, fleshy and glabrous. Embryo straight. *Cephalotomandra* (1–3 spp., Colombia); (*Grajalesia* (1 sp., Mexico); *Guapira* (ca. 70 spp., trop. Amer.); *Neea* (ca. 80 spp., trop. Amer.); *Neeopsis* (1 sp., Guatemala); *Pisonia* (ca. 40 spp., pantropical, but especially diverse in the Americas and SE Asia); *Pisoniella* (1 sp., Mexico, Centr. & S South Amer.).

Tribe 6. Nyctagineae Horan., Char. Ess. Fam.: 106. 1847.

Woody or suffrutescent subshrubs, or annual or perennial herbs, sometimes scandent, unarmed, in some with bands of viscid exudate on internodes. Roots slender and fleshy to tuberous or spongy-woody taproots, rarely rhizomatous. Leaves opposite, frequently anisophyllous, sessile or petiolate, membranous to fleshy, linear, cordate, ovate, or orbicular, glabrous or pubescent, often glandular; base symmetric to asymmetric; margins entire, crenate, undulate or sinuate, glandular pubescent to glabrous. Inflorescences terminal or axillary spikes, cymes, umbels or flowers solitary. Involucres of 3–20 connate or free bracts, or 1–2 often minute, persistent or cauducous bracts subtending individual flowers or terminal cymes; bracts oblong, linear, acuminate, or lanceolate, green or scarious. Flowers actinomorphic or slightly to strongly zygomorphic, campanulate to tubular to salverform, constricted above the ovary (4–)5-lobed (in some cleistogamous flowers often present, these quite reduced). Stamens (1–)2–5(–18), included or exserted, united at base and sometimes inserted on perianth tube. Pollen 100–200 µm, pantoporate with spinulose exine, or tricolpate with reticulate exine. Style included or exserted, filiform, stigma linear, capitate, or peltate. Anthocarp globose, turbinate, clavate, obpyramidal to fusiform, radially symmetric (gibbous in *Cyphomeris*, cymbiform with 2 rows of teeth in *Allionia*), coriaceous (spongy and geocarpous in *Okenia*), 3–10 costate, often with membranaceous wings or viscid glands; sulci smooth, pappilose, or rugose. Embryo hooked. *Abronia* (ca. 20 spp., W and C North Amer.); *Acleisanthes* (17 spp., SW and SC North Amer., with 1 sp. in NE Africa); *Allionia* (1–2 spp., Americas); *Anulocaulis* (5 spp., SC and SW North Amer.); *Boerhavia* (ca. 40 spp., worldwide in tropical and subtropical regions, annuals especially diverse in SW North Amer.); *Commicarpus* (30–35 spp., nearly worldwide in tropical and subtropical regions, especially diverse in Africa and W Asia); *Cyphomeris* (2 spp., SC North Amer.); *Mirabilis* (ca. 60 spp., Americas and 1 sp. in SC Asia); *Nyctaginia* (1 sp., SC North Amer.); *Okenia* (1–2 spp., Mexico, Centr. America); *Tripterocalyx* (4 spp., SW North Amer.).

Tribe 7. Caribeeae Douglas & Spellenb., **tr. nov.** – Type: *Caribea* Alain in Candollea 17: 113. 1960.

Fruticulus perennis valde abbreviatus, depressus; caules numerosi, in base lignescentes, dense foliosi, striate, glandulosi. Folia opposita. Flores 3–5-bracteolati, solitarrii, terminales. Perianthium infundibulare, tubo supra ovarium constricto, limbo 5-lobato. Stamina 2, filamentis capillaribus ad basim perianthii adnati. Ovarium globoso-oblongum, stylus filiformis, exsertus, stigmate capitato. Anthocarpium globoso-oblongum, laeve (adapted from Alain, 1960).

Tufted, compact mat-forming, taprooted perennial. Leaves petiolate, opposite, forming a stipuliform sheath at base, oblanceolate, fleshy or succulent, glabrous, margin entire. Inflorescence terminal, flowers solitary, subtended by an involucre of 3–5 free narrow bracts. Perianth constricted above the ovary, distil portion nearly urceolate, with 5 shallow lobes. Stamens 2, exserted, filiments adenate to perianth base. Style filiform, exserted, stigma capitate. Anthocarp subglobose, smooth. Embryo unknown. *Caribea* (1 sp., Cuba).

■ ACKNOWLEDGEMENTS

The authors wish to thank Alexander Krings and John Strother for helpful discussion during the preparation of this manuscript.

■ LITERATURE CITED

Alain. 1960. Novedades en la flora Cubana, XIII. *Candollea* 17: 113–121.

Bentham, G. & Hooker, J. 1880. *Genera plantarum ad exemplaria imprimis in herbariis kewensibus servata definita*. London: Reeve.

Bittrich, V. & Kühn, U. 1993. Nyctaginaceae. Pp. 473–486 in: Kubitzki, K., Rohwer, J.G. & Bittrich, V. (eds.), *The families and genera of flowering plants*, vol. 2. Berlin: Springer.

Bogle, A.L. 1974. The genera of Nyctaginaceae in the southeastern United States. *J. Arnold Arbor.* 55: 1–37.

Cavanilles Palop, A.J. 1803. *Hortus regius matritensis*. Madrid: Real Jardín Botánico.

Choisy, J.D. 1849. Nyctaginaceae. Pp. 425–458 in: de Candolle, A., (ed.), *Prodromus systematis naturalis regni vegetabilis*, vol. 13. Paris: Masson.

Douglas, N.A. & Manos, P.S. 2007. Molecular phylogeny of Nyctaginaceae: Taxonomy, biogeography, and characters associated with a radiation of xerophytic genera in North America. *Amer. J. Bot.* 94: 856–872.

Fosberg, F.R. 1978. Studies in the genus *Boerhavia* Nyctaginaceae, parts 1–5. *Smithsonian Contr. Bot.* 39: 1–20.

Fowler, B.A. & Turner, B.L. 1977. Taxonomy of *Selinocarpus* and *Ammocodon* (Nyctaginaceae). *Phytologia* 37: 177–208.

Greuter, W., Burdet, H.M., Chaloner, W.G., Demoulin, V., Nicolson, D.H., Silva, P.C. (eds.). 1988. *International code of botanical nomenclature: Adopted by the Fourteenth International Botanical Congress, Berlin July-August 1987.* Regnum Vegetabile 118. Königstein: Koeltz Scientific Books.

Heimerl, A. 1889. Nyctaginaceae. Pp. 14–32 in: Engler, A. & Prantl, K. (eds.), *Die natürlichen Pflanzenfamilien*, vol. 1b. Leipzig: Engelmann.

Heimerl, A. 1897. Beiträge zur Systematik der Nyctaginaceen. *XXIII. Jahresber. K.K. Staats-Oberrealschule im XV. Bezirke Wien (Fünfhaus)* 13: 23.

Heimerl, A. 1901. Monographie der Nyctaginaceen. I. *Denkschr. Kaiserl. Akad. Wiss., Wien, Math.-Naturwiss. Kl.* 70: 97–141.

Heimerl, A. 1934. Nyctaginaceae. Pp. 86–134 in: Engler, A. & Prantl, K. (eds.), *Die natürlichen Pflanzenfamilien*, 2nd ed., vol. 16c. Leipzig: Engelmann.

Horaninov, P.F. 1847. *Characteres essentiales familiarum ac tribuum regni vegetabilis et amphorganici*. Petropoli: Wienhöberianis.

Jussieu, A.L. 1789. *Genera plantarum*. Paris: Herrisant.

Kearney, T.H. & Peebles, R.H. 1964. Nyctaginaceae. Pp. 270–279 in: Kearney, T.H. & Peebles, R.H. (eds.), *Arizona flora*, 2nd ed. Berkeley: Univ. of California Press.

Lagasca y Segura, M. 1816. *Genera et species plantarum*. Madrid: Matriti ex Typographia Regia.

Lawrence, G.H.M. 1951. *Taxonomy of vascular plants*. New York: MacMillan.

Levin, R.A. 2000. Phylogenetic relationships within Nyctaginaceae tribe Nyctagineae: Evidence from nuclear and chloroplast genomes. *Syst. Bot.* 25: 738–750.

Levin, R.A. 2002. Taxonomic status of *Acleisanthes*, *Selinocarpus*, and *Ammocodon* (Nyctaginaceae). *Novon* 12: 58–63.

Mabberley, D.J. 1987. *The plant book*. Cambridge: Cambridge Univ. Press.

McNeill, J., Barrie, F.R., Burdet, H.M., Demoulin, V., Hawksworth, D.L., Marhold, K., Nicolson, D.H., Prado, J., Silva, P.C., Skog, J.E., Wiersema, J.H. & Turland, N.J. (eds.). 2006. *International code of botanical nomenclature (Vienna Code): Adopted by the Seventeenth International Botanical Congress Vienna, Austria, July 2005*. Regnum Vegetabile 146. Ruggell: Gantner.

Meisner, K.F. (ed.). 1841. *Plantarum vascularium genera secundum ordines: Tabulis diagnosticis expositae*. Leipzig: Libraria Weidmannia.

Nowicke, J.W. 1970. Pollen morphology in the Nyctaginaceae, part 1: Nyctagineae, Mirabileae. *Grana* 10: 79–88.

Nowicke, J.W. 1975. Pollen morphology in the order Centrospermae. *Grana* 15: 51–78.

Nowicke, J.W. & Skvarla, J.J. 1979. Pollen morphology: The potential influence in higher order systematics. *Ann. Missouri Bot. Gard.* 66: 633–700.

Reed, C.F. 1969. Nyctaginaceae. Pp. 151–220 in: Lundell, C.L. (ed.), *Flora of Texas*, vol. 2. Renner, Texas: Texas Research Foundation.

Reed, C.F. 1970. Nyctaginaceae. Pp. 574–599 in: Correll, D.S. & Johnston, M.C. (eds.), *Manual of the vascular plants of Texas*, vol. 6. Renner, Texas: Texas Research Foundation.

Spellenberg, R. 2001. Nyctaginaceae. Fasc. 93 of: Rzedowski Rotter, J. & de Rzedowski, G.C. (eds.), *Flora del Bajío y de regiones adyacentes*. Mexico, Patzcuaro: Instituto de Ecología.

Spellenberg, R. 2002. *Boerhavia coulteri* var. *palmeri*, a new varietal combination for *Boerhavia* (Nyctaginaceae) of southwestern North America. *SIDA* 20: 151–155.

Spellenberg, R. 2003a. *Mirabilis*. Pp. 40–56 in: Flora of North America Editorial Committee (ed.), *Flora of North America north of Mexico*, vol. 4. New York: Oxford Univ. Press.

Spellenberg, R. 2003b. Nyctaginaceae. Pp. 14–74 in: Flora of North America Editorial Committee (ed.), *Flora of North America north of Mexico*, vol. 4. New York: Oxford Univ. Press.

Standley, P.C. 1918. Allioniaceae. Pp. 171–254 in: Britton, N.L. (ed.), *North American flora*, vol. 21. New York: New York Botanical Garden.

Standley, P.C. 1931. The Nyctaginaceae of northwestern South America. *Publ. Field Mus. Nat. Hist., Bot. Ser.* 11: 71–126.

Stannard, B.L. 1988. Nyctaginaceae. Pp. 16–37 in: Launert, E. (ed.), *Flora zambesiaca*, vol. 9, part 1. Kew: Royal Botanic Gardens.

Stebbins, G.L. 1974. Evolutionary processes and the origin of higher categories. Pp. 12–28 in: *Flowering plants: Evolution above the species level*. Cambridge: Harvard Univ. Press.

Voss, E.G., Burdet, H.M., Chaloner, W.G., Demoulin, V., Hiepko, P., McNeill, J., Meikle, R.D., Nicolson, D.H., Rollins, R.C., Silva, P.C. & Greuter, W. (eds.). 1983. *International Code of Botanical Nomenclature: Adopted by the Thirteenth International Botanical Congress, Sydney, August 1981.* Regnum Vegetabile 111. Utrecht: Bohn, Scheltema & Holkema.

Weakley, A.S. 2005. *Change over time in our understanding of the flora of the southeastern United States: Implications for plant systematics, bioinformatics, and conservation*. Dissertation, Duke University, Durham, North Carolina.

Willis, J.C. 1922. *Age and area: A study in geographical distribution and origin of species*. Cambridge: Cambridge Univ. Press.

TRIPTEROCALYX CARNEUS (NYCTAGINACEAE) IS SELF-COMPATIBLE

NORMAN A. DOUGLAS

Department of Biology, Duke University, Durham, NC 27705
Present Address: Department of Plant Biology, North Carolina State University,
Campus Box 7612, Raleigh, NC 27695-7612
Correspondent: norman_douglas@ncsu.edu

ABSTRACT—Recent phylogenetic work has demonstrated that the ability of species of the angiosperm family Nyctaginaceae to self-fertilize is evolutionarily labile. However, the potential for further investigation of the evolution of mating systems in the family is limited, because there is no information on reproductive biology for several genera. I performed an experiment on a natural population of *Tripterocalyx carneus* to determine whether this species is self-compatible. Individual flowers were emasculated, bagged to exclude pollinators, or both. Flowers that were bagged but not emasculated set viable fruit in the majority of cases, which demonstrates that this species is self-compatible.

RESUMEN—Un estudio filogenético reciente ha demostrado que la capacidad de autofecundación de las especies de la familia Nyctaginaceae de las angiospermas es evolutivamente cambiable. Sin embargo, el potencial de investigaciones extensas sobre la evolución de sistemas de reproducción en la familia es limitado, porque no hay información sobre la biología reproductiva de varios géneros. Realicé un experimento sobre una población natural de *Tripterocalyx carneus* para determinarse si esta especie es autocompatible. A flores individuales se les removieron los estambres o fueron embolsadas para excluir a los polinizadores, o ambos. El grupo de flores que fue embolsado, pero al que no se le removieron los estambres, produjo frutos viables en la mayoría de los casos, lo cual demuestra que esta especie es autocompatible.

The four o'clock family (Nyctaginaceae) contains both self-compatible and self-incompatible species. As part of a recent phylogenetic study of the family (Douglas and Manos, 2007), we compiled and reviewed all known reports of mating systems of Nyctaginaceae. We found that based on current knowledge, at least three evolutionary gains, or alternatively, six losses of self-incompatibility are required to explain the current distribution of known mating systems in Nyctaginaceae. This is a minimum estimate of the number of transitions in mating systems in the family: we based our inferences on experimental reports for some species, anecdotal reports for others, and certain assumptions (i.e., that species with stylar movements that ensure self-pollination, or that possess cleistogamous flowers, are in fact self-compatible; and that dioecious species are self-incompatible). Data were not available for all genera, nor do all genera have morphological or ''behavioral'' traits (e.g., Spellenberg, 2000) that allow inferences about reproductive systems to be made. Some genera contain both self-compatible and self-incompatible species. For instance, *Mirabilis* is largely self-compatible, but one section, *Quamoclidion*, is comprised of self-incompatible species (Cruden, 1973; Pilz, 1978; Hernández, 1990; Hodges, 1995; Leal et al., 2001)

Of particular interest with regard to evolution of mating system is the tribe Abronieae. Although some authors have considered Abronieae to be monogeneric (Bittrich and Kühn, 1993), most authors have maintained two genera, *Tripterocalyx* and *Abronia* (Galloway, 1975; Spellenberg, 2003). As in *Mirabilis*, both self-incompatible and self-compatible species are known in *Abronia*. Experimental evidence has shown that *Abronia macrocarpa* most likely has a sporophytic self-incompatibility system (Williamson et al., 1994; Williamson, and Bazeer, 1997). The three maritime species, *Abronia latifolia*, *A. maritima*, and *A. umbellata* failed to set seed when experimentally self-pollinated (Tillett, 1967). In a different study, *Abronia umbellata breviflora*, however, was observed to self-pollinate and set viable seed in a greenhouse (McGlaughlin et al., 2002). Finally, Saunders and Sipes (2006) reported a mixed mating system for the narrowly endemic *Abronia ammophila*. Although the mat-

ing system has been experimentally examined in only six species or subspecies of *Abronia*, it is worthwhile to have information on mating systems from the probable sister genus, *Tripterocalyx*, to aid in understanding the evolution of mating systems in *Abronia* specifically, and more generally, to increase confidence and resolution in phylogenetic reconstructions of the evolution of mating systems in Nyctaginaceae.

I examined the ability of *Tripterocalyx carneus* to self-fertilize. Like *Abronia*, *Tripterocalyx* produces umbellate inflorescences of salverform flowers with a showy limb and a narrow tube, which indicates that lepidopterans are likely pollinators. Anthesis occurs in the early evening. The highly fragrant flowers are open through the night and wilt the following morning. Like nearly all other Nyctaginaceae, the fruit is an achene, surrounded by an ''anthocarp,'' a persistent accessory fruit that develops from the lower part of the uniseriate perianth. In *Tripterocalyx*, the anthocarp enlarges greatly, hardens, and develops three or four large, membranaceous wings.

The study was conducted at a site on the U-Bar Ranch, Grant Co., New Mexico. A large population of >40 individual plants was located in sandy loam along the Gila River. The plants were arranged around a circular clearing (result of activity by harvester ants). All plants in the population were well established and each had several umbels producing flowers. I chose 12 plants adjacent to the clearing, and on a single unopened flower in each of four separate inflorescences on each plant, I performed one of the following manipulations: Treatment 1) The flower was slit down the side of the perianth with a razor blade, and with fine forceps, the filaments and anthers were removed. Afterwards, a fine-mesh drawstring bag was placed over the emasculated flower to exclude pollinators. I anticipated that no seed would mature with this treatment. Treatment 2) The flower was emasculated as above, but not bagged, so that seeds that matured would be the result of either geitonogamy (in which an ovule is fertilized by pollen from a different flower on the same plant) or xenogamy (in which the pollen is transferred from a different individual), but not intrafloral self-pollination. Treatment 3) Bags were placed over non-emasculated flowers. Any seeds that resulted from this treatment would be the result of intrafloral self-pollination and would confirm self-compatibility. Treatment 4) Control; flowers were neither emasculated nor bagged, to provide an estimate of the proportion of flowers that would form a viable seed under field conditions.

While hand-pollinations would have been the most straightforward way of testing self-compatibility, it proved to be too difficult to make these manipulations accurately under field conditions. Subsequent to each manipulation, I tied a thread (colored according to treatment) loosely around the lower part of the perianth, which would eventually develop into the accessory fruit. This thread served to identify the fruit resulting from the manipulated flower while allowing me to avoid modifying the surrounding flowers in the umbel. All manipulations took place on the morning of 11 July 2001. I visited the population that evening to verify that the manipulated flowers opened normally, and to observe activity of pollinator in the population. The population was revisited during the morning and evening on 12 and 13 July, at which point the flowers had completely senesced. I then removed the bags. On 5 August, I returned to the population, at which point fruits were fully mature on the manipulated umbels, but had not dispersed. I relocated and collected fruits using the colored threads. In those cases where a fruit failed to develop, the small remnant of the perianth was still surrounded by the thread and all fruits or undeveloped remnants of flowers were recovered while still attached to the maternal plant. Thus, I was able to know the outcome of each manipulation of a flower. I opened all fruits in the lab to ascertain presence or absence of a mature achene. The proportion of anthocarps developing a mature achene were determined for each treatment and compared using a generalized linear model with logit link function (McCulloch and Searle, 2001) to determine whether there was a significant difference in proportion of mature achenes developed under each treatment.

The first evening after flowers were treated, all flowers opened normally. Because, in altered flowers, the slits were longitudinal along the tube of the flower, leaving the limb intact, emasculated flowers were not greatly distorted from the perspective of floral visitors. Presence of bags did not interfere with normal opening of the flower. In *Tripterocalyx*, the stamens and style are included in the floral tube; thus, the bag never contacted either organ. Numerous hawkmoths

TABLE 1—Seeds matured according to pollination treatment, and additional empty fruits produced. Proportions of matured seeds followed by same letter do not differ significantly ($P > 0.05$).

Pollination treatment	*n*	Mature seeds	Additional empty fruits	Proportion of mature seeds
Emasculated, bagged	12	1	2	0.083 a
Emasculated, not bagged	12	10	0	0.833 b
Intact, bagged	12	8	0	0.667 b
Control	12	11	1	0.917 b

(*Hyles lineata*, Sphingidae) visited flowers in the population. Moths visited both emasculated and unmodified flowers (and in several instances, probed the mesh bags). The following morning, all flowers that had opened the previous night were wilted and drooping, but the color had not faded. The second evening, many weakly reopened, including some that had been emasculated. The following evening, no treated flower opened, and the petaloid distal portion of the perianth had fallen from some flowers and was faded in the rest.

When fruits were collected, it was apparent that many had failed to mature. Whereas matured anthocarps were large (20–30 mm in length, with wings ≤12 mm broad), and usually filled with a large (7–10 mm) achene, the undeveloped anthocarps were small and shriveled (7 mm long, with small wings <5 mm broad). These anthocarps were not filled with an achene, and essentially represented remnants of the lower perianth that ceased development after the flowering stage due to the ovule not being fertilized. Tillett (1967) and Galloway (1975) reported that some *Abronia* would mature anthocarps for all flowers on an umbel if even one flower on an umbel were fertilized, although most anthocarps would be empty. In this experiment, a similar phenomenon might have been at work; I found three empty, but mature, anthocarps, although in most cases where an achene failed to mature, the anthocarp was aborted also.

Table 1 shows number of mature achenes produced (and additional empty anthocarps, if any) with respect to treatment applied. The procedure for emasculating and bagging flowers was not 100% effective. In Treatment 1, one flower was still able to produce a normal seed, although this represented a significantly smaller proportion ($P < 0.05$) than in other experimental treatments. Barring apomixis, the only explanation is that I accidentally disrupted an anther sac in the process of removing a stamen from the flower. Nyctaginaceae produce single-seeded fruits, and a single stray pollen grain could effectively pollinate the flower.

Despite limitations of technique, Treatment 2 showed that *T. carneus* is an effective outcrosser, because most emasculated flowers that were open-pollinated set normal fruits. This is not surprising, considering the vigorous attention paid to plants by hawkmoths during the experiment.

Treatment 3 demonstrated convincingly that the species is self-compatible. Intact bagged flowers set fruits in 8 of 12 cases. Mesh on bags I employed was much too fine to be penetrated by the proboscis of *Hyles lineata*. Drawstrings on the mesh bags enabled me to exclude crawling insects, and no insect was found inside bags when they were removed. Therefore, self-pollination is the only explanation for the high number of seeds set. It also is consistent with the unexpected results from Treatment 1; if this species was self-incompatible, I would have had to transfer outcrossed pollen to a stigma enclosed in an unopened bud, rather than simply leaving a grain of self pollen behind.

The proportion of matured achenes in Treatments 2 and 3 did not differ significantly ($P > 0.05$) from each other or from the proportion matured in Treatment 4, the open-pollinated control group. All but one of the unmodified flowers in the control group produced a mature anthocarp filled with a normal achene, indicating that with a combination of self-pollination and outcrossing, *T. carneus* is able to achieve high rates of successful pollination.

These results confirm the ability of *T. carneus* to self-fertilize and suggest that the species possesses a mixed mating system. It is not yet possible to infer the ancestral state for the tribe Abronieae based on current knowledge phylog-

eny and reproductive compatibility. However, as future studies address the phylogeny and reproductive biology of Abronieae, it will be of interest to see whether the apparent lability in mating system is evidenced at fine taxonomic levels, and to investigate correlations with life history, habitat, and pollination ecology.

I thank S. Stoleson (United States Department of Agriculture Forest Service) for facilitating access to the study population, and P. Hernández-Ledesma for providing the Spanish translation of the abstract.

Literature Cited

Bittrich, V., and U. Kühn. 1993. Nyctaginaceae. Pages 473–486 in The families and genera of flowering plants. Volume 2 ((K. Kubitzki, J. G. Rohwer, and V. Bittrich, editors). Springer-Verlag, Berlin, Germany.

Cruden, R. W. 1973. Reproductive biology of weedy and cultivated *Mirabilis* (Nyctaginaceae). American Journal of Botany 60:802–809.

Douglas, N. A., and P. S. Manos. 2007. Molecular phylogeny of Nyctaginaceae: taxonomy, biogeography, and characters associated with a radiation of xerophytic genera in North America. American Journal of Botany 94:856–872.

Galloway, L. A. 1975. Systematics of the North American desert species of *Abronia* and *Tripterocalyx* (Nyctaginaceae). Brittonia 27:328–347.

Hernández, H. 1990. Autopolinizacion en *Mirabilis longiflora* L. (Nyctaginaceae). Acta Botanica Mexicana 12:25–30.

Hodges, S. A. 1995. The influence of nectar production on hawkmoth behavior, self-pollination, and seed production in *Mirabilis multiflora* (Nyctaginaceae). American Journal of Botany 82:197–204.

Leal, A. A., Y. Terada, and M. D. F. P. da Silva Machado. 2001. Floral biology of a population of *Mirabilis jalapa* L. (Nyctaginaceae) from southern Brazil. Acta Scientiarum Universidade Estadual de Maringa 23:587–591.

McCulloch, S. R., and S. R. Searle. 2001. Generalized, linear, and mixed models. John Wiley and Sons, New York.

McGlaughlin, M., K. Karoly, and T. Kaye. 2002. Genetic variation and its relationship to population size in reintroduced populations of pink sand verbena, *Abronia umbellata* subsp. *breviflora* (Nyctaginaceae). Conservation Genetics 3:411–420.

Pilz, G. E. 1978. Systematics of *Mirabilis* subgenus *Quamoclidion* (Nyctaginaceae). Madroño 25:113–132.

Saunders, N. E., and S. D. Sipes. 2006. Reproductive biology and pollination ecology of the rare Yellowstone Park endemic *Abronia ammophila* (Nyctaginaceae). Plant Species Biology 21:75–84.

Spellenberg, R. 2000. Blooming ''behavior'' in five species of *Boerhavia* (Nyctaginaceae). Sida, Contributions to Botany 19:311–323.

Spellenberg, R. 2003. Nyctaginaceae. Pages 14–74 in Flora of North America north of Mexico. Volume 4 (Flora of North America Editorial Committee, editors). Oxford University Press, New York.

Tillett, S. S. 1967. The maritime species of *Abronia* (Nyctaginaceae). Brittonia 19:299–327.

Williamson, P. S., and S. K. Bazeer. 1997. Self-incompatibility in *Abronia macrocarpa* (Nyctaginaceae). Southwestern Naturalist 42:409–415.

Williamson, P. S., L. Muliani, and G. K. Janssen. 1994. Pollination biology of *Abronia macrocarpa* (Nyctaginaceae), an endangered Texas species. Southwestern Naturalist 39:336–341.

Submitted 18 December 2006. Accepted 7 November 2007. Associate Editor was David B. Wester.

American Journal of Botany 94(5): 856–872. 2007.

MOLECULAR PHYLOGENY OF NYCTAGINACEAE: TAXONOMY, BIOGEOGRAPHY, AND CHARACTERS ASSOCIATED WITH A RADIATION OF XEROPHYTIC GENERA IN NORTH AMERICA[1]

NORMAN A. DOUGLAS[2] AND PAUL S. MANOS

Department of Biology, Duke University, Box 90338, Durham, North Carolina 27708 USA

The four o'clock family (Nyctaginaceae) has a number of genera with unusual morphological and ecological characters, several of which appear to have a "tendency" to evolve repeatedly in Nyctaginaceae. Despite this, the Nyctaginaceae have attracted little attention from botanists. To produce a phylogeny for the Nyctaginaceae, we sampled 51 species representing 25 genera (of 28–31) for three chloroplast loci (*ndhF*, *rps16*, *rpl16*, and nrITS) and included all genera from North America. Parsimony, likelihood, and Bayesian methods were used to reconstruct the phylogeny for the family. The family is neotropical in origin. A radiation of woody taxa unites *Pisonia* and *Pisoniella* with the difficult tropical genera *Neea* and *Guapira*, which also form a clade, though neither appears to be monophyletic. This group is sister to a clade containing *Bougainvillea*, *Belemia*, and *Phaeoptilum*. A dramatic radiation of genera occurred in the deserts of North America. The tribe Nyctagineae and its subtribes are paraphyletic, due to over-reliance on a few homoplasious characters, i.e., pollen morphology and involucre presence. Two notable characters associated with the desert radiation are cleistogamy and edaphic endemism on gypsum soils. We discuss evolutionary trends in these traits in light of available data about self-incompatibility and gypsum tolerance in Nyctaginaceae.

Key words: biogeography; cleistogamy; gypsophily; homoplasy; mating system; Nyctaginaceae; phylogeny; pollen morphology.

Nyctaginaceae Juss. is a family of 28–31 genera and 300–400 species, that contains the familiar cultivated four o'clocks (*Mirabilis jalapa*) and bougainvillea (*Bougainvillea* spp.). Nyctaginaceae has long been known to be one of the core groups of families of Caryophyllales (Centrospermae) on the basis of the presence of betalain pigments, free-central placentation, p-type sieve tube elements, and the presence of perisperm, as well as molecular evidence (Bittrich and Kühn, 1993; Bremer et al., 2003). Within this group, the modern consensus is that Nyctaginaceae are closely related to certain monocarpellate members of a paraphyletic Phytolaccaceae, especially subfam. Rivinoideae (Rodman et al., 1984; Rettig et al., 1992; Downie and Palmer, 1994; Behnke, 1997; Downie et al., 1997; Cuenoud et al., 2002), although *Sarcobatus* (Sarcobataceae) has also been implicated as a close relative of this group (Behnke, 1997; Cuenoud et al., 2002).

Nyctaginaceae have a uniseriate petaloid perianth, usually interpreted as sepalous in origin (Rohweder and Huber, 1974). In most taxa the lower part of the perianth is fleshy or coriaceous and encloses the superior ovary, giving it the appearance of an inferior ovary. This accessory fruit is persistent and accrescent around the mature achene. While technically a diclesium (Bogle, 1974; Spellenberg, 2003), it is typically referred to as an "anthocarp."

Most genera can be recognized on the basis of fruit structure alone. In *Boldoa*, *Cryptocarpus*, and *Salpianthus*, the perianth is persistent but not accrescent, and thus these taxa lack the anthocarp (Bittrich and Kühn, 1993). In *Andradea*, *Leucaster*, and *Reichenbachia*, the perianth is variously accrescent but is not expanded (Bittrich and Kühn, 1993). However, in the remaining genera the anthocarp completely encloses the fruit and takes many forms (Willson and Spellenberg, 1977; Bittrich and Kühn, 1993). In taxa in which anthocarps are ribbed, the 3–10 ribs can be elaborated into wings (*Phaeoptilum*, *Grajalesia, Tripterocalyx*, *Abronia*, and some *Colignonia*, *Acleisanthes*, and *Boerhavia*), covered by viscid glandular hairs or warts (*Pisonia*, *Pisoniella*, *Cyphomeris*, *Commicarpus*, and some *Boerhavia* and *Acleisanthes*), or unelaborated, to leave an essentially gravity-dispersed fruit (*Mirabilis*, *Anulocaulis*, *Nyctaginia*, and some *Colignonia* and *Boerhavia*). Fleshy anthocarps are probably bird-dispersed in *Neea* and *Guapira*. They are also found in *Okenia*, though this genus is geocarpic and the seeds generally germinate at the spot where they are "planted" by the maternal individual (N. Douglas, personal observation). The unusual anthocarps of *Allionia* are boat-shaped, with two rows of inward-pointing teeth lining the concave side, suggesting possible exozoochory or wind dispersal, though no observations on this are available. In herbaceous taxa, at least, species-level characters are often found in this structure (Willson and Spellenberg, 1977; Spellenberg, 2003).

The family was treated by Heimerl in *Die Natürlichen Pflanzenfamilien* (Heimerl, 1889, 1934) and by Standley in several papers (Standley, 1909, 1911, 1918, 1931a, b) by which time most of the currently recognized genera had been described. Standley (1931a) formally transferred *Oxybaphus* L'Hér. ex Willd., *Hesperonia* Standl., *Quamoclidion* Choisy, and *Allionella* Rydb. into *Mirabilis*, though this has been overlooked in some floras (e.g., Kearney and Peebles, 1960).

[1] Manuscript received 22 June 2006; revision accepted 28 February 2007.

The authors thank the members of the Manos Lab at Duke University, D. Stone, and E. Christine Davis for discussion and helpful comments on the manuscript; F. Kauff for translations from German; M. Hilda Flores Olvera and P. Hernández-Ledezma for assistance in the field; the curators of DUKE, NMC, and NY for permission to sample from their collections; R. Wilbur, Strybing Arboretum, and the Duke University Botany Greenhouses for providing plant material used in this study; and Grand Canyon National Park for permission to collect. The Department of Biology (Duke), the Consortium in Latin American & Caribbean Studies (Duke), and the National Science Foundation provided funding for this project. The authors especially thank R. Spellenberg, who provided invaluable assistance in many aspects of this project.

[2] Author for correspondence (e-mail: nad@duke.edu)

TABLE 1. Classification of Nyctaginaceae and estimates of species number.

Tribe Subtribe	Genus	Species number	Spellenberg (2003), if different	Distribution
Leucastereae	*Leucaster* Choisy	1		SA
	Reichenbachia Spreng.	2		SA
	Andradea Fr. Allemão	1		SA
	Ramisia Glaz. ex Baillon	1		SA
Boldoeae	*Boldoa* Cav. ex Lagasca	1		SA, CA
	Salpianthus Humb. & Bonpl.	1		CA
	Cryptocarpus H.B.K.	1		SA
Abronieae	*Abronia* Juss. (incl. *Tripterocalyx* Hook. ex Standl.)	33	24*	NA
Nyctagineae				
Colignoniinae	*Colignonia* Endl.	6		SA
	Pisoniella (Heimerl) Standl.	1		SA, CA
Boerhaviinae	*Boerhavia* L.	20	ca. 40	Pantropical/subtropical
	Anulocaulis Standl.	4–5	5	NA
	Cyphomeris Standl.	2		NA
	Commicarpus Standl.	25	30–35	Pantropical/subtropical
	Caribea Alain	1		Cuba
	Acleisanthes A. Gray	7	17**	NA
	Selinocarpus A. Gray (incl. *Ammocodon* Standl.)	10	—	NA, Africa
	Okenia Schldl. & Cham.	1–2		NA, CA
Nyctagininae	*Mirabilis* L.	54	ca. 60	NA, Asia
	Cuscatlania Standl.	1		CA
	Allionia L.	2		NA, CA, SA
	Nyctaginia Choisy	1		NA
Phaeoptilinae	*Phaeoptilum* Radlk.	1		Africa
Bougainvilleeae	*Bougainvillea* Comm. ex Juss.	18		SA
	Belemia Pires	1		SA
Pisonieae	*Pisonia* L.	40	10–50	Pantropical/subtropical
	Guapira Aubl.	70	10–50	SA, CA
	Neea Ruiz & Pavon	83		SA, CA
	Neeopsis Lundell	1		CA
	Cephalotomandra Karst. & Triana	1–3		SA
	Grajalesia Miranda	1		CA

Note: Classification scheme according to Bittrich and Kühn (1993) and estimates of species number. For those genera treated in *Flora of North America* (Spellenberg, 2003), species number reflects newly described species and taxonomic readjustments. SA = South America; CA = Central America; NA = North America. * As 20 spp. *Abronia* and four spp. *Tripterocalyx*. ** Including *Selinocarpus* & *Ammocodon* (Levin, 2002).

Heimerl (1934) synthesized the family as it was known, including in his classification genera that had been recently described by Standley (i.e., *Pisoniella*, *Cuscatlania*). He based his supergeneric classification on a combination of plant habit, indumentum, linear vs. capitate stigma, straight vs. curved embryo, sex distribution, pollen grain morphology, and the occurrence of bracts or involucre (Bittrich and Kühn, 1993; Heimerl, 1934). Bittrich and Kühn (1993) provided the most recent summary of the classification at the tribal and subtribal level (Table 1). Their treatment broadly followed that of Heimerl (1934), adjusting ranks and incorporating genera described after 1934, i.e., *Caribea*. It recognized six tribes, two of which, Pisonieae and Nyctagineae, contain the majority of genera and species (Table 1, Pisonieae: six genera, ca. 200 spp.; Nyctagineae: 14 genera, ca. 100 spp.).

Whereas the bulk of diversity of Pisonieae resides in three highly similar arborescent genera with poorly differentiated species, Nyctagineae sensu Bittrich and Kühn (1993) is a diverse, mainly herbaceous, group recognized largely on the basis of very large (100–200 μm in diameter), pantoporate pollen grains, among the largest known in angiosperms (Stevens, 2001). The original formulation of tribe Mirabileae subtribe Boerhaaviinae (Heimerl, 1934), the antecedent of tribe Nyctagineae, was partly diagnosed by the presence of pantoporate pollen grains. Of four currently recognized subtribes, the Nyctagininae comprises those taxa with involucres, which may be of connate or distinct bracts. In contrast, the largest subtribe, Boerhaviinae, is composed of eight genera united primarily by their lack of involucral bracts. Four of these (*Boerhavia*, *Anulocaulis*, *Cyphomeris*, *Commicarpus*) have occasionally been treated as a single *Boerhavia* (Fosberg, 1978). This seems merely to reflect a preference for fewer large genera, because the four segregate genera are as distinct from each other as any other given pair of genera in the herbaceous group. The others, *Caribea*, *Okenia*, *Acleisanthes*, and *Selinocarpus* (including *Ammocodon*), were placed in Boerhaviinae on the basis of pollen morphology and the absence of involucres subtending flowers or inflorescences (though small subtending bracts may be present). The remaining two subtribes, Colignoniiae and the monospecific Phaeoptilinae, have aberrant morphology compared to Nyctagininae and Boerhaviinae, for example, pollen grains in *Colignonia* and *Pisoniella* are dramatically smaller, and in *Phaeoptilum* they are pantocolpate. In *Pisoniella*, the embryo is straight, typical of Pisoneae, instead of a hooked embryo that encircles the perisperm as found in the remaining Nyctagineae (Bittrich and Kühn, 1993). Additionally, the shrubby, scandent or lianoid growth habits of these taxa are rare in the other subtribes, which are mostly perennial herbs. Though Heimerl placed *Colignonia* in a monogeneric tribe Colignonieae, Bittrich and Kühn (1993) include subtribe Colignoniinae (including *Pisoniella*) in tribe Nyctagineae, uniting all taxa with

pantoporate grains and *Phaeoptilum* (with its pantocolpate grains) in one tribe.

Two major centers of distribution have been noted for the Nyctaginaceae (Standley, 1909). The first is in the neotropics and Caribbean, characterized by arborescent genera such as *Neea*, *Guapira*, *Pisonia*, and *Bougainvillea*, as well as the herbaceous *Colignonia* and *Salpianthus*. The second is in arid western North America, where several herbaceous or suffrutescent genera are native, including *Boerhavia*, *Mirabilis*, *Abronia*, *Acleisanthes* sensu Levin (2002), and *Commicarpus*. A few genera are widespread in tropical and subtropical regions of the world (*Boerhavia*, *Commicarpus*, *Pisonia*): *Mirabilis* is present in North and South America with one species in Asia, and *Acleisanthes* contains the disjunct *A. somalensis* from Somalia. *Mirabilis* (*M. jalapa*, *M. oxybaphoides*) and *Bougainvillea* (*B. glabra*, *B. spectabilis*, *B. peruviana*, and numerous hybrid cultivars) are naturalized in many parts of the world. Only one genus is restricted to the Old World, the monospecific *Phaeoptilum* of southwestern Africa.

The first molecular phylogenetic study of Nyctaginaceae was presented by Levin (2000). The focus was on species in certain genera of tribe Nyctagineae sensu Bittrich and Kühn (1993), including genera in subtribes Nyctagininae (*Allionia*, *Mirabilis*) and Boerhaviinae (*Acleisanthes*, *Selinocarpus*, *Boerhavia*), as well as *Abronia* and *Pisonia*. The study justified the formal combination of *Acleisanthes*, *Selinocarpus*, and *Ammocodon* (Levin, 2002; Spellenberg and Poole, 2003), but due to limited sampling of genera, it was not possible to evaluate the monophyly of the subtribes of Nyctagineae (Levin, 2000). The *Flora of North America* treatment of Nyctaginaceae (Spellenberg, 2003), while not referring to tribal classification, reflected these and other taxonomic changes for the genera and species that occur in North America north of Mexico (Table 1).

In the herbaceous taxa of Nyctaginaceae found in the deserts of North America, several unusual characters occur with notable frequency. As indicated by the common name for the family, species in several genera (*Anulocaulis*, *Cyphomeris*, *Acleisanthes*, *Mirabilis*, *Abronia*, and *Tripterocalyx*) flower in the evening and are adapted to moth pollination (Baker, 1961; Grant, 1983; Grant and Grant, 1983; Hernández, 1990; Hodges, 1995; Levin et al., 2001). Internodal bands of viscid secretions, which may discourage aphid colonization (McClellan and Boecklen, 1993), are present in *Anulocaulis*, *Cyphomeris*, and some species of *Boerhavia*. As mentioned, anthocarp morphology is also variable, with wings and viscid glands being common modifications.

Because these characters are often polymorphic at the generic level, they would seem to represent evolutionary "tendencies." Sanderson (1991) discussed evolutionary tendencies in explicit phylogenetic terms: a tendency is a concentrated distribution of homoplasy within a tree. The main objection to the study of tendencies is the difficulty in defining the taxonomic scope at which they operate, in other words, it is "... biologically inappropriate [when investigating a hypothesized tendency] to include taxa that cannot under any circumstances exhibit the states of interest" (Sanderson, 1991, p. 357). Thus, when considering whether a character has a tendency to evolve, it is first necessary to evaluate the range of taxa in which it could potentially appear. In some cases, it may be possible to identify another character upon which the evolution of the character of interest is dependent. If this other trait is itself uniquely derived, its occurrence will define the group in which the tendency may conceivably exhibit itself. If the independent character is itself derived multiple times, then the problem is pushed back so that the challenge is first to explain the tendency for the *independent* character to evolve in the group.

In the case of tendencies in Nyctaginaceae, it is not immediately obvious what sorts of traits may be required to enable, for instance, a shift to nocturnal pollination or the development of viscid bands on stem internodes. There are two traits, however, that seem to have a tendency to evolve in Nyctaginaceae and that we can reasonably assume are contingent on other traits: the evolution of cleistogamy is improbable without prior self-compatibility, and lineages that specialize on gypsum are unlikely to have arisen from lineages with no latent or expressed gypsum tolerance.

Cleistogamous (closed, self-fertilizing) flowers are produced in addition to chasmogamous (open) flowers in four genera of Nyctaginaceae: *Acleisanthes*, *Cyphomeris*, *Nyctaginia*, and some *Mirabilis* (Cruden, 1973; Spellenberg and Delson, 1974; Fowler and Turner, 1977; Levin, 2002). Though species with cleistogamous flowers have evolved in a number of angiosperm families, only in much larger families, e.g., Poaceae, Fabaceae, and Malpighiaceae, is this trait found in as many genera (Lord, 1981). Despite a long awareness of this phenomenon generally (Darwin, 1884), the evolution of this character has only rarely been investigated with phylogenetic methods (Desfeux et al., 1996; Bell and Donoghue, 2003).

Second, as in many caryophyllid families, e.g., Amaranthaceae and Portulacaeae, there is a propensity in many Nyctaginaceae to be tolerant of, or specialists of, gypseous soils. Outcrops of gypsum (hydrous calcium sulfate) are quite common in arid North America, especially in the Chihuahuan Desert. These areas have a flora characterized by gypsophiles, which never occur on other substrates, and gypsum-tolerant species, which are found on both gypseous and nongypseous soils (Waterfall, 1946; Parsons, 1976; Meyer, 1986). In the United States and Mexico, Nyctaginaceae are well represented in gypsum communities (Parsons, 1976). At least 25 species in seven genera are known to occur on gypsum. Of these, roughly half are known gypsophiles, found only on gypsum soils (Johnston, 1941; Waterfall, 1946; Fowler and Turner, 1977; Turner, 1991, 1993; Spellenberg, 1993, 2003; Mahrt and Spellenberg, 1995; Harriman, 1999; Levin, 2002).

Although gypsum soils support a distinct flora, the evolution of gypsophily is not understood as well as other cases of edaphic endemism. Gypsum is not an inherently poor substrate for plants in the same way as soil with, for instance, toxic levels of heavy metals (Cockerell and Garcia, 1898; Johnston, 1941; Loomis, 1944; Parsons, 1976; Meyer, 1986; Oyonarte et al., 2002). Recent experimental work has pointed toward mechanical, rather than chemical, factors to explain the limited flora of gypsum soils: seedlings of nongypsophiles are unable to penetrate the hard crust typical of gypseous soils. This indicates that adaptations of gypsum-tolerant taxa primarily act to enhance survival in the establishment stage (Meyer, 1986; Meyer et al., 1992; Escudero et al., 1997, 1999, 2000; Romao and Escudero, 2005).

Edaphic-endemic species are sometimes found to be related to species that are merely tolerant: in the case of a serpentine endemic species of *Layia* (Asteraceae), certain populations of a non-endemic progenitor species were found to tolerate serpentine soils (Baldwin, 2005). Thus, even in the case of highly toxic soils, saltational speciation (Antonovics, 1971;

TABLE 2. Primer sequences used and original publication.

Region Primer name	Sequence	Reference
ITS		
ITS4	TCCTCCGCTTATTGATATGC	White et al., 1990
ITS5a	CCTTATCATTTAGAGGAAGGAG	Stanford et al., 2000
ndhF		
Nyct_ndhF1F	TGCCTGGATTATACCCTTCA	This study
NdhF972F	ATGTCTCAATTGGGTTATATGATG	Olmstead and Sweere, 1994
Nyct_ndhF13R	CAFCBGGATTACYGCATTT	This study
Nyct_ndhF22R	CTTGTAACGCCGAAACCATT	This study
Nyct_ndhF6F	AACGGGBAGTTTYGARTTTG	This study
Nyct_ndhF8R	AGTAGGCCCCTCCATAGCAT	This study
Nyct_ndhF14F	TCAATCGTTGCAATCCTTCT	This study
Nyct_ndhF16R	TTTCCGATTCATGAGGATATGA	This study
rps16		
rpsF	GTGGTAGAAAGCAACGTGCGA	Oxelman et al., 1996 (modified)
Rps2R	TCGGGATCGAACATCAATTGCAAC	Oxelman et al., 1996
rpl16		
F71	GCTATGCTTAGTGTGTGACTCGTTG	Jordan et al., 1996
R1661	CGTACCCATATTTTTCCACCACGAC	Jordan et al., 1996

Kruckeberg, 1986) is not required to explain edaphic endemism. These lines of evidence, and the fact that roughly half of the species of Nyctaginaceae found on gypsum are not restricted to it, make it reasonable to assume that an underlying ability to survive in gypsum soils is an early stage in the evolution of this type of edaphic endemism in Nyctaginaceae.

In principle, for both of these examples, the evolution of both the independent and contingent characters can be reconstructed on a phylogeny. With an understanding of the distribution of homoplasy in Nyctaginaceae, we will have a more robust framework for asking questions about character evolution and adaptation to xeric environments. In this phylogenetic study we comprehensively sample the genera of Nyctaginaceae, with the following goals: (1) to evaluate the existing classification of Bittrich and Kühn (1993), (2) to understand the biogeographic history of the family, and (3) to have a basis for understanding the evolutionary history of characters of historical taxonomic importance and the potential adaptive significance as manifested in their "tendency" to evolve repeatedly in lineages occurring in the deserts of North America.

MATERIALS AND METHODS

Sampling—Fifty-one species representing 25 genera of Nyctaginaceae were sampled. Taxa, voucher information, and GenBank numbers are given in Appendix 1. Our sampling is nearly comprehensive at the generic level, with representative species of every genus except *Neeopsis*, *Cephalotomandra*, *Grajalesia*, *Cuscatlania*, *Boldoa*, and *Cryptocarpus*. The genera omitted are monotypic, rarely collected, and/or of dubious distinction. For example, *Boldoa purpurascens* is often included in *Salpianthus* (Pool, 2001). All tribes and subtribes recognized by Bittrich and Kühn (1993) are included. Because different taxa have been found to be sister to Nyctaginaceae (Rettig et al., 1992; Behnke, 1997; Downie et al., 1997; Cuenoud et al., 2002), outgroups were selected from both Phytolaccaceae and Sarcobataceae. More distantly related taxa in the "core Caryophyllales," i.e., Aizoaceae, Molluginaceae, and Stegnospermataceae (Cuenoud et al., 2002), were also included to enable us to test the monophyly of Nyctaginaceae and to identify which taxa are sister to the family. For four species, data were obtained from two different accessions, and for two, GenBank sequences were used for some loci. "*Phytolacca*" is a composite of one GenBank sequence from *P. acinosa* and three new sequences from *P. americana*.

Molecular data—Genomic DNA was extracted from fresh, silica-dried, or air-dried (herbarium) leaf tissue using either Qiagen DNAeasy Plant Mini Kits or a modified CTAB method (Doyle and Doyle, 1987). Internal transcribed spacer (ITS) sequences were obtained using primers ITS4 and ITS5a (White et al., 1990; Stanford et al., 2000), which amplifies ITS1, 5.8S, and ITS2. Chloroplast *ndhF* sequences were obtained as two overlapping fragments using primers Nyct-ndhF1, ndhF972, Nyct-ndhF13R, and Nyct-ndhF22R. With the exception of ndhF972 (Olmstead and Sweere, 1994), these were designed based on GenBank *ndhF* sequences for Nyctaginaceae and Phytolaccaceae. Many samples, especially those from herbarium materials, were recalcitrant to PCR of long (>1 kb) fragments due to DNA degradation; for these, four additional primers (Nyct-ndhF6F, Nyct-ndhF8R, Nyct-ndhF13F, and Nyct-ndhF16R) were designed, based on sequences for Nyctaginaceae and Phytolaccaceae, and used in conjunction with the aforementioned primers, so that the gene was amplified in four overlapping fragments. The chloroplast intron *rps16* was amplified using primers rpsF and rps2R (Oxelman et al., 1997), and *rpl16* was obtained using primers F71 and R1661 (Jordan et al., 1996). Primer sequences and references are given in Table 2. PCR products were cleaned with Qiaquick columns (Qiagen, Valencia, California, USA). Cycle sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit, and sequences were determined with an ABI 3700 DNA Analyzer (Applied Biosystems, Foster City, California, USA) in the Genetic Analysis facility in the Department of Biology at Duke University. Raw chromatograms were edited and assembled in Sequencher 4.1 (Gene Codes Corp., Ann Arbor, Michigan, USA). Sequence alignment was performed either by eye (*ndhF*) or in ClustalX (Thompson et al., 1997) (other regions) followed by manual adjustment in Se-Al (Rambaut, 1996). Across the entire data set, ITS1 and ITS2 were too variable to be confidently aligned, although the 5.8S region was highly conserved. Ambiguously aligned regions were excluded from further analyses of the entire data set, though they were used in analyses of more restricted taxon sets (see *Restricted analyses*).

Caribea littoralis Alain, a Cuban endemic, has been collected only once. The collection locality is in southeastern Cuba in a dry coastal habitat. The morphology of the plant is difficult to interpret because it is highly distinct from any other member of the Nyctaginaceae, and the leaves and flowers are highly reduced. Few details are clearly visible on the specimen, though the description appears to have been based on fresh material (Alain, 1960). Due to the age of the collection, only about 25% of an *ndhF* sequence was obtainable. This sequence was unique in our data set, and a BLAST search found that this sequence fragment was most similar to an existing *Bougainvillea ndhF* sequence (GenBank no. AF194825). Preliminary phylogenetic analysis (see *Data analysis*) placed this taxon as sister to either *Pisoniella* or *Belemia*. These last two are not closely related to each other, resulting in substantial loss of resolution in the clade including these taxa. Therefore, *Caribea* was excluded from all further analysis, and while this result confirms that this enigmatic taxon belongs in Nyctaginaceae, further study must await rediscovery of this species. Unfortunately, repeated attempts to relocate the population at the type locality

TABLE 3. Summary of sequence statistics by partition for the molecular matrix.

	Full analysis								
Partition	*ndhF* (entire)	*ndhF* (1st pos.)	*ndhF* (2nd pos.)	*ndhF* (3rd pos.)	*rps16*	*rpl16*	5.8S	Entire	Chloroplast
No. taxa (full matrix = 58)	54	54	54	54	51	42	55	55	55
Aligned length	2193	731	731	731	1237	1367	157	5505	4797
Analyzed length	2205	669	668	668	780	792	157	3734	3577
Constant	1348	474	532	342	520	552	136	2556	2420
Uninformative	272	83	69	120	110	139	5	526	521
Parsimony-informative	385	112	67	206	150	101	16	652	636
ML model	GTR+I+Γ	GTR+I+Γ	GTR+I+Γ	GTR+Γ	GTR+Γ	GTR+Γ	SYM+I+Γ	GTR+I+Γ	GTR+I+Γ

Note: Maximum-likelihood (ML) model estimated by ModelTest (Posada and Crandall, 1998): Full, Entire; Sensitivity, Entire; Restricted I & II, Entire. ML model for remaining partitions (used in Bayesian analyses "B2," "B4," and "B6," see text) estimated with MrModelTest (Nylander, 2004). Numbers in parentheses are number of informative characters gained from the inclusion of ITS1 and ITS2 in restricted analyses.

in Cuba have proved unsuccessful (D. Stone, Duke University, personal communication).

Data analysis—Initial maximum parsimony (MP), maximum likelihood (ML), and Bayesian analyses were performed for each of the four loci. The 5.8S, not surprisingly, had low variation and produced poorly resolved trees; however, examination of the support values for the topology favored by each locus revealed no supported nodes in conflict. Therefore, the data sets were combined for further analyses.

MP analysis was performed using PAUP* version 4.0b10 (Swofford, 2002). A heuristic search was performed, with 1000 replicates of 10 random-addition sequences, tree-bisection-reconnection (TBR) branch swapping, MAXTREES set to autoincrease, MULTREES = yes. Support was evaluated using 1000 bootstrap replicates of 10 random addition sequences, TBR branch swapping, MULTREES = YES.

For the ML analysis, the data set was first examined using ModelTest 2.0 (Posada and Crandall, 1998), which selected a complex model of evolution (GTR + I + Γ). Ten random-addition replicates (TBR, MAXTREES set to autoincrease, MULTREES = yes) were run in PAUP*. Maximum-likelihood bootstrap support values were obtained by 100 replicates of single random-addition sequences, TBR branch swapping, MULTREES = yes.

Bayesian analysis was performed using MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). For exploring the effect of different models for different partitions of the data, best-fit models for each partition were estimated in MrModelTest (Nylander, 2004), which selects the best-fit model from those available in MrBayes. The partitions were as follows: 1, all loci together; 2, nuclear 5.8S; 3, all chloroplast loci; 4, *rpl16*; 5, *rps16*; 6, *ndhF*; and 7, 8, 9, first, second, and third positions of *ndhF*, respectively. The models selected by MrModelTest for each partition are given in Table 3. Bayesian searches were then performed on the entire data set using four partition/model combinations: "B1," single model for all partitions, (1); "B2," nuclear and chloroplast, (2 and 3); "B4," all loci, (2, 4, 5, and 6); and "B6," all loci with separate models for each codon position of *ndhF* (2, 4, 5, 7, 8, and 9). For each combination, we executed four independent runs of 1×10^6 generations each, sampling every 100th tree. After discarding trees from the burn-in (determined by visualizing the plateau in –lnL scores, approximately after 50 000 generations), we compared the posterior tree sets from each run by computing a 50% majority rule tree in PAUP*. No strongly supported topological differences (at posterior probability ≥95%) were found between the four runs of each model set. Therefore, the four posterior tree files for each set of models were combined into a single posterior tree file for purposes of assessing support values yielded by each set of models. These preliminary analyses were conducted including the partial *ndhF* sequence for *Caribea*; however, the B6 analysis was repeated without this sequence.

Sensitivity analyses—Due primarily to the inclusion of GenBank sequences for outgroup taxa and the failure of certain loci to amplify (mostly from herbarium material), approximately 17.7% of the data matrix was coded as "missing." The potential impact of this was investigated by deleting from the analysis 18 taxa (Appendix 1) for which one or more sequences were entirely missing and by combining sequences from *Bougainvillea glabra* and *B. infesta* into a composite operational taxonomic unit "*Bougainvillea*." "*Phytolacca*" and *Rivina humilis* were the only remaining outgroups in this analysis, which allowed us to examine the effect of including distant outgroups. The MP, ML, and corresponding bootstrap searches were performed with the same settings as in the analysis of the full matrix. The resulting trees were compared to the topology from the full analysis to see whether the exclusion of missing data led to a preferred topology that differed substantively from the topology or levels of support in the analysis of the full matrix.

Restricted analyses—To gain resolution within and between closely related genera, our selection of loci encompassed a large range of sequence variation. Because both the ITS1 and ITS2 regions had to be excluded from the analysis of the complete data set due to questionable alignment (though the highly conserved 5.8S region was kept in the full matrix), following the analysis of the full data set, two restricted data sets were constructed to allow us to increase the number of included characters (Table 3) by reducing the taxon sampling to two distinct clades found in the full analysis. These restricted data sets comprised all included nucleotide positions in the full data set, plus sites that were unalignable across the breadth of taxa included in the full data set, but that were alignable within each of the restricted sets of taxa. The first restricted analysis group was comprised of North American herbs representing all taxa in the sister group to *Allionia*, whereas the second corresponded to the Pisonieae, *Bougainvillea*, *Belemia*, and *Phaeoptilum* (the "B&P" clade from the full analysis). The MP, ML, and corresponding bootstrap analyses were performed in the same fashion as in the full matrix and sensitivity analyses, with the exception that the ML models were reestimated in ModelTest.

Character data—The historical taxonomic significance given to pollen morphology and involucral bracts led us to examine these characters in a phylogenetic context. Pollen data follows the scheme of Nowicke, who identified four types in Nyctaginaceae (Nowicke, 1968, 1970, 1975; Nowicke and Luikart, 1971; Reyes-Salas and Martínez-Hernández, 1982; Chavez et al., 1998). Pollen type was coded as a multistate, unordered character. In many cases, the exact species included in our study were not examined in the published studies. If there was no indication of within-genus pollen polymorphism, that pollen type was assigned to all species in this analysis. However, multiple pollen types were recorded within *Neea* and *Pisonia*. Thus, only *N. psychotrioides*, which was examined by Nowicke, was coded unambiguously; other species of *Neea* and *Pisonia* were coded as polymorphic (states "1&3" and "1&4," respectively) to reflect this uncertainty in the assignment of ancestral states. The presence of involucral bracts was scored as present/absent. If only small subtending bracteoles occur (common in many taxa), this character was coded as "absent," mirroring the usage of this character in defining subtribe Nyctagininae. The occurrence of cleistogamous flowers was scored based primarily on literature sources (Spellenberg and Delson, 1974; Bittrich and Kühn, 1993; Levin, 2002; Spellenberg, 2003). Gypsophilic taxa were identified in literature sources (Waterfall, 1946; Parsons, 1976; Fowler and Turner, 1977; Turner, 1991; Harriman, 1999; Levin, 2002; Spellenberg, 2003; N. Douglas, personal observation). Taxa were identified as full gypsophiles (recorded only from gypseous soils), gypsum tolerant (recorded from both gypseous and nongypseous soils), or nongypsophilic. Taxa that do not occur in areas with gypsum outcrops were considered to be nongypsophilic. It is unlikely that transitions to or from full gypsophily could evolve with no intermediate gypsum-tolerant step; therefore, this character was analyzed as both unordered and ordered, with two steps required between

TABLE 3. Extended.

Sensitivity analysis	Restricted analysis I	Restricted analysis II
Entire	Entire	Entire
39	19	15
5359	4887	4914
3396	4278	4082
2597	3883	3526
412	160	382
387	235 (122)	174 (76)
GTR+I+Γ	GTR+I+Γ	GTR+I+Γ

nongypsophily and full gypsophily. Parsimony ancestral states of all characters were reconstructed with the program Mesquite 1.6 (Maddison and Maddison, 2006). Those terminals that were not assigned a single state, and branches that were not unambiguously resolved, are depicted as "equivocal."

RESULTS

Data matrix—The entire data matrix (Table 3) had a length of 5505 bp, of which 1771 were excluded due to ambiguous alignment, mainly due to the presence of length variation in ITS1 and ITS2 and in the two chloroplast introns, *rpl16* and *rps16*. Of the remaining 3734 characters, 652 were parsimony informative.

Phylogenetic analysis of the complete dataset—The MP analysis resulted in 36 shortest trees (length: 2287, consistency index [CI]: 0.657, retention index: 0.809, rescaled CI: 0.531); however, the strict consensus (tree not shown) resolved all but two ingroup nodes. Thirty-nine nodes were supported with parsimony bootstrap values (MPBS) $\geq$70.

The best-fit model as determined by ModelTest (Table 3) using both a hiearchical liklihood ratio test (HLRT) and the Akaike information criterion (AIC) was a general-time-reversible model with a proportion of invariant sites and a gamma shape parameter (GTR+I+Γ). The ML search returned a single ML tree, which was nearly identical to the MP topology, except in the placement of the genus *Colignonia*. This taxon is placed as sister to the large clade containing *Acleisanthes* and *Boerhavia* in the MP analysis (MPBS = 80) and is not resolved with strong support in any ML or Bayesian analysis. Overall, 38 nodes in the ML analysis were supported with likelihood bootstrap values (MLBS) $\geq$70.

Models determined by MrModelTest for each data partition in the Bayesian analyses are given in Table 3. On the basis of our preliminary examination of partitioned models, the signal in the data set apparently is strong, and the topology is not contingent on model selection: the tree topologies produced by the Bayesian B1, B2, B4, and B6 searches were consistent. The principal difference between them is in the level of support for the topology, with 37, 39, 40, and 40 nodes, respectively, supported by posterior probabilities (PP) $\geq$95%. Deletion of *Caribea* led to the resolution, with support of two additional nodes in the repeated B6 search, for a total of 42 nodes supported at greater than 95% PP. The topology of this Bayesian B6 consensus tree is identical to the ML tree. All further Bayesian support values refer to the B6 analysis.

The Nyctaginaceae are supported as monophyletic by ML (MLBS = 71) and Bayesian (PP = 100) analyses (Fig. 1). Interestingly, in the MP bootstrap analysis of this matrix, the monophyly of the Nyctaginaceae is not supported. Despite the inclusion of several outgroups, no single sister lineage emerges with strong support.

Leucastereae, a tribe of four South American genera (*Andradea*, *Leucaster*, *Ramisia*, and *Reichenbachia*), is supported as the earliest branching lineage in Nyctaginaceae (Fig. 1) followed by Boldoeae, represented by *Salpianthus*. A clade containing largely neotropical trees and shrubs, and the African genus *Phaeoptilum*, receives support from MP and B6 analyses, though not from ML. We will refer to this group as the Bougainvilleeae and Pisonieae ("B&P") clade, (Fig. 2), recognizing that it also includes *Phaeoptilum* and *Pisoniella*, which are currently classified in Nyctagineae. *Bougainvillea*, *Belemia*, and *Phaeoptilum* form a clade within this group, which is sister to a clade containing the Pisonieae and the genus *Pisoniella*. Within the Pisonieae, *Neea* and *Guapira* together form a clade but neither genus appears to be monophyletic.

Strong support is found in all analyses for a clade including mostly North American xerophytic genera. For the purposes of this paper, we refer to it as the North American Xerophytic ("NAX" clade) (Fig. 2). The NAX clade is well defined by geography, habit, and habitat, but it has never been recognized formally. The earliest branch within this clade leads to *Acleisanthes* sensu Levin (2000). It is followed by a clade representing *Abronia* and *Tripterocalyx* (tribe Abronieae). Phylogenetic relationships of the remaining genera in the NAX clade are mostly well resolved, with the exception of low support values for the placement of *Commicarpus* and *Allionia*. Two pairs of genera in this clade are not resolved as monophyletic. *Anulocaulis* includes *Nyctaginia*, and *Boerhavia* includes *Okenia*, though support in both of these cases is weak or lacking. Examination of the branch lengths (Fig. 2) makes it clear that *Anulocaulis* and *Nyctaginia* are at least very closely related.

The position of *Colignonia* is not resolved in the ML and Bayesian analyses. The ML analysis resolves *Colignonia* sister to the B&P and NAX clades but with weak support. A position sister to only the NAX clade is supported in the MP analysis.

Sensitivity analyses—The deletion of taxa with significant missing data resulted in a matrix of 39 taxa with only 3.1% missing data, as compared to 58 taxa with 17.7% missing data in the full analysis (See Appendix 1). The MP/ML analyses of this matrix yielded trees (not shown) that had no well-supported nodes conflicting with the topology of the tree from the full matrix. The support for the monophyly of the Nyctaginaceae increased to 94/95 MPBS/MLBS, from −/71 in the analysis of the full data set. The high level of support found in this analysis for the monophyly of Nyctaginaceae indicates that the inclusion of many outgroups in the full matrix, including the quite distant *Stegnosperma*, may have affected the level of support in the MP analysis. Alternatively, high levels of missing data in the full data set may be responsible for low support values at this key node. Support for the placement of *Cyphomeris* decreased to 70/66 relative to the full analysis. *Commicarpus* and *Allionia* increased to 73/67 and 87/77, respectively; these nodes had not received strong support in any analysis of the full data set. The remainder of the comparable nodes were similarly supported between the full and sensitivity analyses.

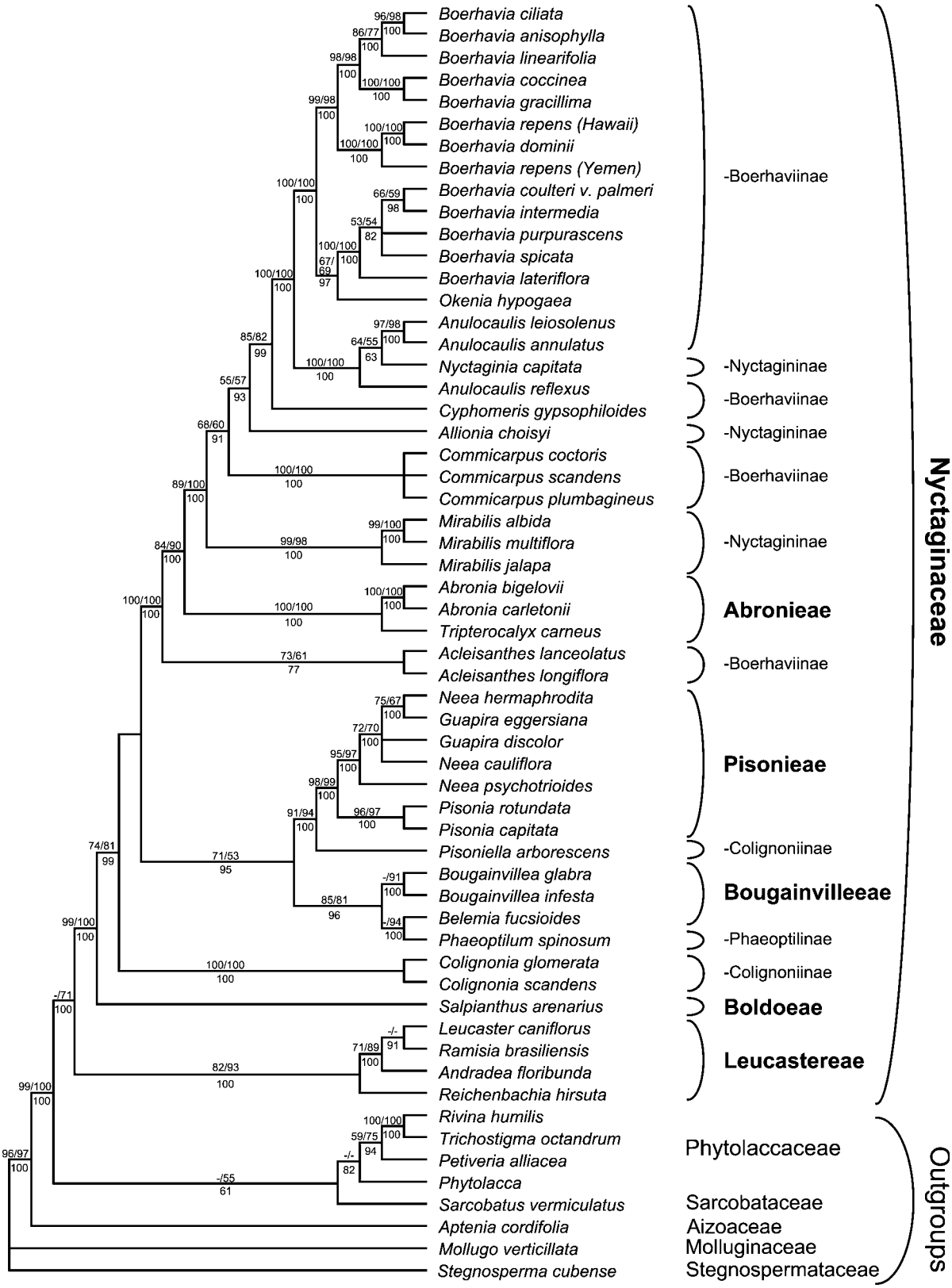

Fig. 1. Maximum-likelihood (ML) topology from the analysis of the entire data set. Parsimony bootstrap/ML bootstrap support values above branches, Bayesian posterior probability from the "B6" analysis below branches, "-" indicates bootstrap support value <50. tribes of Nyctaginaceae according to Bittrich and Kühn (1993) are in bold. "-" before unbold name signifies a subtribe of tribe Nyctagineae.

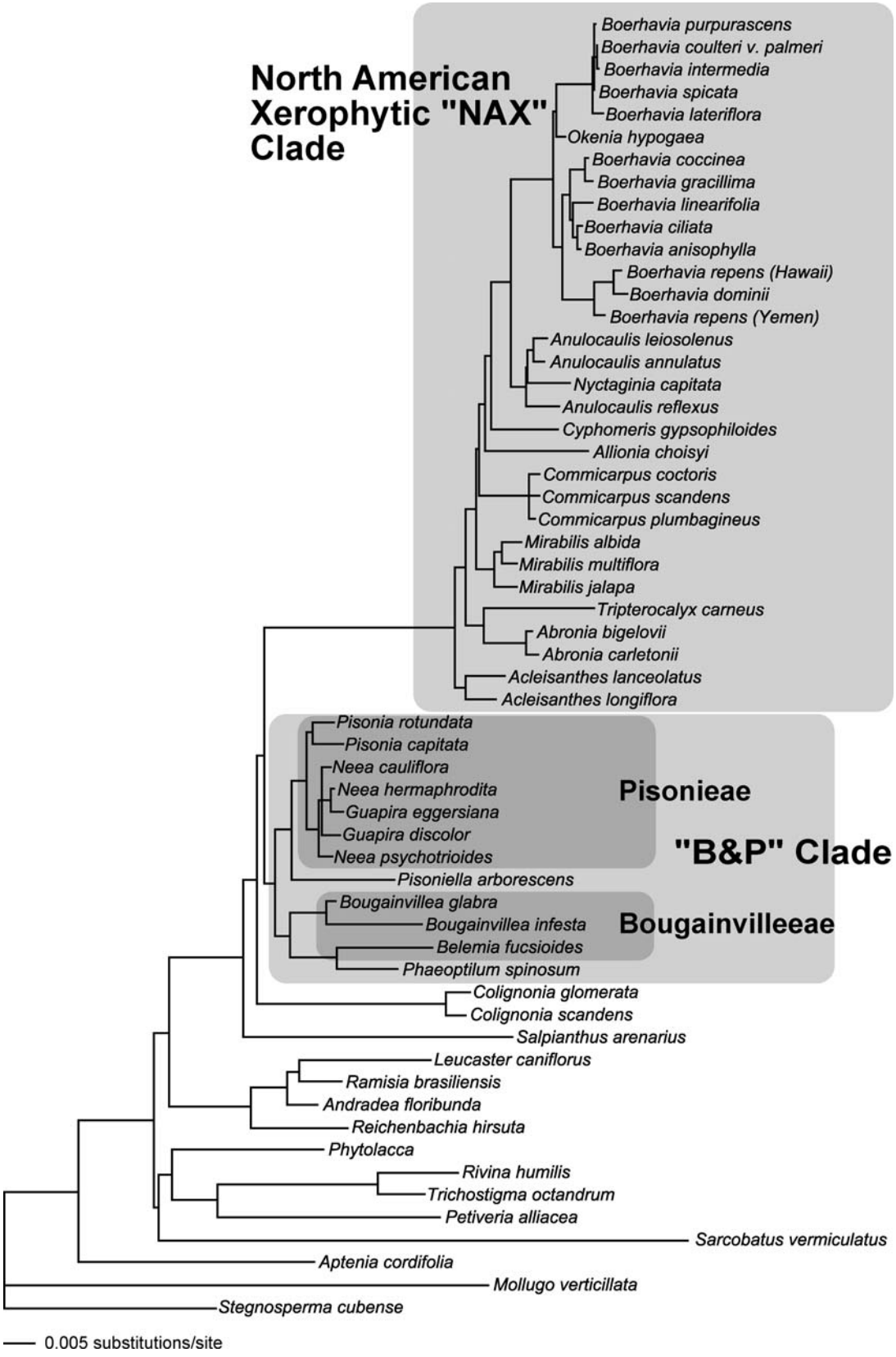

Fig. 2. Phylogram of the maximum-likelihood topology from Fig. 1. Major clades referred to in text are highlighted.

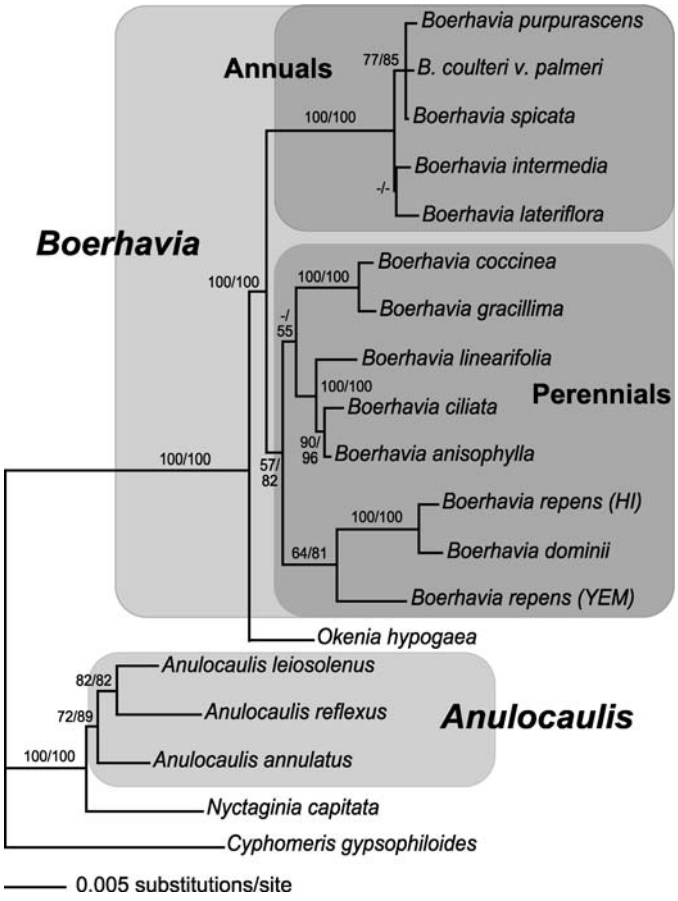

Fig. 3. Phylogram of the maximum-likelihood (ML) topology from the first restricted analysis. MP bootstrap/ML bootstrap support values are shown. *Anulocaulis* and *Boerhavia* are each supported as monophyletic.

Restricted analyses—For the two restricted analysis groups, 122 and 76 additional informative characters were gained with the inclusion of ITS1 and ITS2, respectively (Table 3). A small number of additional sites were gained from the chloroplast introns *rps16* and *rpl16* (<5 characters in either data set). ModelTest 3.7 selected a GTR+I+Γ model for each (Table 3) data set. For the first group (all taxa in the sister group to *Allionia*), MP and ML analyses produced a tree (Fig. 3) with improved resolution in the *Anulocaulis* + *Nyctaginia* clade. Though the placement of *A. annulatus* differs between the full matrix topology and the restricted analysis, the monophyly of *Anulocaulis* was well supported with bootstrap values of 72/89 MPBS/MLBS. Similarly, the full analysis resolves *Okenia* within a paraphyletic *Boerhavia* with low MP and ML bootstrap support, but 97% Bayesian posterior probability, yet the restricted analysis found *Boerhavia* strongly supported as monophyletic and sister to *Okenia* with high support (100/100). *Boerhavia* consists of two clades, corresponding to annual and perennial species, that were also found in the full analysis.

The restricted analysis of the B&P clade produced a tree (not shown) that did not conflict with the topology of this clade in the full matrix analysis. Support values were generally slightly lower, probably due to the concentration of missing data in this group and the lower number of additional characters from the ITS region. Support values remained high for the nodes uniting *Guapira eggersiana* and *Neea hermaphrodita*, for the placement of *G. discolor* in the clade sister to *N. psychotrioides*, and for the monophyly of *Neea* + *Guapira* (MPBS/MLBS bootstrap support of 64/73, 82/94, and 94/92, respectively).

Character reconstructions—For each character reconstructed (Fig. 4), multiple state transitions are inferred. Tricolpate-spinulose pollen (Fig. 4a) appears to be the ancestral condition in the group, transitioning to a pantoporate-spinulose condition subsequent to the divergence of *Salpianthus* from the main lineage. The latter condition is found in nearly all members of the NAX clade, yet appears to predate that group. At least eight transitions among the four pollen types have occurred in the Nyctaginaceae. Considering the small number of *Neea* and *Pisonia* examined and the polymorphism exhibited by these genera, the number of transitions could be higher. Reconstruction of involucral bracts shows five gain/loss steps. This character is fixed within genera, thus this interpretation is likely to be affected only by the future inclusion of the remaining genera in the family. Only the inclusion of *Cuscatlania*, which has an involucre, could conceivably change the number of steps required. Cleistogamous flowers are uniquely derived in four genera. Gypsophily requires nine or 13 steps to explain, depending on whether it is considered to be an unordered or an ordered character. Reconstructions were performed only on the ML topology from the full analysis. Adjusting the positions of *Okenia* and *Nyctaginia* to reflect the topology from the restricted analysis (Fig. 3) results in the branches leading to *Nyctaginia* + *Anulocaulis* and *Nyctaginia* + *Anulocaulis* + *Okenia* + *Boerhavia* being resolved as nongypsophilic. Treating gypsophily as an unordered character has the same result. Otherwise, the alternative topology has no substantive effect on the conclusions we make regarding the degree of homoplasy shown by the remaining three characters shown in Fig. 4.

DISCUSSION

Phylogeny of Nyctaginaceae—The earliest branching lineage in Nyctaginaceae, the Leucastereae (Fig. 1), had been previously recognized as a natural group on the basis of arborescence, a stellate indumentum, and tricolpate pollen (Heimerl, 1934; Bittrich and Kühn, 1993). The Boldoeae, an herbaceous group native from the Galapagos to northwestern Mexico and the Caribbean, are represented in this study by *Salpianthus*. These two lineages had been predicted to be basal or outside of Nyctaginaceae on the basis of apparent pleisomorphies such as alternate leaves and bisexual flowers (Bittrich and Kühn, 1993). The anthocarp structure is absent in Leucastereae and Boldoeae, although the unexpanded perianth does persist around the fruit. Persistent tepals are also found in many Phytolaccaceae. However, the perianth consists of free tepals in most Phytolaccaceae and all of subfamily Rivinoideae (except *Hilleria*, in which three of four tepals are partially fused, (Rohwer, 1993)). In Nyctaginaceae, including Leucastereae and Boldoeae, tepals are fully connate.

Within the B&P clade (Fig. 2), *Phaeoptilum* is found to be sister to *Belemia*, rendering the Bougainvilleeae paraphyletic. The Pisonieae are found to be sister to *Pisoniella*, which had been included in that tribe by Heimerl (1934) but was removed

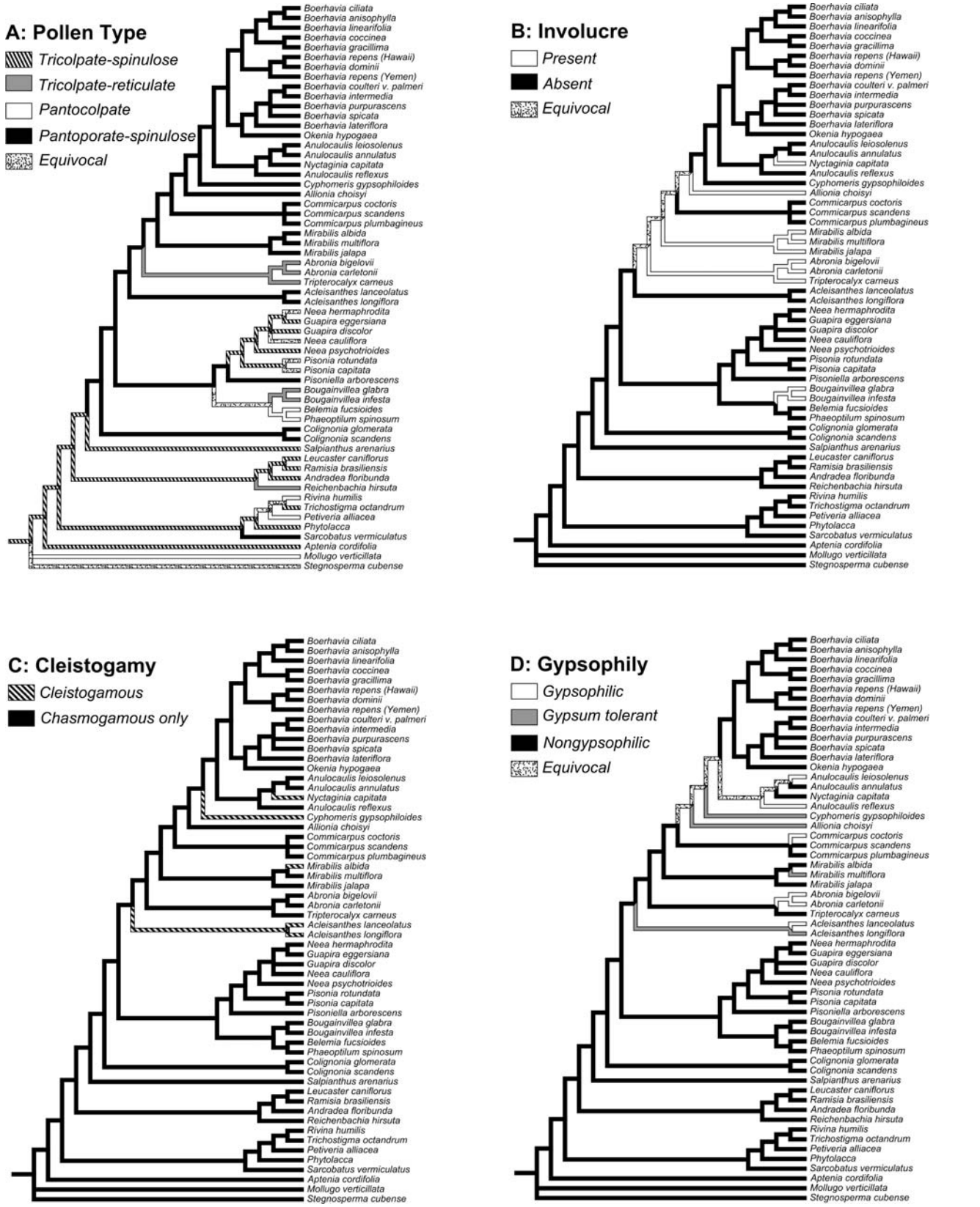

Fig. 4. Parsimony reconstruction of (A) pollen morphology, (B) involucre presence, (C) cleistogamous flowers, and (D) gypsophilic habit (based on ordered characters). See text for sources of characters used in reconstructions.

to subtribe Colignoniinae by Bittrich and Kühn (1993) following the suggestion of Bohlin (1988). The reasoning behind this move is mysterious, and in light of our results, it appears to have been unwarranted. *Pisoniella* possesses a straight embryo like other Pisonieae, and the large coriaceous anthocarps are provided with viscid glands along the ribs, much like those in *Pisonia* (Heimerl, 1934).

Within Pisonieae, *Neea* and *Guapira* form a clade (Fig. 2). These genera are distinguished primarily by whether the stamens are included (*Neea*) or exserted (*Guapira*). Our sampling is extremely limited in these two large genera, with only five accessions to represent ca. 150 species, though we were able to include accessions from geographically disparate locales. Neither genus forms a monophyletic group. This conclusion has been occasionally anticipated (e.g., Pool, 2001). It is unclear whether our sampling simply happened to include misclassified species in otherwise good genera, or whether this paraphyly is representative of *Neea* and *Guapira* generally. Much more intensive sampling is clearly needed to understand the relationships of the species in these genera, and it would be imprudent to attempt to reclassify them until a more detailed study is made including phylogenetic, morphological, and distributional data. Unfortunately, collections of these dioecious trees often do not include individuals of both sexes. Also, the tendency of many Pisonieae to oxidize when dried has left many descriptions lacking crucial information concerning the color of fruits. Therefore, the taxonomic literature is quite confused and species limits are known not much better than when Standley (1931a, p. 73) wrote that, "I know of few groups of plants in which specific differences are so unstable and so baffling ... particularly in *Neea*, *Torrubia* [= *Guapira*] and *Mirablis*, no single character seems to be constant." Finally, in this study we did not attempt to infer the ages of lineages, yet it appears that the branch lengths in the *Neea* + *Guapira* clade are comparatively short, especially considering that this clade can be expected to accommodate as many as 150 species (Fig. 2). A similar pattern has been noted in other radiations of neotropical trees, e.g., *Inga* (Fabaceae) (Richardson et al., 2001). If the pattern of relatively short branches inferred between species was upheld with the inclusion of a larger sample of taxa and more rapidly evolving markers, it would point to this clade as another example of rapid diversification in the neotropics.

Tribe Nyctagineae is broadly paraphyletic. As mentioned, *Pisoniella* and *Phaeoptilum* are not found in this study to be closest relatives of any other Nyctagineae. Based on pollen morphology, Bohlin (1988) has suggested that *Colignonia* (subtribe Colignoniinae) has affinities to the tribe Mirabileae of Heimerl (1934), which roughly corresponds to the tribe Nyctagineae and the NAX clade. *Colignonia* may in fact be sister to the NAX clade as suggested by the MP analysis or to the NAX + B&P clade as suggested by the ML analysis (Fig. 1). Tribe Nyctagineae also does not include *Abronia* or *Tripterocalyx* (tribe Abronieae, Fig. 1). There are certain characters of the Abronieae that are anomalous within the Nyctagineae (and the NAX clade) and that justified recognition at a higher taxonomic level, namely, tricolpate pollen and linear stigmas. The two genera in the tribe have long been thought to be a natural group and are often synonymized (Heimerl, 1934; Bittrich and Kühn, 1993), though most authors have maintained the two genera (Galloway, 1975; Spellenberg, 2003). The *Abronia* + *Tripterocalyx* clade is characterized by the combination of an umbellate inflorescence of salverform flowers with included stamens and style, an involucre, anthocarps with typically well-developed wings or lobes, and a mature embryo with a single cotyledon.

Anulocaulis and *Nyctaginia* are classified in different subtribes in the classification of Bittrich and Kühn (1993), presumably based on the presence of an involucre in *Nyctaginia*. Both genera are succulent perennial herbs, and the turbinate fruits with umbonate apices of *Nyctaginia capitata* strongly resemble those of *Anulocaulis eriosolenus*. They differ in many characters, including flower color (red-orange in *Nyctaginia* vs. white to pink in *Anulocaulis*) and flowering time (flowers of *Nyctaginia* are open during the day, while in *Anulocaulis* anthesis is at sunset or later and flowers wilt in the morning). While the full matrix ML tree (Fig. 1) indicates that *Anulocaulis* may not be monophyletic, this relationship is poorly supported (MPBS/MLBS/PP = 64/55/63). In the restricted MP and ML analysis (Fig. 3), however, a monophyletic *Anulocaulis* is more strongly supported (MPBS/MLBS = 72/89). Therefore, we see no compelling reason to question the taxonomic status of *Anulocaulis*.

Anulocaulis + *Nyctaginia* are sister to a strongly supported clade containing *Boerhavia* and *Okenia*. Like the previous instance, *Okenia* resolves within *Boerhavia* in the full matrix ML topology (Fig. 1), but support for this relationship is only moderately significant in the Bayesian analysis of the full data set (PP = 97) and weakly supported by MPBS and MLBS (67/69). Conversely, *Boerhavia* is strongly supported as a monophyletic group in the MP and ML analyses of the restricted data set (MPBS/MLBS = 100/100, Fig. 3). Vegetatively, *Okenia* strongly resembles most *Boerhavia* in its decumbent habit, and subequal opposite leaves with sinuate or undulate margins. The flowers of *Okenia*, though larger, are similar in color to some perennial *Boerhavia* from the Chihuahuan Desert. Finally, *Okenia* is annual, a condition found in one clade of *Boerhavia*. However, *Okenia* is strikingly different than *Boerhavia* in its unique reproductive biology: it produces aerial flowers, but the large, spongy fruits are geocarpic, with peduncles elongating greatly after fertilization and the fruits maturing several centimeters belowground. The relationship between these two genera is deserving of more study.

Biogeographical patterns—The basal lineages of Nyctaginaceae (Boldoeae, Leucastereae, *Colignonia*, Bougainvilleeae, and Pisonieae [including *Pisoniella*]) are fundamentally South American. Though some taxa have representatives or populations in (sub)tropical North America, (*Salpianthus*, *Neea*, *Guapira*, *Pisonia*, *Pisoniella*), their distributions all include the neotropics, and phylogenetically they are interspersed with neotropical endemics. The widespread tropical genus *Pisonia* possesses extremely viscid anthocarps, which aid dispersal, frequently by seabirds (Burger, 2005). The sole genus not native to the Americas is *Phaeoptilum*, endemic to arid southwestern Africa. This monospecific genus is closely related to *Belemia* and *Bougainvillea*, both from eastern and southern South America. *Phaeoptilum* is morphologically quite distinct from its sister taxon *Belemia*, though vegetatively it resembles the xeric-adapted *Bougainvillea spinosa*. The early Cretaceous date (130–90 Ma) for the opening of the south Atlantic (Smith et al., 1994) makes vicariance an unlikely explanation for this disjunction. Dispersal seems more likely, and while there is no specialized dispersal structure on the anthocarp of *Belemia*, both *Bougainvillea* and *Phaeoptilum*

have compelling (albeit different) adaptations for wind dispersal. *Phaeoptilum* produces winged anthocarps highly similar to those found in *Tripterocalyx* and some species of *Acleisanthes*. In *Bougainvillea*, most species display three showy bracts, each fused to a solitary flower. In fruit each involucral bract remains fused to a fruit and acts as a wing, the structure functioning as a unit of dispersal (Ridley, 1930).

The North American Xerophytic Clade has diversified in the deserts of the southwestern United States and northwestern Mexico. Every genus is confined to or has representatives in this region. Widespread taxa in this clade, namely *Commicarpus* and *Boerhavia*, possess glandular fruits, which have most likely aided bird-dispersal in a manner similar to that of *Pisonia*. Two red-flowered *Boerhavia*, *B. coccinea* and the similar *B. diffusa* are widespread in most tropical and subtropical areas. *Boerhavia diffusa* appears to have naturally dispersed from the Americas, though the confused taxonomy of this species and *B. coccinea* in regional floras makes this difficult to evaluate, and both of these species are frequently transported by human activity. The "repens" complex in *Boerhavia* (*B. repens* and related species) is widespread in coastal habitats throughout the tropical Pacific and Indian oceans to the Arabian Peninsula, along with *B. dominii* from Australia. Like the red-flowered perennial *Boerhavia* mentioned, these species also have viscid glandular anthocarps. *Okenia* is found in deep sand dune habitat along the Pacific and Caribbean coasts of Mexico and Central America, with a disjunct population in southern Florida. Other authors (Heimerl, 1934; Fowler and Turner, 1977; Thulin, 1994; Levin, 2002; Spellenberg and Poole, 2003) have discussed the remarkable disjunctions of *Acleisanthes somaliensis* and *Mirabilis himalaicus* from east Africa and southern Asia, respectively. These appear to be attributable to long-distance dispersal events, due to their derived position within otherwise exclusively American clades (Levin, 2000; N. Douglas, unpublished data).

Pollen and involucre evolution—Tribal and subtribal classifications (Table 1) of the Nyctaginaceae have relied heavily on a few characters, such as pollen morphology and the development of an involucre. However, divisions based on these characters are not supported by our results because these characters have a high degree of homoplasy among genera.

Parsimony reconstruction of pollen type across Nyctaginaceae (Fig. 4a) shows that substantial homoplasy exists (11 changes), involving three of the four types diagnosed by Nowicke (Nowicke, 1970, 1975; Nowicke and Luikart, 1971). Pantocolpate grains may constitute a synapomorphy for *Belemia* + *Phaeoptilum*. It has been noted that large, desiccation-resistant, pantoporate pollen grains, equipped with pore plates, were found primarily in the herbaceous desert taxa (Nowicke and Luikart, 1971). Specific correlations between large and/or polyaperturate grains and habitat in angiosperms have not been adequately investigated. In a study of ecological correlates of pollen morphology in a wide selection of angiosperms (Lee, 1978), there was an extremely weak correlation of pore number with width and with "temperature." According to our reconstructions, the origin of pantoporate-spinulose pollen predates the major radiation of desert taxa in the NAX clade. However, *Colignonia* and *Pisoniella* have much smaller grains than do the remaining taxa with pantoporate-spinulose pollen (*Colignonia* = 25–35 μM, *Pisoniella* = 30–37 μM, Nowicke and Luikart, 1971; N. Douglas, unpublished data). Therefore, it would seem best to consider grain size as a variable separate from grain shape and exine structure.

Within Nyctagineae, the subtribes Nyctagininae and Boerhaviinae were separated by the presence or absence of an involucre subtending the inflorescence. In subtribe Nyctagininae, the involucre of *Mirabilis* is comprised of fused bracts; the remaining genera possess involucres of distinct bracts. The involucre in *Bougainvillea* is distinctive; fruits of *Bougainvillea* retain a large involucral bract as discussed. Involucres have no known dispersal function in any of the other taxa; they likely serve merely to protect the flower buds and developing fruits or discourage nectar-robbing insects (Cruden, 1970). Parsimony reconstruction of this character on the molecular topology (Fig. 4b) indicates that, for involucres, there are at least five gain/loss steps in the family, four in the NAX clade, which contains the members of the Nyctagineae-Nyctagininae, Nyctagineae-Boerhaviinae, and Abronieae, reflecting the artificial nature of this classification. In this analysis, the character was treated in a very simplistic fashion, reflecting nothing more than taxonomic convention. Comparative developmental studies may shed light on deeper homologies or convergences, especially as they relate to the subtending bracts found in many genera. The selective benefits involved in the expression of this structure could be revealed by appropriate ecological investigations.

Self-compatibility and cleistogamy—The production of obligately selfing flowers is obviously contingent on the ability of plants to self-pollinate and produce fertile progeny. Our incomplete knowledge of reproductive systems in Nyctaginaceae means that an unambiguous reconstruction of self-compatibility is not currently possible. However, several studies have addressed mating systems in select Nyctaginaceae: sporophytic self-incompatibility (SI) is known in *Bougainvillea* (Zadoo et al., 1975; López and Galetto, 2002). Some *Mirabilis* (sect. *Quamoclidion*) and *Abronia macrocarpa* fail to set seed when self-pollinated (Cruden, 1973; Williamson et al., 1994), but the basis for incompatibility is not known in these genera. The Pisonieae are usually dioecious and are thus self-incompatible, although in these genera there are occasional monoecious or hermaphroditic species (e.g., *Pisonia brunoniana*) for which the mating system has not been studied (Sykes, 1987). Evidence suggests that many genera in the NAX clade are self-compatible: in addition to the production of cleistogamous flowers in four genera, *Boerhavia* and some *Mirabilis* are known to have a delayed self-pollination mechanism whereby the style curls and encounters the anthers as the flower wilts (Chaturvedi, 1989; Hernández, 1990; Spellenberg, 2000). Finally, flowers protected from pollinators have set viable seed in *Abronia umbellata* Lam. (McGlaughlin et al., 2002) and *Colignonia* (Bohlin, 1988).

Reasoning from these data, we can make certain inferences regarding the evolution of mating systems in Nyctaginaceae. Explanations for current distribution of mating systems family must incorporate one, or some combination of both, of the following scenarios. Which one is preferred depends on the likelihood of self-compatible lineages giving rise to lineages with an inability to self-fertilize, and the implications of either scenario are interesting.

One scenario, and the most parsimonious given our current knowledge, is that there have been at least three independent derivations of SI from a self-compatible ancestor. A single

change can account for the Pisonieae and *Bougainvillea*, one for the derived *Mirabilis* sect. *Quamoclidion*, and one for *Abronia macrocarpa*. It is often assumed that outcrossing species are not derived from selfing ancestors and that selfing lineages are an evolutionary "dead end" (Fisher, 1941; Stebbins, 1974; Lande and Schemske, 1985). In the case of Nyctaginaceae, however, the question is whether it is possible that *self-incompatible* species have arisen from *self-compatible* ancestors. It would seem that populations making this transition would be subject to most of the forces that affect the balance of selfing and outcrossing in self-compatible populations. A recent study of *s*-locus polymorphism in Solanaceae (Igic et al., 2006) has shown that losses of SI are irreversible in that family. The "cost" of developing the complex genetic systems necessary for SI would be added to the transmission advantage of alleles promoting self-fertilization (Uyenoyama et al., 1993); these factors must count against a hypothesis of multiple transitions to SI in one family.

Conversely, if we assume that SI is ancestral and has been lost repeatedly, transitions from SI to self-compatibility have occurred a minimum of six times (in *Colignonia*, *Acleisanthes*, some *Abronia*, two or more times in *Mirabilis*, and finally in the clade sister to *Mirabilis*). This represents a doubling of the number of evolutionary steps required to explain the distribution of known Nyctaginaceae mating systems. Other authors have discussed the merits of parsimony weighting schemes or maximum-likelihood approaches to testing the irreversibility of selfing (Barrett et al., 1996; Bena et al., 1998; Takebayashi and Morrell, 2001). In these cases, it may not be possible to escape a circular argument employing *only* phylogenetic evidence, because a weighting scheme favoring losses of SI assumes the conclusion. In Solanaceae (Igic et al., 2006), evidence of ancient polymorphism at the incompatibility locus itself was required to demonstrate the irreversibility of the loss of SI. In our case, the most convincing resolution will come when SI is characterized in *Mirabilis* sect. *Quamoclidion* and SI *Abronia*. If in these taxa and any others that may yet be discovered to be self-incompatible the genetic basis for SI can be identified, homology could be assessed and the ancestral functionality of the underlying mechanism could be tested.

Assuming the derived state is self-compatibility, of these six lineages, three have given rise to cleistogamous/chasmogamous lineages, and four gains of cleistogamy are required to explain the distribution of the character in Nyctaginaceae (Fig. 4c). Interestingly, the cleistogamous genera are all perennial, which should be less susceptible to selection pressure for reproductive assurance than annuals (Barrett et al., 1996). Alternatively, cleistogamous flowers can function to maximize seed set when resources, rather than pollinators, are limiting (Schemske, 1978). These hypotheses are both applicable to the cleistogamous Nyctaginaceae, though distinguishing between them may be difficult, because pollinators in desert environments tend to be scarce when water is scarce. Spellenberg and Delson (1974) found that *Acleisanthes* (*Ammocodon*) *chenopodioides*, with a generalized flower morphology and a diurnal pollinator fauna, produced roughly equal numbers of seeds from cleistogamous and chasmogamous flowers, and did not have a strong seasonal pattern in the production of cleistogamous flowers. In contrast, *Acleisanthes longiflora*, a species with large, specialized hawkmoth-pollinated flowers, produced the majority of a season's seeds from cleistogamous flowers produced preferentially in the dry early summer when sphingid moths are less active. This may suggest that cleistogamy in this genus is insurance against reproductive failure due to the absence of pollinators in some years.

Gypsophily—Parsimony reconstruction of gypsophily in Nyctaginaceae (Fig. 4d) indicates that gypsophiles and gypsum-tolerant species are widely dispersed in the NAX clade. With the current sampling, the ancestor of this clade is inferred to be nongypsophilic (whether or not the character is considered "ordered"), indicating that gypsum tolerance is derived multiple times. This conclusion is tenuous for two reasons. First, gypsum outcrops are common in the Chihuahuan Desert but less so in other parts of the ranges of the NAX genera. We are unable to rule out the possibility that taxa coded in this analysis as "nongypsophilic" are actually gypsum-tolerant, but simply do not occur in areas with gypsum soils.

Second, there are two *Mirabilis* [*M. nesomii* Turner and *M. linearis* (Pursh) Heimerl] which are gypsophilic (Turner, 1991) and gypsum-tolerant (R. Spellenberg, New Mexico State University, personal communication), respectively. These species, both in section *Oxybaphus*, are close relatives of the oxybaphoid *M. albida*, a nongypsophile included in this study. It is possible to add gypsophilic taxa as sisters to *M. albida* on our topology, so that the resolution of the ancestor of the NAX clade becomes equivocal, with ACCTRAN reconstruction as gypsum-tolerant, and DELTRAN as nongypsophilic. The same reconstruction would be made for the ancestors of *Commicarpus* and *Abronia* + *Tripterocalyx*. The sensitivity of the reconstruction at these key nodes to sampling artifacts indicates that in order to reconstruct the history of gypsophily in this clade, it will be necessary to undertake more intensive phylogenetic sampling at the species level, investigating an appropriate sample of nongypsophilic taxa closely related to known gypsophiles.

Even if we cannot know the gypsum tolerance of the ancestor of the NAX clade based on existing data, it is evident that there are at least four instances of strong gypsophily evolving in the family. It would be profitable to investigate the ecology of these gypsophytes and their relatives in the NAX clade. An experimental approach investigating whether or not seedlings of nongypsphiles have the latent ability to establish on gypseous crusts would disentangle the expression of gypsum tolerance from biogeographic complications, clarify the phylogenetic distribution of gypsum tolerance and perhaps reveal the nature of the adaptation(s) involved.

It is possible that establishment on gypsum is facilitated by some sort of modification to the radicle. Alternatively, because germination in a desert environment is always risky, adaptations to gypsum soils may differ little from germination strategies of desert taxa generally. Possible strategies could serve to optimize the timing of germination, minimize the risk of all seedlings perishing or increase the length of time a seedling has to establish itself. These could include high germination rate at low temperatures and various forms of bet-hedging, such as seed heteromorphism and variable seed dormancy (Escudero et al., 1997). The production of mucilage upon wetting by the seed coat presumably increases the local availability of water and upon drying, anchors the seed (Romao and Escudero, 2005). Some of these traits are known in Nyctaginaceae. For instance, production of mucilage by the anthocarp is common in both gypsophilic and nongypsophilic taxa in the NAX clade (Spellenberg, 2003), and fruit/seed heteromorphism is known in *Abronia* and *Tripterocalyx* (Wilson, 1974).

Understanding when in their history Nyctaginaceae became gypsum-tolerant will clarify whether homoplasy is best explained by answering the question "how do species become gypsum-tolerant?" or "why are certain species found only on gypsum?" If it turned out that gypsum tolerance was ancestral in the NAX clade, then experiments may reveal the reasons full gypsophiles do not occur on more typical soils.

The tendency of Nyctaginaceae to evolve cleistogamy and gypsophily has been shown to the extent that we have demonstrated that the high level of homoplasy for these traits is restricted to the NAX clade. In neither case are we able to conclusively identify the largest group capable of evolving the trait. Largely due to the phylogenetic position of *Acleisanthes* (with gypsophilic, cleistogamous species), we infer that it is possible that the ancestor of the entire NAX clade was predisposed to evolve these traits. In the case of cleistogamy, the topology indicates either that SI mechanisms develop easily in Nyctaginaceae, or that once self-compatibility emerges, there is a high chance of cleistogamy following. If the latter situation is correct, the explanation for the large number of cleistogamous species in the NAX clade must ultimately rely on explaining the frequent loss of SI, though the proximate cause is more likely related to resource or pollinator limitation in xeric environments. With gypsophily, it remains to be seen what trait(s) allow for tolerance of gypsum soils and when they evolved and what factors act exclude to gypsophiles from nongypsum soils.

The present study is the first to provide a comprehensive genus-level examination of the phylogeny of Nyctaginaceae. Though sampling of *Caribea*, *Cuscatlania*, *Cephalotomandra*, *Grajalesia*, and *Neeopsis* would be desirable, the current level of sampling is sufficient to draw several useful conclusions with bearing on future studies of the family. Aside from providing a framework for future taxonomic revisions, it raises interesting evolutionary questions regarding biogeography, reproductive biology, and edaphic endemism. To a degree, this work may be considered a case study into the practical issues that may arise in an investigation of tendencies in character evolution. New insights will be gained with a combination of phylogenetic work at finer taxonomic scales and experimental data to better understand the natural history of individual species, especially those in the xerophytic clade.

LITERATURE CITED

ALAIN, H. 1960. Novedades en la flora cubana, XIII. *Candollea* 17: 113.

ANTONOVICS, J. 1971. Heavy metal tolerance in plants. *Advances in Ecological Research* 7: 1–85.

BAKER, H. G. 1961. The adaptation of flowering plants to nocturnal and crepuscular pollinators. *Quarterly Review of Biology* 36: 64–73.

BALDWIN, B. G. 2005. Origin of the serpentine-endemic herb *Layia discoidea* from the widespread *Layia glandulosa* (Compositae). *Evolution* 59: 2473–2479.

BARRETT, S. C. H., L. D. HARDER, AND A. C. WORLEY. 1996. The comparative biology of pollination and mating in flowering plants. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 351: 1271–1280.

BEHNKE, H. D. 1997. Sarcobataceae—a new family of Caryophyllales. *Taxon* 46: 495–507.

BELL, C. D., AND M. J. DONOGHUE. 2003. Phylogeny and biogeography of Morinaceae (Dipsacales) based on nuclear and chloroplast DNA sequences. *Organisms Diversity & Evolution* 3: 227–237.

BENA, G., B. LEJEUNE, J.-M. PROSPERI, AND I. OLIVIERI. 1998. Molecular phylogenetic approach for studying life-history evolution: the ambiguous example of the genus *Medicago* L. *Proceedings of the Royal Society of London, B, Biological Sciences* 265: 1141–1151.

BITTRICH, V., AND U. KÜHN. 1993. Nyctaginaceae. *In* K. Kubitzki, J. G. Rohwer, and V. Bittrich [eds.], The families and genera of flowering plants, 473–486. Springer-Verlag, Berlin, Germany.

BOGLE, A. L. 1974. The genera of Nyctaginaceae in the southeastern United States. *Journal of the Arnold Arboretum* 55: 1–37.

BOHLIN, J. E. 1988. A monograph of the genus *Colignonia* (Nyctaginaceae). *Nordic Journal of Botany* 8: 231–252.

BREMER, B., K. BREMER, M. W. CHASE, J. L. REVEAL, D. E. SOLTIS, P. S. SOLTIS, P. F. STEVENS, A. A. ANDERBERG, M. F. FAY, P. GOLDBLATT, W. S. JUDD, M. KÄLLERSJÖ, J. KAREHED, K. A. KRON, J. LUNDBERG, D. L. NICKRENT, R. G. OLMSTEAD, B. OXELMAN, J. C. PIRES, J. E. RODMAN, P. J. RUDALL, V. SAVOLAINEN, K. J. SYTSMA, M. VAN DER BANK, K. WURDACK, J. Q. Y. XIANG, AND S. ZMARZTY. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.

BURGER, A. E. 2005. Dispersal and germination of seeds of *Pisonia grandis*, an Indo-Pacific tropical tree associated with insular seabird colonies. *Journal of Tropical Ecology* 21: 263–271.

CHATURVEDI, S. K. 1989. A new device of self pollination in *Boerhaavia diffusa* L. Nyctaginaceae. *Beitraege zur Biologie der Pflanzen* 64: 55–58.

CHAVEZ, R. P., R. F. NAVA, E. GRAFSTROM, M. VON PFALER, AND S. NILSSON. 1998. On the pollen of *Salpianthus* (Nyctaginaceae)—a morphological and image analysis approach. *Grana* 37: 352–357.

COCKERELL, T. D., AND F. GARCIA. 1898. Preliminary note on the growth of plants in gypsum. *Science* 8: 119–121.

CRUDEN, R. W. 1970. Hawkmoth pollination of *Mirabilis multiflora* Nyctaginaceae. *Bulletin of the Torrey Botanical Club* 97: 89–91.

CRUDEN, R. W. 1973. Reproductive biology of weedy and cultivated *Mirabilis* Nyctaginaceae. *American Journal of Botany* 60: 802–809.

CUENOUD, P., V. SAVOLAINEN, L. W. CHATROU, M. POWELL, R. J. GRAYER, AND M. W. CHASE. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany* 89: 132–144.

DARWIN, C. 1884. The different forms of flowers on plants of the same species. J. Murray, London, UK.

DESFEUX, C., S. MAURICE, J. P. HENRY, B. LEJEUNE, AND P. H. GOUYON. 1996. Evolution of reproductive systems in the genus *Silene*. *Proceedings of the Royal Society of London, B, Biological Sciences* 263: 409–414.

DOWNIE, S. R., D. S. KATZ-DOWNIE, AND K. J. CHO. 1997. Relationships in the Caryophyllales as suggested by phylogenetic analyses of partial chloroplast DNA ORF2280 homolog sequences. *American Journal of Botany* 84: 253–273.

DOWNIE, S. R., AND J. D. PALMER. 1994. A chloroplast DNA phylogeny of the Caryophyllales based on structural and inverted repeat restriction site variation. *Systematic Botany* 19: 236–252.

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.

ESCUDERO, A., L. F. CARNES, AND F. PEREZ-GARCIA. 1997. Seed germination of gypsophytes and gypsovags in semi-arid central Spain. *Journal of Arid Environments* 36: 487–497.

ESCUDERO, A., J. M. IRIONDO, J. M. OLANO, A. RUBIO, AND R. C. SOMOLINOS. 2000. Factors affecting establishment of a gypsophyte: the case of *Lepidium subulatum* (Brassicaceae). *American Journal of Botany* 87: 861–871.

ESCUDERO, A., R. C. SOMOLINOS, J. M. OLANO, AND A. RUBIO. 1999. Factors controlling the establishment of *Helianthemum squamatum*, an endemic gypsophile of semi-arid Spain. *Journal of Ecology* 87: 290–302.

FISHER, R. A. 1941. Average excess and average effect of a gene substitution. *Annals of Eugenics* 11: 53–63.

FOSBERG, F. R. 1978. Studies in the genus *Boerhavia* Nyctaginaceae, parts 1–5. *Smithsonian Contributions to Botany*: 1–20.

FOWLER, B. A., AND B. L. TURNER. 1977. Taxonomy of *Selinocarpus* and *Ammocodon* Nyctaginaceae. *Phytologia* 37: 177–208.

GALLOWAY, L. A. 1975. Systematics of the North American desert species of *Abronia* and *Tripterocalyx* Nyctaginaceae. *Brittonia* 27: 328–347.

GRANT, V. 1983. The systematic and geographical distribution of hawk moth flowers in the temperate North American flora. *Botanical Gazette* 144: 439–449.

GRANT, V., AND K. A. GRANT. 1983. Hawkmoth pollination of *Mirabilis longiflora* (Nyctaginaceae). *Proceedings of the National Academy of Sciences, USA, Biological Sciences* 80: 1298–1299.

HARRIMAN, N. A. 1999. Synopsis of New World *Commicarpus* (Nyctaginaceae). *SIDA Contributions to Botany* 18: 679–684.

HEIMERL, A. 1889. Nyctaginaceae. *In* A. Engler and K. Prantl [eds.], Die Natürlichen Pflanzenfamilien, 14–32. Engelmann, Leipzig, Germany.

HEIMERL, A. 1934. Nyctaginaceae. *In* A. Engler and K. Prantl [eds.], Die Natürlichen Pflanzenfamilien, 86–134. Engelmann, Leipzig, Germany.

HERNÁNDEZ, H. 1990. Autopolinización en *Mirabilis longiflora* L. (Nyctaginaceae). *Acta Botanica Mexicana* 12: 25–30.

HODGES, S. A. 1995. The influence of nectar production on hawkmoth behavior, self-pollination, and seed production in *Mirabilis multiflora* (Nyctaginaceae). *American Journal of Botany* 82: 197–204.

IGIC, B., L. BOHS, AND J. R. KOHN. 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proceedings of the National Academy of Sciences, USA* 103: 1359–1363.

JOHNSTON, I. M. 1941. Gypsophily among Mexican desert plants. *Journal of the Arnold Arboretum* 22: 145–170.

JORDAN, W. C., M. W. COURTNEY, AND J. E. NEIGEL. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *American Journal of Botany* 83: 430–439.

KEARNEY, T. H., AND R. H. PEEBLES. 1960. Nyctaginaceae. *In* T. H. Kearney and R. H. Peebles [eds.], Arizona flora, 270–279. University of California Press, Berkeley, California, USA.

KRUCKEBERG, A. R. 1986. The stimulus of unusual geologies for plant speciation—an essay. *Systematic Botany* 11: 455–463.

LANDE, R., AND D. SCHEMSKE. 1985. The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* 39: 24–40.

LEE, S. 1978. A factor analysis study of the functional significance of angiosperm pollen. *Systematic Botany* 3: 1–19.

LEVIN, R. A. 2000. Phylogenetic relationships within Nyctaginaceae tribe Nyctagineae: evidence from nuclear and chloroplast genomes. *Systematic Botany* 25: 738–750.

LEVIN, R. A. 2002. Taxonomic status of *Acleisanthes*, *Selinocarpus*, and *Ammocodon* (Nyctaginaceae). *Novon* 12: 58–63.

LEVIN, R. A., R. A. RAGUSO, AND L. A. MCDADE. 2001. Fragrance chemistry and pollinator affinities in Nyctaginaceae. *Phytochemistry* 58: 429–440.

LOOMIS, W. E. 1944. Effect of heavy applications of gypsum on plant growth. *Plant Physiology* 19: 706–708.

LÓPEZ, H. A., AND L. GALETTO. 2002. Flower structure and reproductive biology of *Bougainvillea stipitata* (Nyctaginaceae). *Plant Biology* 4: 508–514.

LORD, E. M. 1981. Cleistogamy: a tool for the study of floral morphogenesis, function, and evolution. *Botanical Review* 47: 421–449.

MADDISON, W. P., AND D. R. MADDISON. 2006. Mesquite: a modular system for evolutionary analysis, version 1.1. Website: http://mesquiteproject.org.

MAHRT, M., AND R. SPELLENBERG. 1995. Taxonomy of *Cyphomeris* (Nyctaginaceae) based on multivariate analysis of geographic variation. *SIDA Contributions to Botany* 16: 679–697.

MCCLELLAN, Y., AND W.-J. BOECKLEN. 1993. Plant mediation of ant–herbivore associations: the role of sticky rings formed by *Boerhavia spicata*. *Coenoses* 8: 15–20.

MCGLAUGHLIN, M., K. KAROLY, AND T. KAYE. 2002. Genetic variation and its relationship to population size in reintroduced populations of pink sand verbena, *Abronia umbellata* subsp. *breviflora* (Nyctaginaceae). *Conservation Genetics* 3: 411–420.

MEYER, S. E. 1986. The ecology of gypsophile endemism in the eastern Mojave Desert. *Ecology* 67: 1303–1313.

MEYER, S. E., E. GARCÍA-MOYA, AND L. D. LAGUNES-ESPINOZA. 1992. Topographic and soil surface effects on gypsophile plant community patterns in central Mexico. *Journal of Vegetation Science* 3: 429–438.

NOWICKE, J. 1968. Palynotaxonomic study of the Phytolaccaceae. *Annals of the Missouri Botanical Garden* 55: 294–364.

NOWICKE, J. W. 1970. Pollen morphology in the Nyctaginaceae. Part 1: Nyctagineae, Mirabileae. *Grana* 10: 79–88.

NOWICKE, J. W. 1975. Pollen morphology in the order Centrospermae. *Grana* 15: 51–78.

NOWICKE, J. W., AND T. J. LUIKART. 1971. Pollen morphology of the Nyctaginaceae. Part 2: Colignonieae, Boldoeae, and Leucastereae. *Grana* 11: 145–150.

NYLANDER, J. A. A. 2004. MrModeltest, version 2. Program distributed by the author, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden. Website: http://www.abc.se/~nylander/mrmodeltest2/mrmodeltest2.html.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics—an empirical approach using 3 molecular data sets in the Solanaceae. *Systematic Biology* 43: 467–481.

OXELMAN, B., M. LIDEN, AND D. BERGLUND. 1997. Chloroplast *rps*16 intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* 206: 393–410.

OYONARTE, C., G. SANCHEZ, M. URRESTARAZU, AND J. J. ALVARADO. 2002. A comparison of chemical properties between gypsophile and non-gypsophile plant rhizospheres. *Arid Land Research and Management* 16: 47–54.

PARSONS, R. F. 1976. Gypsophily in plants—a review. *American Midland Naturalist* 96: 1–20.

POOL, A. 2001. Nyctaginaceae. *In* W. D. Stevens, C. Ulloa U., A. Pool, and O. M. Montie [eds.], Flora de Nicaraugua, 1581–1592. Missouri Botanical Garden Press, St. Louis, Missouri, USA.

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

RAMBAUT, A. 1996. Se-Al: sequence alignment editor. Available at http://evolve.zoo.ox.ac.uk/.

RETTIG, J. H., H. D. WILSON, AND J. R. MANHART. 1992. Phylogeny of the Caryophyllales—gene sequence data. *Taxon* 41: 201–209.

REYES-SALAS, M., AND E. MARTÍNEZ-HERNÁNDEZ. 1982. Palynological catalog for the flora of Veracruz, Mexico. 8. Nyctaginaceae family. *Biotica (Mexico)* 7: 423–456.

RICHARDSON, J. E., R. T. PENNINGTON, T. D. PENNINGTON, AND P. M. HOLLINGSWORTH. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293: 2242–2245.

RIDLEY, H. N. 1930. The dispersal of plants throughout the world. L. Reeve & Co., Ashford, UK.

RODMAN, J. E., M. K. OLIVER, R. R. NAKAMURA, J. U. MCCLAMMER JR., AND A. H. BLEDSOE. 1984. A taxonomic analysis and revised classification of Centrospermae. *Systematic Botany* 9: 297–323.

ROHWEDER, O., AND K. HUBER. 1974. Centrospermen-Studien. 7. Beobachtungen und Anmerkungen zur Morphologie und Entwicklungsgeschichte einiger Nyctaginaceen. *Botanishe Jahrbücher* 94: 327–359.

ROHWER, J. G. 1993. Phytolaccaceae. *In* K. Kubitski, J. G. Rohwer, and V. Bittrich [eds.], The families and genera of flowering plants, 506–515. Springer-Verlag, Berlin, Germany.

ROMAO, R. L., AND A. ESCUDERO. 2005. Gypsum physical soil crusts and the existence of gypsophytes in semi-arid central Spain. *Plant Ecology* 181: 127–137.

RONQUIST, F., AND J. P. HUELSENBECK. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

SANDERSON, M. J. 1991. In search of homoplastic tendencies—statistical inference of topological patterns in homoplasy. *Evolution* 45: 351–358.

SCHEMSKE, D. 1978. Evolution of reproductive characters in *Impatiens* (Balsaminaceae): the significance of cleistogamy and chasmogamy. *Ecology* 59: 596–613.

SMITH, A. G., D. G. SMITH, AND B. M. FUNNELL. 1994. Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press, Cambridge, UK.

SPELLENBERG, R. 1993. Taxonomy of *Anulocaulis* (Nyctaginaceae). *SIDA Contributions to Botany* 15: 373–389.

SPELLENBERG, R. 2000. Blooming "behavior" in five species of *Boerhavia* (Nyctaginaceae). *SIDA Contributions to Botany* 19: 311–323.

SPELLENBERG, R. 2003. Nyctaginaceae. *In* F. o. N. A. E. Committee [ed.], Flora of North America north of Mexico, 14–74. Oxford University Press, New York, New York, USA.

SPELLENBERG, R., AND R. K. DELSON. 1974. Aspects of reproduction in Chihuahuan Desert Nyctaginaceae. Transactions of the Symposium on the Biological Resources of the Chihuahuan Desert Region, United States and Mexico, 273–287. Sul Ross State University, Alpine, Texas, USA.

SPELLENBERG, R., AND J. M. POOLE. 2003. Nomenclatural adjustments and comments in *Abronia* and *Acleisanthes* (Nyctaginaceae). *SIDA Contributions to Botany* 20: 885–889.

STANDLEY, P. C. 1909. The Allioniaceae of the United States with notes on Mexican species. *Contributions from the United States National Herbarium* 12: 303–389.

STANDLEY, P. C. 1911. The Allioniaceae of Mexico and Central America. *Contributions from the United States National Herbarium* 13: 377–430.

STANDLEY, P. C. 1918. Allioniaceae. *In* N. L. Britton [ed.], North American Flora, 171–254. New York Botanical Garden Press, Bronx, New York, USA.

STANDLEY, P. C. 1931a. Studies of American plants: Nyctaginaceae. *Field Museum Botanical Series* 8: 304–311.

STANDLEY, P. C. 1931b. The Nyctaginaceae of northwestern South America. *Field Museum Botanical Series* 11: 71–126.

STANFORD, A. M., R. HARDEN, AND C. R. PARKS. 2000. Phylogeny and biogeography of Juglans (Juglandaceae) based on *matK* and ITS sequence data. *American Journal of Botany* 87: 872–882.

STEBBINS, G. L. 1974. Flowering plants. Evolution above the species level. Belknap Press of Harvard University Press, Cambridge, Massachusetts, USA.

STEVENS, P. F. 2001-onwards. Nyctaginaceae. Angiosperm phylogeny website http://www.mobot.org/MOBOT/research/APweb/.

SWOFFORD, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, Massachusetts, USA.

SYKES, W. R. 1987. The parapara, *Pisonia brunoniana* (Nyctaginaceae). *New Zealand Journal of Botany* 25: 459–466.

TAKEBAYASHI, N., AND P. L. MORRELL. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *American Journal of Botany* 88: 1143–1150.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.

THULIN, M. 1994. Aspects of disjunct distributions and endemism in the arid parts of the Horn of Africa, particularly Somalia. Proceedings of the XIIIth plenary meeting AETFAT, 2: 1105–1119. National Herbarium and Botanic Gardens of Malawi, Zomba, Malawi.

TURNER, B. L. 1991. A new gypsophilic species of *Mirabilis* (Nyctaginaceae) from Nuevo León, Mexico. *Phytologia* 70: 44–46.

TURNER, B. L. 1993. A new species of *Anulocaulis* (Nyctaginaceae) from southern Coahuila, Mexico. *SIDA Contributions to Botany* 15: 613–615.

UYENOYAMA, M. K., K. E. HOLSINGER, AND D. M. WALLER. 1993. Ecological and genetic factors directing the evolution of self-fertilization. Oxford University Press, Oxford, UK.

WATERFALL, U. T. 1946. Observations on the gypsum flora of southwestern Texas and adjacent New Mexico. *American Midland Naturalist* 36: 456–466.

WHITE, T. J., T. BRUNS, S. LEE, AND J. TAYLOR. 1990. Amplifications and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* M. Innis, D. Gelfand, J. Sninsky, and T. White [eds.], PCR protocols: a guide to methods and applications, 315–322. Academic Press, San Diego, California, USA.

WILLIAMSON, P. S., L. MULIANI, AND G. K. JANSSEN. 1994. Pollination biology of *Abronia macrocarpa* (Nyctaginaceae), an endangered Texas species. *Southwestern Naturalist* 39: 336–341.

WILLSON, J., AND R. SPELLENBERG. 1977. Observations on anthocarp anatomy in the subtribe Mirabilinae (Nyctaginaceae). *Madrono* 24: 104–111.

WILSON, R. C. 1974. *Abronia.* Part 2. Anthocarp polymorphism and anatomy for 9 *Abronia* species found in California. *Aliso* 8: 113–128.

ZADOO, S. N., R. P. ROY, AND T. N. KHOSHOO. 1975. Cytogenetics of cultivated bougainvilleas. Part 2. Pollination mechanism and breeding system. *Proceedings of the Indian National Science Academy, Part B, Biological Sciences* 41: 498–502.

APPENDIX 1. Taxa, GenBank accession numbers, and voucher information used in this study. Regions not sampled are indicated by a dash. Cultivated plants were obtained from the following sources: DUBG = Duke University Botany Greenhouses, Durham; STRYB = Strybing Arboretum, San Francisco. Vouchers are deposited at the following herbaria: DUKE = Duke University, NMC = New Mexico State University, NY = New York Botanical Garden. † Accession included in "sensitivity" analysis. If sequences were downloaded from GenBank, then no voucher information is given.

Species—GenBank accession numbers: ITS, *ndhF*, *rpl16*, *rps16*, *Voucher specimen*, Locality, Year, Herbarium

Abronia bigelovii Heimerl†—EF079455, EF079510, EF079564, EF079606, *Douglas 2088,* New Mexico, USA, 2001, DUKE; ***Abronia carletonii*** J.M. Coult. & Fisher†—EF079456, EF079511, EF079565, EF079607, *Douglas 2091,* New Mexico, USA, 2001, DUKE; ***Acleisanthes lanceolatus*** (Wooton) R.A. Levin†—EF079454, EF079509, EF079563, EF079605, *Douglas 2072,* New Mexico, USA, 2001, DUKE; ***Acleisanthes longiflora*** A. Gray—EF079457, EF079512, —, EF079608, *Douglas 2098,* New Mexico, USA, 2001, DUKE; ***Allionia choisyi*** Standl.†—EF079467, EF079519, EF079574, EF079618, *Douglas 2187,* Coahuila, Mexico, 2002, DUKE; ***Andradea floribunda*** Allemão—EF079491, EF079545, —, EF079639, *Amorim 2294,* Brazil, 1998, NY; ***Anulocaulis annulatus*** (Coville) Standl.†—EF079503, EF079557, EF079599, EF079650, *Spellenberg 3162,* California, USA, 1993, NMC; ***Anulocaulis leiosolenus*** (Torr.) Standley ***v. leiosolenus*** Spellenberg—EF079464, EF079517, —, EF079615, *Douglas 2122,* Arizona, USA, 2002, DUKE; ***Anulocaulis reflexus*** I.M. Johnst.†— EF079468, EF079520, —, —, *Douglas 2192,* Chihuahua, Mexico, 2002, DUKE; ***Anulocaulis reflexus*** I.M. Johnst.†— —, —, EF079586, EF079629, *Spellenberg 10739,* Chihuahua, Mexico, 1990, NMC; ***Belemia fucsioides*** Pires—EF079488, EF079542, —, —, *Belem 3796,* Brazil, 1968, NY; ***Boerhavia anisophylla*** Torr.†— EF079469, EF079521, EF079575, EF079619, *Douglas 2194,* Durango, Mexico, 2002, DUKE; ***Boerhavia ciliata*** Brandegee— EF079465, —, EF079572, EF079616, *Douglas 2145,* Texas, USA, 2002, DUKE; ***Boerhavia coccinea*** Mill.†— EF079472, EF079525, EF079579, EF079622, *Spellenberg 13275,* Arizona, USA, 2001, DUKE; ***Boerhavia coulteri*** (S. Wats.) ***v. palmeri*** Spellenb.†— EF079471, EF079524, EF079578, EF079621, *Spellenberg 13273,* Arizona, USA, 2001, DUKE; ***Boerhavia dominii*** Meikle & Hewson†—EF079487, EF079540, EF079594, EF079638, *Smyth 42,* Australia, 1997, MO; ***Boerhavia gracillima*** Heimerl†—EF079479, EF079533, EF079587, EF079630, *Spellenberg 12447 ,* Texas, USA, 1997, NMC; ***Boerhavia intermedia*** M.E. Jones†—EF079474, EF079527, EF079581, EF079624, *Spellenberg 13279,* Arizona, USA, 2001, DUKE; ***Boerhavia lateriflora*** Standl.†—EF079466, EF079518, EF079573, EF079617, *Douglas 2161,* Sonora, Mexico, 2002, DUKE; ***Boerhavia linearifolia*** A. Gray†—EF079459, EF079514, EF079567, EF079610, *Douglas 2102,* New Mexico, USA, 2001, DUKE; ***Boerhavia purpurascens*** A. Gray†—EF079470, EF079523, EF079577, EF079620, *Spellenberg 13261,* Arizona, USA, 2001, DUKE;

Boerhavia repens L.†—EF079480, EF079534, EF079588, EF079631, *Spellenberg 7183,* Sana, Yemen , 1983, NMC; ***Boerhavia repens*** L. †—EF079477, EF079531, EF079584, EF079627, *Rose 2,* Oahu, Hawaii, USA, 2001, DUKE; ***Boerhavia spicata*** Choisy†—EF079473, EF079526, EF079580, EF079623, *Spellenberg 13276,* Arizona, USA, 2001, DUKE; ***Bougainvillea glabra*** Choisy†—EF079463, —, EF079571, EF079614, *Douglas 2121,* North Carolina, USA (DUBG), 2002, DUKE; ***Bougainvillea infesta*** Griseb.—EF079498, EF079551, —, EF079644, *Nee 51442,* Bolivia, 2000, NY; ***Caribea litoralis*** Alain— —, EF079530, —, —, *A. H. Liogier 7013,* Cuba, 1959, NY; ***Colignonia glomerata*** Griseb.—EF079495, EF079549, —, EF079642, *Nee 52523,* Bolivia, 2003, NY; ***Colignonia scandens*** Benth.†—EF079502, EF079556, EF079598, EF079648, *Grantham 63,* Lojas, Ecuador (STRYB), 2003, DUKE; ***Commicarpus coctoris*** N.A. Harriman†—EF079481, EF079535, EF079589, EF079632, *Spellenberg 12883,* Oaxaca, Mexico, 1998, NMC; ***Commicarpus plumbagineus*** (Cav.) Standl.†—EF079504, EF079558, EF079600, EF079651, *Spellenberg 7374,* Ta'izz, Yemen, 1983, NMC; ***Commicarpus scandens*** (L.) Standl.†—EF079482, EF079536, EF079590, EF079633, *Spellenberg 12887,* Puebla, Mexico, 1998, NMC; ***Cyphomeris gypsophiloides*** (M. Martens & Galeotti) Standl.†—EF079458, EF079513, EF079566, EF079609, *Douglas 2100,* New Mexico, USA, 2001, DUKE; ***Guapira discolor*** (Spreng.) Little†—EF079476, EF079529, EF079583, EF079626, *Spellenberg 13294,* Florida, USA, 2001, DUKE; ***Guapira eggersiana*** (Heimerl) Lundell—EF079496, EF079550, —, EF079643, *Mori 25542/40,* French Guiana, 2003, NY; ***Leucaster caniflorus*** (Mart.) Choisy— —, EF079541, —, —, *Pirani 3602,* Brazil, 1995, NY; ***Leucaster caniflorus*** (Mart.) Choisy—EF079497, —, —, —, *Hatschbach 50421,* Brazil, 1993, NY; ***Mirabilis albida*** (Walter) Heimerl†—EF079451, EF079506, EF079560, EF079602, *Douglas 2035,* Arizona, USA, 2001, DUKE; ***Mirabilis jalapa*** L.†—EF079461, EF079515, EF079569, EF079612, *Douglas 2119,* North Carolina, USA (DUBG), 2002, DUKE; ***Mirabilis multiflora*** (Torr.) A. Gray†—EF079452, EF079507, EF079561, EF079603, *Douglas 2037,* Arizona, USA, 2001, DUKE; ***Neea cauliflora*** Heimerl—EF079493, EF079547, —, —, *Schanke S15106,* Peru, 2002, NY; ***Neea hermaphrodita*** S. Moore—EF079489, EF079543, —, —, *Nee 51426,* Bolivia, 2000, NY; ***Neea psychotrioides*** Donn. Sm.†—EF079505, EF079559, EF079601, EF079652, *Wilbur 63654,* Heredia, Costa Rica, 1995, DUKE; ***Nyctaginia capitata*** Choisy†—EF079478, EF079532, EF079585, EF079628, *McIntosh 2049,* New Mexico, USA, 1992, NMC; ***Okenia hypogaea*** Schltdl. & Cham.†—EF079483, —, —, EF079634, *TR & RK Van Devender 92–1069,* Sonora, Mexico, 1992, NMC; ***Okenia hypogaea*** Schltdl. & Cham.†——, EF079522, EF079576, —, *Douglas 2206,* Veracruz, Mexico, 2002, DUKE; ***Phaeoptilum spinosum*** Radlk.†—EF079490, EF079544, —, —, *Seydel 4077,* Namibia, 1964, NY; ***Pisonia capitata*** (S. Watson) Standl.†—EF079484, EF079537, EF079591, EF079635, *AL Reina G. (2000—193),* Sonora, Mexico, 2000, NMC; ***Pisonia rotundata*** Griseb. †—EF079475, EF079528, EF079582, EF079625, *Spellenberg 13293,* Florida, USA, 2001, DUKE; ***Pisoniella arborescens*** (Lag. & Rodr.) Standl.†—EF079485, —, EF079592, EF079636, *LeDuc 231,* Oaxaca, Mexico, 1992, NMC; ***Pisoniella arborescens*** (Lag. & Rodr.) Standl. †— —, EF079539, —, —, *Anderson 13522,* Oaxaca, Mexico, 1988, NY; ***Ramisia brasiliensis*** Oliv.— EF079492, EF079546, —, EF079640, *Jardim 1507,* Brazil, 1998, NY; ***Reichenbachia hirsuta*** Spreng.†—EF079494, EF079548, EF079595, EF079641, *Nee 51972,* Bolivia, 2002, NY; ***Salpianthus arenarius*** Humb. & Bonpl.†—EF079486, EF079538, EF079593, EF079637, *Spellenberg 12903,* Michoacan, Mexico, 1999, NMC; ***Tripterocalyx carneus*** (Greene) L. A. Galloway†—EF079453, EF079508, EF079562, EF079604, *Douglas 2060,* New Mexico, USA, 2001, DUKE; ***Outgroups*: *Aptenia cordifolia*** (L. f.) Schwantes— —, AF194824, —, —, *;* ***Mollugo verticillata*** L.— —, AF194827, —, —, *;* ***Mollugo verticillata*** L.— —, —, —, EF079649, *Wilbur 77788,* North Carolina, USA, 2004, DUKE; ***Petiveria alliacea*** L.—EF079499, EF079552, —, —, *AL Reina G. 98–2048,* Sonora, Mexico, 1998, NY; ***Phytolacca americana*** Roxb.—EF079460, —, EF079568, EF079611, *Douglas 2118,* North Carolina, USA, 2002, DUKE; ***Phytolacca acinosa*** L— —, AF194828, —, —, *;* ***Rivina humilis*** L.†—EF079462, EF079516, EF079570, EF079613, *Douglas 2120,* North Carolina, USA (DUBG), 2002, DUKE; ***Sarcobatus vermiculatus*** (Hook.) Torr. †—EF079501, EF079555, EF079597, EF079647, *Spellenberg 13312,* Nevada, USA, 2002, DUKE; ***Stegnosperma cubense*** A. Rich.—EF079500, EF079554, EF079596, EF079646, *Salas–M. 2649,* Oaxaca, Mexico, 1999, NY; ***Trichostigma octandrum*** (L.) H. Walter— —, EF079553, —, EF079645, *Acevedo–Rodriguez 5447* Virgin Islands, USA 1993, NY.