EVALUATING THE MONOPHYLY AND BIOGEOGRAPHY OF *CRYPTANTHA* (BORAGINACEAE)

_____

A Thesis

Presented to the

Faculty of

San Diego State University

_____

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Biology

with a Concentration in

Evolutionary Biology

_____

by

Makenzie E. Mabry

Summer 2015

SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Makenzie E. Mabry:

Evaluating the Monophyly and Biogeography of *Cryptantha* (Boraginaceae)

Michael G. Simpson, Chair
Department of Biology

Marshal Hedin
Department of Biology

Robert Edwards
Department of Computer Science

April 20th, 2015
Approval Date

There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.

—*Charles Darwin*
The Origin of Species

ABSTRACT OF THE THESIS

Evaluating the Monophyly and Biogeography of *Cryptantha*
(Boraginaceae)
by
Makenzie E. Mabry
Master of Science in Biology with a Concentration in
Evolutionary Biology
San Diego State University, 2015

Cryptantha Lehmann ex G. Don, an herbaceous plant genus of the Boraginaceae family, is found in western North America and western South America, but not in the tropics between. This amphitropical distribution has long puzzled scientists. In a previous study, *Cryptantha* was found to be paraphyletic and was split into five genera, including a weakly supported, potentially non-monophyletic *Cryptantha*. In all subsequent studies of the Amsinckiinae, the subtribe to which *Cryptantha* belongs, interrelationships of *Cryptantha* are generally not well-supported and have a low sample size. Next generation sequencing methods, such as genome skimming, allow for the acquisition of significantly more data at relatively low costs. Use of the complete ribosomal cistron, nearly complete chloroplast genome, and twenty-three mitochondrial genes, as well as a greatly increased sample size, has allowed for inference of relationships within this complex with strong support. The occurrence of a non-monophyletic *Cryptantha* is confirmed, with two clades, termed here the Albidae Clade and the Maritimae Clade, strongly supported as independent of the remainder of the genus. From these phylogenetic analyses, assessment of classification, character evolution, and the phylogeographic history that elucidates the current amphitropical distribution of the group, is performed. Revealing the timing, direction, and number of times of dispersal between North and South America gives insight as to the origin of the great biodiversity of these regions.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

INTRODUCTION

Boraginaceae, the forget-me-not family, has been the focus of many recent phylogenetic studies (Hasenstab-Lehman and Simpson 2012; Nazaire and Hufford 2012; Cohen 2013; Weigend et al. 2013). This family of herbs, shrubs, and trees has been subject to differing circumscriptions over the years, being classified as one large family (Boraginaceae s.l, in the broad sense), with up to five subfamilies (Mabberley 2008, APGIII 2009), or treated more narrowly (Boraginaceae s.s., in the strict sense), with the subfamilies elevated to family status. In this study, I elect to treat the Boraginaceae as the latter (s.s.), based, e.g., on recent work by Cohen (2013) and Weigend et al. (2013).

Boraginaceae s.s. has numerous diagnostic characteristics, including: hirsute to hispid vestiture, a usually circinate scorpioid cyme inflorescence, mostly actinomorphic flowers, a strongly four-lobed ovary, and a fruit that is a schizocarp of nutlets. Within the Boraginaceae s.s., depending on the author and morphological characters used, there are from four to thirteen named tribes (Cohen 2013). Most recently, Cohen (2013) and Weigend et al. (2013) recognized five tribes in the family as defined here, based on their respective molecular phylogenetic studies. From these recent phylogenetic analyses, the genus *Cryptantha* has been consistently recovered in a well-supported clade containing the genera *Amsinckia, Cryptantha*, *Dasynotus*, *Eremocarya*, *Greeneocharis*, *Harpagonella*, *Johnstonella*, *Oncaglossum*, *Orecoarya*, *Pectocarya*, *Plagiobothrys*, and three North American species of *Cynoglossum* (Hasenstab-Lehman and Simpson 2012; Nazaire and Hufford 2012; Cohen 2013; Weigend et al. 2013). Given that this clade resides in the tribe Cynoglosseae (Cohen 2013; Weigend et al. 2013), the first available name to designate it at the rank of subtribe is Amsinckiinae Brand (1931). Thus, subtribe Amsinckiinae is used in this study to designate this clade.

Using one chloroplast and one nuclear marker in their study of the Amsinckiinae [their Cryptanthinae Brand, ined.], Hasenstab-Lehman and Simpson (2012) recovered

Cryptantha as polyphyletic and split it into five genera, the four resurrected genera *Eremocarya*, *Greeneocharis*, *Johnstonella*, and *Oreocarya*, plus a newly delimited *Cryptantha*, a classification accepted here (Figure 1). In the parsimony analysis presented by Hasenstab-Lehman and Simpson (2012), *Cryptantha* s.s. (in the strict sense) was recovered as a monophyletic group with weak support (BS=71). In their maximum likelihood and Bayesian trees, *Cryptantha* was found to be polyphyletic and split between two clades termed *Cryptantha* s.s. 1 and *Cryptantha* s.s. 2, but with weak support (Figure 1). In all recent studies of the Amsinckiinae, interrelationships of species within both clades of *Cryptantha* are generally poorly resolved (Hasenstab-Lehman and Simpson 2012; Cohen 2013; Weigend et al. 2013).

Previous to these recent molecular phylogenetic analyses, studies assessing interrelationships within *Cryptantha* used only morphological characteristics and phenetic assessments. In 1925, Johnston described 15 series of *Cryptantha* occurring in North America (Table 1). These series were circumscribed based on the number of nutlets per fruit (1-4), nutlet sculpturing (generally smooth or "rough," the latter having minute tubercles), and, if more than one nutlet, whether the nutlets are similar (homomorphic) or different in size and/or sculpturing (heteromorphic). Johnston's series *Angustifoliae*, *Circumscissae*, and *Maritimae* are partially or entirely comprised of the newly resurrected genera *Eremocarya*, *Greeneocharis*, and *Johnstonella* (Hasenstab-Lehman and Simpson 2012). The remaining series mostly comprise taxa from the genus *Cryptantha*, as treated here. In his classification, Johnston characterized series *Affines* as having a fruit with one or four smooth, asymmetrical nutlets. Series *Albidae*, containing only *Cryptantha albida* (Kunth) I.M. Johnston, is characterized by a fruit with four homomorphic nutlets that are dark and triangular-ovate in shape. *Ambiguae* is united by the presence of one to four smooth to papillate homomorphic nutlets per fruit. *Barbigerae* has a fruit with one to four homomorphic, dorsally convex nutlets that are laterally rounded or obtuse. *Flaccidae* is described as having one smooth, ovate nutlet per fruit. Series *Graciles*, containing only *Cryptantha gracilis* Osterhout, has one smooth, lanceolate nutlet per fruit. Series *Leiocarpae* is similar to the former in having smooth, homomorphic nutlets, but is differnt in having one to four smooth homomorphic nutlets per fruit. Series *Maritimae* is characterized as having one to four nutlets per fruit that are typically heteromorphic with the odd nutlet maturing larger than the three consimilar

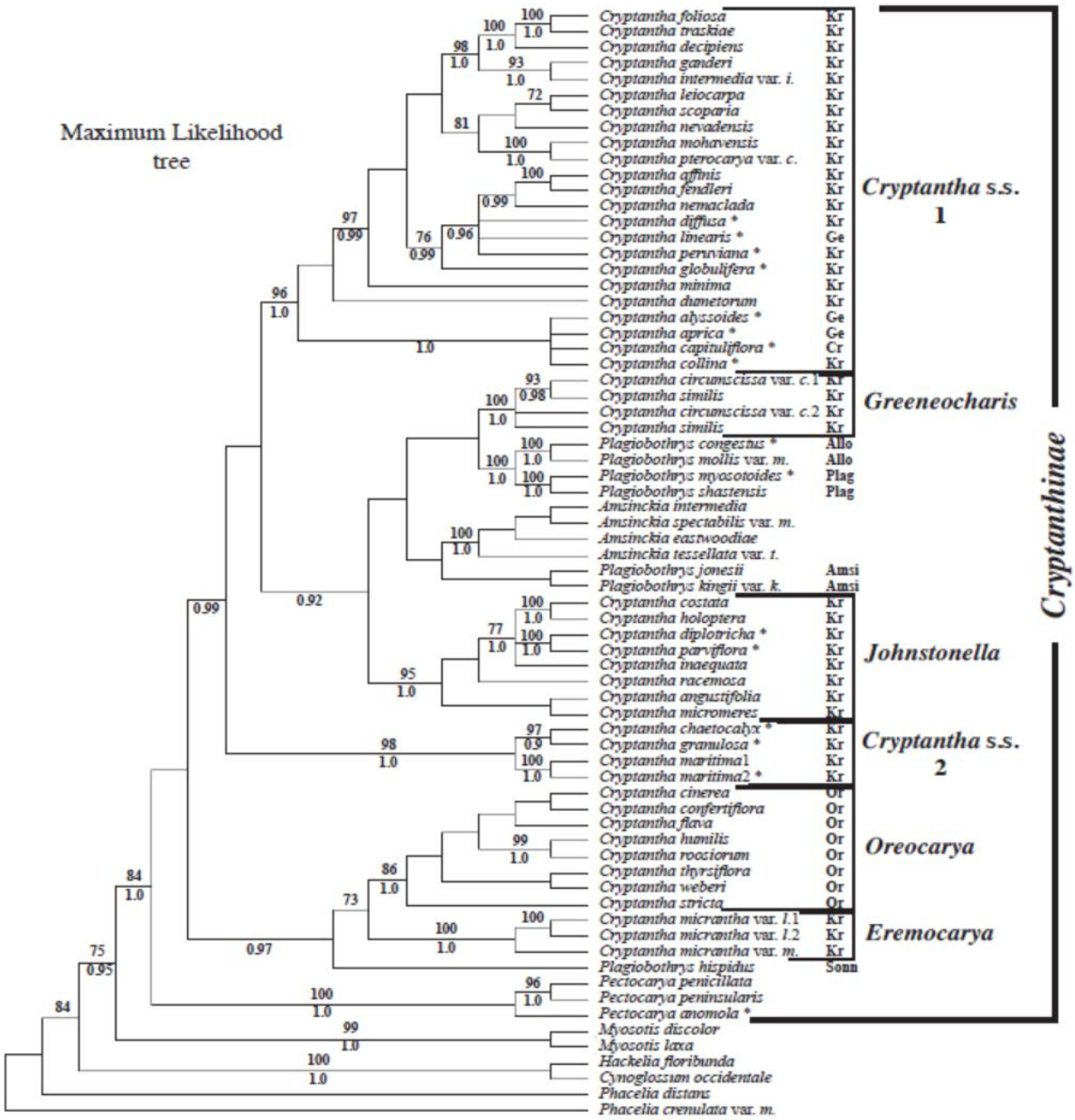

Figure 1. Maximum likelihood tree from Hasenstab-Lehman and Simpson (2012) showing the newly resurrected genera and the polyphyletic *Cryptantha* s.s.1 and *Cryptantha* s.s. 2 clades. ML bootstrap values shown above lineage and Bayesian Posterior probabilities show below. *Cryptantha* sections (Johnston 1927) abbreviations; Cr=*Cryptantha*; Ge=*Geocarya*; Kr=*Krynitzia*; Or=*Oreocarya*. *Plagiobothys* section abbrevions: Allo=*Allocarya*; Amsi=*Amsinckiopsis*; Plag=*Plagiobothrys*; Sonn=*Sonnea*. Note: Cryptanthinae is equivalent to Amsinckiinae.

Table 1. Johnston's (1925) series with described taxa

Johnston's series	Species	Current genus (Hasenstab-Lehman and Simpson 2012)
Affines	*C. affinis**	*Cryptantha s.s.*
	C. glomeriflora	*Cryptantha s.s.*
Albidae	*C. albida**	*Cryptantha s.s.*
Ambiguae	*C. ambigua**	*Cryptantha s.s.*
	C. crinita	*Cryptantha s.s.*
	*C. echinella**	*Cryptantha s.s.*
	C. excavata	*Cryptantha s.s.*
	C. hendersonii	*Cryptantha s.s.*
	*C. mariposae**	*Cryptantha s.s.*
	*C. simulans**	*Cryptantha s.s.*
	*C. torreyana**	*Cryptantha s.s.*
	C. traskiae	*Cryptantha s.s.*
Angustifoliae	*C. angelica*	***Johnstonella***
	*C. angustifolia**	***Johnstonella***
	C. costata	***Johnstonella***
	C. grayi	***Johnstonella***
	C. holoptera	***Johnstonella***
	C. inaequata	***Johnstonella***
	C. micrantha	***Eremocarya***
	C. pusilla	***Johnstonella***
	*C. racemosa**	***Johnstonella***
Barbigerae	*C. barbigera**	*Cryptantha s.s.*
	*C. decipens**	*Cryptantha s.s.*
	C. foliosa	*Cryptantha s.s.*
	*C. intermedia**	*Cryptantha s.s.*
	*C. nevadensis**	*Cryptantha s.s.*
	C. patula	*Cryptantha s.s.*
	*C. scoparia**	*Cryptantha s.s.*
Circumscissae	*C. circumscissa**	***Greeneocharis***
Flaccidae	*C. flaccida**	*Cryptantha s.s.*
	C. rostellata	*Cryptantha s.s.*
	*C. spariflora**	*Cryptantha s.s.*
Graciles	*C. gracilis**	*Cryptantha s.s.*
Leiocarpae	*C. abramsii =C. clevelandii var. clevelandii**	*Cryptantha s.s.*
	*C. brandegei=C. clevelandii var. clevelandii**	*Cryptantha s.s.*
	*C. clevelandii**	*Cryptantha s.s.*
	*C. hispidissima=C. clevelandii var. florosa**	*Cryptantha s.s.*
	*C. leiocarpa**	*Cryptantha s.s.*
	*C. microstachys**	*Cryptantha s.s.*
	*C. nemaclada**	*Cryptantha s.s.*
Maritimae	*C. dumetorum**	*Cryptantha s.s.*
	C. echinosepala	***Johnstonella***
	*C. martitma**	*Cryptantha s.s.*
	C. micromeres	***Johnstonella***

	*C. recurvata**	*Cryptantha s.s.*
Mohavenses	*C. mohavensis**	*Cryptantha s.s.*
	*C. watsonii**	*Cryptantha s.s.*
Muricatae	*C. muricata**	*Cryptantha s.s.*
Pterocaryae	*C. oxygona**	*Cryptantha s.s.*
	*C. pterocarya**	*Cryptantha s.s.*
	*C. utahensis**	*Cryptantha s.s.*
Ramulosissimae	*C. fendleri**	*Cryptantha s.s.*
Texanae	*C. crassisepala**	*Cryptantha s.s.*
	*C. kelseyana**	*Cryptantha s.s.*
	*C. minima**	*Cryptantha s.s.*
	C. pattersonii	*Cryptantha s.s.*
	*C. texana**	*Cryptantha s.s.*

Notes: Last column lists the genus that the species is currently recognized as (Hasenstab-Lehman and Simpson 2012). * Indicates submitted taxa. **Bold**= newly resurrected genera.

nutlets. Series *Mohavenses* has four smooth, lance-ovate or lanceolate homomorphic nutlets per fruit. Series *Muricatae* is only represented by one species, *Cryptantha muricata* (Hooker & Arnott) A. Nelson & J. F. Macbride, which has a fruit with four homomorphic nutlets that are coarsely tuberculate. Series *Pterocaryae* has a fruit with one to four rough, winged nutlets, which can be heteromorphic or homomorphic; if heteromorphic, the odd nutlet typically lacks a wing, having a thin margin. In series *Ramulosissimae*, containing only *Cryptantha fenderi* (A. Gray) Greene, the fruit has four smooth, homomorphic, lanceolate nutlets. Lastly, series *Texanae* has one to four heteromorphic nutlets per fruit; the odd nutlet in this series is typically larger and more roughened than the consimilar nutlets.

Johnston (1927) later studied the South American Boraginaceae, including the genus *Cryptantha*. In this work, he named three sections of South American *Cryptantha*: *Eucryptantha*, *Geocarya*, and *Krynitzkia* (Table 2). *Krynitzkia* is distinguished in having only chasmogamous (also termed "chasmogamic") flowers, which open to expose the sexual organs of the plant, potentially allowing for cross pollination. This section comprises all 55 North American, and most (24 of 44) South American *Cryptantha* species; two species, *Cryptantha albida* and *Cryptantha maritima* (Greene) Greene, are found in both North and South America. Members of the other two sections, in addition to forming typical chasmogamous flowers in the upper parts of the plant, develop cleistogamous (also termed "cleistogamic") flowers, in which the perianth does not open up and the pollen produced within that flower self-pollinates the ovary. One reason for this characteristic of plants is that they are neotenic, meaning when at maturity these clesitogamous flowers look like immature

Table 2. Johnston's (1927) South American Sections for Cryptantha s.s. taxa

Johnston's sections	Species
Eucryptantha (Cryptantha)	*C. alfalfalis**
	*C. calycotricha**
	*C. capituliflora**
	*C. glomerata**
	*C. glomerulifera**
	C. halpostachya
	C. longifolia
	C. spathulata
Geocarya	*C. alyssoides**
	C. aprica
	*C. cynoglossoides**
	C. dolichophylla
	C. dimorpha
	C. gayi
	C. involucrata
	*C. kingii**
	C. linearis
	C. volckmannii
Krynitzkia	*C. argentea*
	C. calycina
	C. chaetocalyx
	*C. diffusa**
	C. filaginea
	C. filiformis
	*C. globulifera**
	C. grandulosa
	C. limensis
	C. maritima
	C. patagonica
	*C. peruviana**
	C. romanii
	*C. subamplexicaulis**
	C. taltalensis
	*C. gnaphalioides**
	C. dichita
	*C. hispida**
	*C. phaceloides**

* Indicates submitted taxa.

chasmogamous flowers. This is caused by a reduced rate of development for the perianth, but not for the anthers and carpels, the sexual parts of the plant. Members of section *Eucryptantha*, comprising 10 species restricted to South America, bear cleistogamous flowers in leaf axils of the middle part of the plant and in the extreme lower portion of the upper inflorescence units; these cleistogamous flowers form fruits similar in morphology to those of the extreme upper chasmogamous ones. In section *Geocarya*, consisting of 12

species also restricted to South America, cleistogamous flowers similar to those of section *Eucryptantha* are produced. However, all members of section *Geocarya* develop more specialized cleistogamous flowers at the extreme base of the plant, these termed "cleistogenes" (Grau 1983). The fruits of these cleistogenes in *Geocarya* are different morphologically, being typically larger, reduced in number, and having a different sculpturing pattern from either the chasmogamous or cleistogamous flowers above (Johnston 1927; Grau 1983).

The distribution of *Cryptantha* species, restricted to the non-tropical regions of western North America and western South America (Figure 2), is found in several other plant groups. The cause of this "amphitropical" (or "amphitropic") distribution has long been debated by researchers (Raven 1963; Raven and Axelrod 1974; Moore et al. 2006). Possible explanations have included both vicariance and long-distance dispersal (Raven 1963; Raven and Axelrod 1974). The most recent accepted explanation for amphitropical distribution is via long-distance dispersal by migratory birds (Raven 1963; Moore et al. 2006). Hasenstab-Lehman and Simpson (2012) found that the distribution of the Amsinckiinae is best explained by several unidirectional dispersal events from North to South America. However, they had a limited sample size of South America taxa and recovered one incident of dispersal from South to North America in their *Cryptantha* s.s 1clade.

To better assess the phylogenetic history of *Cryptantha* species, a larger sample size and considerably more sequence data are necessary. Next generation sequencing genome skimming methods (Straub et al. 2011; Straub et al. 2012) allow for the acquisition of millions of base pairs. Genome skimming, also called shallow sequencing, can be used for obtaining near complete sequences of high copy regions, such as the chloroplast (cpDNA), mitochondria (mtDNA), and the ribosomal cistron (nrDNA) (Straub et al. 2011). This method of sampling of the genome has been shown to increase the resolution and support for phylogenetic hypotheses in plant groups (Straub et al. 2012). Work on the genus Oreocarya, a close relative of Cryptantha, has also proven this technique to be successful in greatly improving resolution in phylogenetic analyses (Ripma et al. 2014).

Figure 2. Distribution of *Cryptantha* showing the distributions in western North America and in western South America.

GOALS AND OBJECTIVES

The main goal of this study is to infer a well-supported phylogeny for the genus *Cryptantha*. This phylogeny will be used to address three major objectives. First, the monophyly of the genus and of the *Cryptantha* s.s. 1 and *Cryptantha* s.s. 2 clades recovered by Hasenstab-Lehman and Simpson (2012) will be tested, and phylogenetic interrelationships within *Cryptantha* will be inferred. Second, character evolution will be assessed for diagnostic morphological traits that Johnston used to describe his series and sections, including nutlet number, fruit heteromorphism, nutlet sculpturing, plant duration, evolution of cleistogamy, and stem vestiture. Third, biogeographic history will be assessed by inferring the number, timing, and direction of possible intercontinental dispersals.

MATERIALS AND METHODS

TAXON SAMPLING AND DNA ISOLATION

A total of 81 taxa were used for phylogenetic analyses, except for a coalescent species tree analysis in which the sample size was reduced to 50 (Table 3; see *Phylogenetic Analysis*). Samples of *Cryptantha* were obtained from both existing herbarium and recent field collections. For the latter, fresh leaf material was collected and dried in silica gel to prepare it for DNA extraction. All field collections have herbarium voucher specimens deposited at San Diego State University herbarium (SDSU). Duplicates of these collections, where available, are deposited at other accredited herbaria (SD, UCR).

To test the monophyly of *Cryptantha*, representatives of the closely related genera of subtribe Amsinckiinae were selected based on previous phylogenetic studies of the group (Hasenstab-Lehman and Simpson 2012; Cohen 2013; Weigend et al. 2013). Taxa include representatives of *Amsinckia*, North American *Cynoglossum, Dasynotus, Greeneocharis*, *Johnstonella*, *Oreocarya*, *Pectocarya*, and *Plagiobothys. Microula tibetica* Benth., found in the clade sister to the Amsinckiinae (Weigend et al. 2013) is used to root the tree.

Using leaf material, total genomic DNA was extracted and purified using a modified three-day version of the CTAB (cetyl trimethyl ammonium bromide) protocol (Doyle and Doyle 1987). RNaseA was added for degradation of single-stranded RNA for more efficient downstream analyses (Hasenstab-Lehman pers. comm.). Whole genomic DNA was quantified using NanoDrop spectroscopy (Thermo Fisher Scientific) and viewed for presence using gel electrophoresis, before sending out for library preparation.

DNA Sequencing and Quality Control

Whole genomic DNA was sent to Global Biologics (Columbia, Missouri, USA) for library preparation and barcoding for multiplexing to be used for Genome skimming methods (Straub et al. 2011; Straub et al. 2012). High throughput sequencing was performed on an

Table 3. Taxa Included for Phylogenetic Interference, Including Accession Number, Continent Locality, and Series/Section Placement by Johnston (1925, 1927, 1961)

Genus	Species	Variety	Location	Series	Section	Accession
Amsinckia	intermedia		North America			SDSU20756
Amsinckia	**tessellata**		North America			SDSU20350
Cryptantha	affinis		North America	Affines	Krynitzkia	SD199070
Cryptantha	**albida**		North America	Albidae	Krynitzkia	SDSU20612
Cryptantha	alfalfalis		South America	Glomeratar	Cryptantha	CONC163659
Cryptantha	alyssoides		South America	Alyssoides	Geocarya	CONC156553
Cryptantha	ambigua		North America	Ambiguae	Krynitzkia	SDSU20524
Cryptantha	**aspera**		South America		Cryptantha	MO4317599
Cryptantha	barbigera		North America	Barbigerae	Krynitzkia	SDSU20349
Cryptantha	**calycotricha**		South America	Halplostachyae	Cryptantha	CONC150898
Cryptantha	capituliflora		South America	Capituliflora	Cryptantha	CONC166914
Cryptantha	**clevelandii**	**var. florosa**	North America	Leiocarpae	Krynitzkia	RSA 710334
Cryptantha	clevelandii	var. florosa	North America	Leiocarpae	Krynitzkia	SDSU18342
Cryptantha	**clevelandii**	**var.clevelandii**	North America	Leiocarpae	Krynitzkia	SDSU20782
Cryptantha	**clokeyi**		North America	Muricatae	Krynitzkia	UCR164170
Cryptantha	**corollata**		North America	Barbigerae	Krynitzkia	SDSU20775
Cryptantha	crassisepala		North America	Texanae	Krynitzkia	SDSU20623
Cryptantha	crinita		North America		Krynitzkia	SDSU20823
Cryptantha	**cynoglossoides**		South America	Dimorphae	Geocarya	SI87776
Cryptantha	decipens		North America	Barbigerae	Krynitzkia	SDSU20014
Cryptantha	**diffusa**		South America	Barbigerae	Krynitzkia	MERL56799
Cryptantha	**dumetorum**		North America	Maritimae	Krynitzkia	SDSU 18694
Cryptantha	echinella		North America	Ambiguae	Krynitzkia	SDSU 19611
Cryptantha	fendleri		North America	Ramulosissimae	Krynitzkia	SDSU20114
Cryptantha	**flaccida**		North America	Flaccidae	Krynitzkia	SDSU19846
Cryptantha	ganderi		North America	Leiocarpae	Krynitzkia	SDSU20345
Cryptantha	globulifera		South America	Barbigerae	Krynitzkia	CONC163475
Cryptantha	**globulifera**		South America	Barbigerae	Krynitzkia	SGO147985
Cryptantha	globulifera		South America	Dimorphae	Geocarya	SGO146942
Cryptantha	globulifera		South America	Lineares	Geocarya	SGO147688
Cryptantha	**glomerata**	**var. glomerata**	South America	Glomeratae	Cryptantha	SGO146941
Cryptantha	glomeruifera		South America	Glomeruliferae	Cryptantha	CONC166867
Cryptantha	**gnaphalioides**		South America	Gnaphalioides	Krynitzkia	SGO146002
Cryptantha	**gracilis**		North America	Gracilis	Krynitzkia	UCR217631
Cryptantha	**hispida**		South America	Phaceloides	Krynitzkia	CONC150914
Cryptantha	**incana**		North America	Barbigerae	Krynitzkia	UCR227031
Cryptantha	**intermedia**	**var. intermedia**	North America	Barbigerae	Krynitzkia	SDSU20037
Cryptantha	kelseyana		North America	Texanae	Krynitzkia	SDSU20630
Cryptantha	**kingii**		South America	Virentes	Geocarya	SGO123832
Cryptantha	leiocarpa		North America	Leiocarpae	Krynitzkia	SDSU20759
Cryptantha	mariposae		North America	Ambiguae	Krynitzkia	SDSU20826
Cryptantha	**maritima**		North America	Maritimae	Krynitzkia	SDSU 20050
Cryptantha	**martirensis**		North America		Krynitzkia	SDSU18625
Cryptantha	**mexicana**		North America	Albidae	Krynitzkia	SDSU20610
Cryptantha	**microstachys**		North America	Leiocarpae	Krynitzkia	SD216851
Cryptantha	minima		North America	Texanae	Krynitzkia	SDSU20629
Cryptantha	mohavensis		North America	Mohavenses	Krynitzkia	SDSU20877
Cryptantha	**muricata**	**var. muricata**	North America	Muricatae	Krynitzkia	SDSU20749
Cryptantha	nemaclada		North America	Leiocarpae	Krynitzkia	SDSU20774
Cryptantha	**nevadensis**	**var. nevadensis**	North America	Barbigerae	Krynitzkia	SDSU20393
Cryptantha	**nevadensis**	**var. rigida**	North America	Barbigerae	Krynitzkia	SDSU20766
Cryptantha	oxygona		North America	Pterocaryae	Krynitzkia	RSA685321
Cryptantha	**peruviana**		South America	Barbigerae	Krynitzkia	SGO140959
Cryptantha	phaceloides		South America	Phaceloides	Krynitzkia	SGO146206
Cryptantha	**pterocarya**	**var.pterocarya**	North America	Pterocaryae	Krynitzkia	SDSU20355
Cryptantha	**recurvata**		North America	Maritimae	Krynitzkia	UCR225245
Cryptantha	scoparia		North America	Barbigerae	Krynitzkia	UCR211150
Cryptantha	simulans		North America	Ambiguae	Krynitzkia	SDSU20390

Cryptantha	sparsiflora	North America	Flaccidae	Krynitzkia	UCR184326
Cryptantha	**subamplexicaulis**	South America	Barbigerae	Krynitzkia	SGO129437
Cryptantha	**texana**	North America	Texanae	Krynitzkia	SDSU20611
Cryptantha	**torreyana**	North America	Ambiguae	Krynitzkia	SDSU20124
Cryptantha	**utahensis**	North America	Pterocaryae	Krynitzkia	SDSU20348
Cryptantha	**watsonii**	North America	Mohavenses	Krynitzkia	UCR226737
Cryptantha	wigginsii	North America	Leiocarpae	Krynitzkia	SDSU 20082
Cynoglossum	**grande**	North America			SDSU19197
Dasynotus	**daubenmirei**	North America			SDSU20343
Eremocarya	**micrantha**	North America	Angustifoliae	Krynitzkia	SDSU18956
Greeneocharis	**simulis**	North America			SDSU20605
Johnsontella	**angustifolia**	North America	Angustifoliae	Krynitzkia	RSA 731212
Johnsontella	**racemosa**	North America	Angustifoliae	Krynitzkia	SDSU 18710
Microula	**tibetica**	China			GH00466293
Oreocarya	**flavoculata**	North America			SDSU20030
Oreocarya	setosissma	North America			SDSU20242
Oreocarya	**virgata**	North America			SDSU20117
Pectocarya	**penicillata**	North America			UC1965571
Plagiobothrys	**fuluvs**	North America			
Plagiobothrys	**greenei**	North America			
Plagiobothrys	**hispidus**	North America			JEPS87508
Plagiobothrys	**jonesii**	North America			UCR215416
Plagiobothrys	**kingii**	North America			UC1876874

Note: **Bold** taxa were used for the reduced analyses.

Illumina HiSeq2000 (Illumina, San Diego, California, USA) at the Institute for Integrative Genome Biology (IIGB) Instrumentation Facilities at the University of California, Riverside or on an Illumina HiSeq2500 (Illumina, San Diego, California, USA) at Global Biologics (Columbia, Missouri, USA). Runs at both facilities yielded 101 base-pair single-end reads.

Quality control of reads was performed using PRINSEQ (Schmieder and Edwards 2011). Any read less than 50 base pairs in length with a mean quality Phred score below 30 and more than one N was removed. Both the 5' and 3' ends of reads were trimmed using a quality Phred score of 30 and a window size of 1. Lastly, all exact and reverse complement sequence duplicates were removed. Reads were then imported into the program Geneious (version 8.0, *Biomatters*) in FASTQ file format for all further analyses (Kearse et al. 2012). Geneious, a powerful research tool, is used extensively in the following assemblies using the protocol of Ripma et al. (2014).

PLASTOME ASSEMBLY AND MODEL SELECTION

De novo assemblies were done using Geneious, with default settings on the largest read pools to recover nearly complete plastomes (Ripma et al. 2014). The *de novo* assembly of *Cryptantha barbigera* (A. Gray) Greene produced a 125,000 bp partial plastome sequence. To ensure this sequence was cpDNA, the Find Annotations function in Geneious was used to

transfer annotations from the *Solanum lycopersium* L. (AM087200) chloroplast sequence from GenBank (Benson et al. 2005) with 50% or greater similarity. The newly annotated, partial plastome sequence of *C. barbigera* was then used as a reference for a reference guided assembly in Geneious, with default settings and 25 iterations (Ripma et al. 2014). A consensus contig was saved for each sample with a 75% threshold. Areas with no coverage were coded as a gap, and areas with less than 20x coverage were masked with an N (Ripma et al. 2014). Sequences were aligned using the MAFFT plugin (version 7.017, Misawa and Miyata 2002) with default settings and examined for misalignments by eye. If portions could not be realigned with confidence, they were excluded. After visual realignments, the Strip Alignments function in Geneious was used to remove any ambiguity codes. The AIC criteria (Aikaike 1974) in PartitionFinder (Lanfear et al. 2012), was used to find the best model of evolution for each codon position of the plastome (Table 4). Any codon position with the same model of evolution was then grouped into the same partition.

Table 4. Results for the Best Model of Evolution for Each Partition as Determined Using the AIC Criteria (Aikaike 1974) in PartitionFinder (Lanfear et al. 2012)

Regions	Partitions	Model of Evolution
nrDNA		
	ETS	TVM+I+G
	18S	K80
	ITS1, ITS2	TrNef+I+G
	5.8S, 26S	TrN+I+G
mtDNA		
	atp6, ccmC, cox2exon1, cox2exon2, nad1exon1, nad4exon1, nad5exon4, nad5exon5, nad9, orfBcodon1	TVM+I+G
	atp9, cob, cox3, nad2exon4, nad4L, nad4exon3, nad5exon2, nad6, nad7exon4, orf214	TVM+I+G
	nad1exon3, nad7exon3, orf142	HKY+I+G
cpDNA		
	cpDNACodon1, cpDNACodon2, cpDNACodon3	GTR+I+G

CISTRON ASSEMBLY AND MODEL SELECTION

Using the ITS sequence of *Cryptantha alyssoides* (D.C.) Reiche (JQ513396) from GenBank, a reference guided assembly was done using Geneious with default setting and 100 iterations. To assure that the whole cistron (ETS, 18S, ITS1, 5.8S, ITS2, and the 26S) had been captured through these iterations, the Transfer Annotations function from *Solanum lycopersium* (AM087200) with 50% or greater similarity was used. Once the complete

cistron was verified, it was used as a reference for a second reference guided assembly. Paralogs of the ITS regions that may be present due to incomplete homogenization were removed using a strict 75% matching consensus sequence requirement and removing any base pair position with an ambiguity code. Sequences were aligned using the MAFFT plugin with default settings and edited following the same protocol as described in the plastome assembly section above. To find the best model of evolution for the coding and non-coding regions of the cistron, the AIC criteria (Aikaike 1974) in PartitionFinder (Lanfear et al. 2012) was used (Table 4). Any region with the same model of evolution was then grouped into the same partition.

MITOCHONDRIAL GENE ASSEMBLY AND MODEL SELECTION

To assemble mitochondrial genes, a reference guided assembly using the *Nicotiana tabacum* L. (BA000042) mitochondrial sequence from GenBank was performed in Geneious. Resulting consensus contigs were annotated from the *Nicotiana tabacum* (BA000042) sequence and saved as a custom BLAST database. A file of mitochondrial genes extracted from *Nicotiana* (Ripma et al. 2014) was then used to perform a sequence search on the consensus contigs. Mitochondrial genes found in all taxa were aligned and edited using the protocol described above. The AIC criteria (Aikaike 1974) in PartitionFinder (Lanfear et al. 2012), was used to find the best model of evolution for each gene region (Table 4). Any gene with the same model of evolution was then grouped into one partition.

PHYLOGENETIC ANALYSIS

Maximum likelihood (ML) analyses were performed using RAxML (Stamatakis et al. 2008), implemented in Geneious for each of the three regions separately as well as concatenated. Regions were partitioned as stated above, and statistical support was assessed with 1,000 Bootstrap replicates using the GTR+I+G model of evolution.

Bayesian inference (BI) was also performed for each of the three regions separately and concatenated using BEAST (version 1.8.0, Drummond et al. 2012). For the separate analyses, each region was partitioned and run under the model of evolution as determined in PartitionFinder (Lanfear et al. 2012; Table 4). Analyses were run for 100 million generations and duplicated six times. The concatenated analysis was partitioned the same as in the ML

concatenated analysis using the GTR+I+G model of evolution and run for 250 million generations. Results were viewed in Tracer (Rambaut et al. 2014) to ensure convergence, then combined in LogCombiner (version 1.8.0, Drummond et al. 2012), annotated in TreeAnnotator (version 1.8.0, Drummond et al. 2012), and viewed in FigTree (Rambaut 2014).

Coalescent species tree estimates were performed in *BEAST (version 1.8.0, Drummond et al. 2012) on both the full dataset and a dataset with reduced (50) taxa for 250 million generations. The 50 taxa were selected to represent what are thought to be representatives of all major genera or clades. For both analyses, runs were duplicated six times. Results were viewed in Tracer (Rambaut et al. 2014) to ensure convergence, then combined in LogCombiner (version 1.8.0, Drummond et al. 2012), annotated in TreeAnnotator (version 1.8.0, Drummond et al. 2012), and viewed in FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Multi-species coalescence methods such as *BEAST co-estimate gene trees and the species tree, because of this they are computationally intensive and their application is hindered with large datasets (Liu et al. 2009). Due to the large sample size of this study, species tree estimates were also done using summary statistic coalescent methods, STAR (Liu et al. 2009) and ASTRAL (Mirarab et al. 2014). The three gene trees resulting from the ML analysis were used as input trees for these methods. For STAR (Liu et al. 2009), which requires rooted trees, *Microula tibetica* was designated as the outgroup.

CHARACTER EVOLUTION

Character evolution was assessed in Mesquite (Maddison and Maddison 2010), using maximum likelihood ancestral state reconstruction and the resulting concatenated maximum likelihood tree as input. The concatenated maximum likelihood tree was chosen as input for further analyses because it had the most nodes recovered with strong support (see *Results*). The MK1 probability model was chosen as best fit for the data considering that all characters had more than 2 states. Characters included were: 1) nutlet number per fruit: one, one to two, three to four, or four; 2) fruit heteromorphism: homomorphic (all nutlets similar), heteromorphic (at least one nutlet different), or both; 3) nutlet sculpturing: rough, smooth, or both; 4) plant duration: annual, perennial, or either; 5) cleistogamy: no cleistogamy

(chasmogamy), cleistogamy, or cleistogamy and cleistogenes; and 6) upper stem axis vestiture, specifically trichome orientation: spreading, appressed, or both.

BIOGEOGRAPHIC INFERENCE

Biogeographic analyses were performed using BioGeoBEARS (Matzke 2012, 2013) to determine patterns of dispersal. The program BioGeoBEARS evaluates phylogeography models used by the programs LAGRANGE (Ree and Smith 2008), DIVA (Ronquist 1997), and BAYAREA (Landis et al. 2013). It then provides a common statistical framework in order to judge which models are preferred for the input dataset. The concatenated ML tree of the complete dataset (81taxa) was used as the input tree file, and areas were set using the Global Ecological Zones published by the Forestry Department of the Food and Agriculture Organization of the United Nations (2001; Figure 3A, 3B). These Global Ecological Zones were described using the vegetation, climate and physiography of the world. Of the 22 defined zones, *Cryptantha* occurs in 11. In North America *Cryptantha* occurs in subtropical desert (SBWh), subtropical dry forest (SCs), subtropical mountain system (SM), subtropical steppe (SBSh), temperate desert (TeBWk), and temperate mountain system (TeM). In South America, *Cryptantha* occurs in subtropical dry forest (SCs), subtropical mountain system (SM), subtropical steppe (SBSh), tropical desert (TBWh), and tropical mountain system (TM) (Table 5). Species ranges within these zones were determined using herbarium records (CONC, LP, MO, SDSU, SGO) for South America and the Biota of North America Program (BONAP) for North America (Kartesz 2014). To limit computational load for analyses to run, North America subtropical dry forest and subtropical mountains zones were combined into one area (CA) and in South America, subtropical steppe and subtropical dry forest were combined (SBShCs). A total of nine areas were used, with any individual species occurring in up to a maximum of six areas (Table 5).

DIVERGENCE TIME ESTIMATION

Approximation of divergence times and divergence dates of major clades was performed in BEAST (version 1.8.0, Drummond et al. 2012). Both published rates of nucleotide substitutions and fossil records were used as calibrations for separate analyses. The average of the published rate of nucleotide substitution for angiosperm ITS data

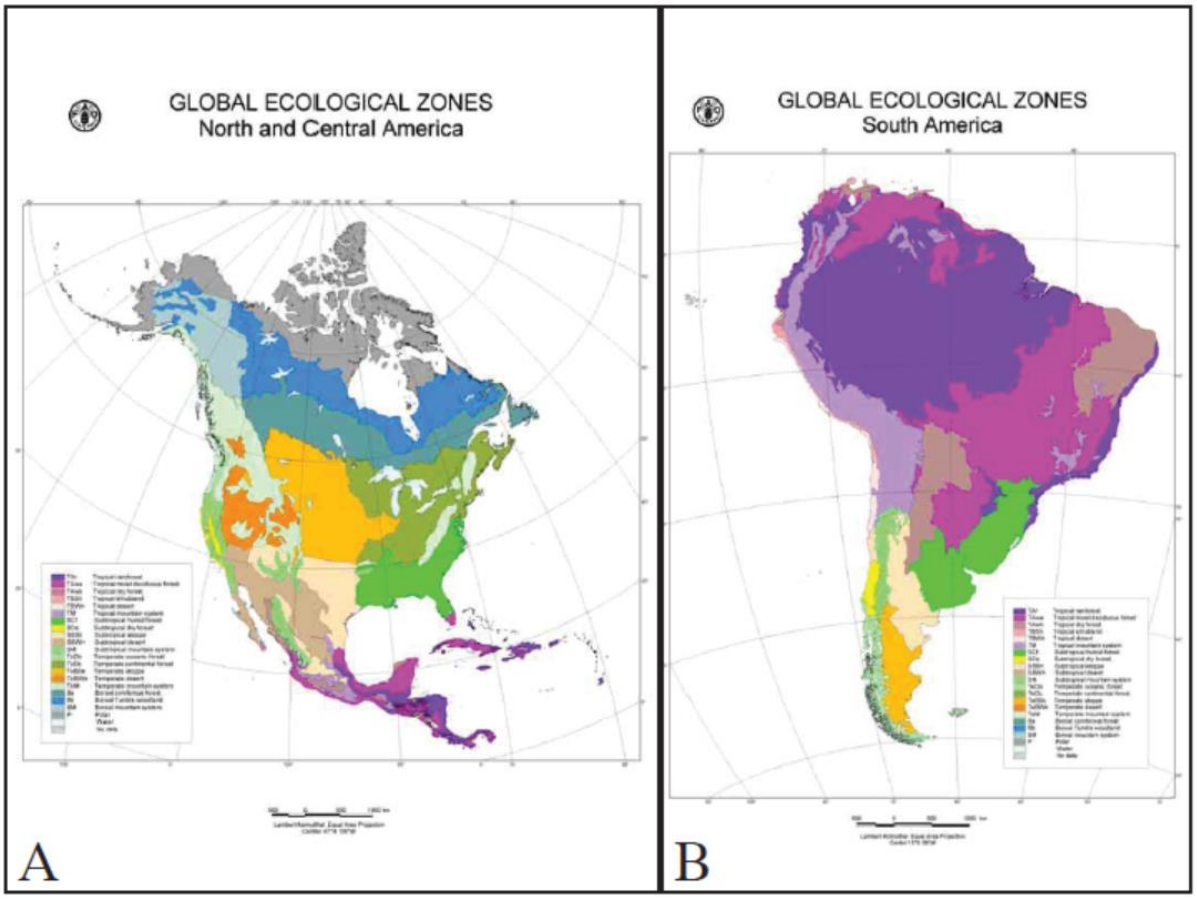

Figure 3. Global Ecological Zones of North and South America (Forestry Department of the Food and Agriculture Org. of the United Nations 2000) used for determining species boundaries for BioGeoBEARS (Matzke 2012, 2013). A. North America ranges: yellow and green= California region consisting of subtropical dry forest and the subtropical mountain system, brown= subtropical desert, peach= subtropical steppe, dark orange= temperate desert, seafoam green = temperate mountain system. B. South America ranges: purple = tropical mountain system, light peach = tropical desert, peach and yellow= subtropical steppe and dry forest, blue green = temperate oceanic forest.

(0.00413 substitutions/site/million years, Kay et al. 2006) was used as the rate of evolution for the ITS1 and ITS 2 partition with a normal distribution and a lognormal clock. Clocks for all other partitions were estimated also using a lognormal clock. A separate analysis utilized fossil *Cryptantha* taxa to constrain nodes. Three fossil *Cryptantha* relatives have been discovered*; Cryptantha auriculata* (M.K. Elias) Segal, *Cryptantha chaneyi* (M.K. Elias) Segal, and *Cryptantha coroniformis* (M.K. Elias) Segal (Elias 1942; Segal 1964, 1966; Figure 4). *Cryptantha chaneyi*, although it does not resemble any extant member of

Table 5. Species occurrences using the Global Ecological Zones (Forestry Depart. of the Food and Agriculture Org. of the U.N. 2000) for both North and South America.

Taxa	North America					South America			
Global Ecological Zones	CA	SBWh	SBSh	TeBWk	TeM	TM	TBWh	SBShCs	SM
Region Name in Tree (Fig.21; Fig. 22)	A	B	C	D	E	F	G	H	I
A_intermedia_SDSU20756_	X	X		X	X				
A_tessellata_SDSU20350_	X	X	X	X	X				
C_affinis_SD199070_	X			X	X				
C_albida_SDSU20612_	X	X	X			X			
C_alfalfalis_CONC163659_								X	
C_alyssoides_CONC156553_								X	
C_ambigua_SDSU20524_	X			X	X				
C_aspera_MO4317599_							X		
C_barbigera_SDSU20349_	X	X	X	X	X				
C_calycotricha_CONC150898_							X		
C_capituliflora_CONC166914_									X
C_clevelandii_RSA710334_	X								
C_clevelandii_SDSU18342_	X								
C_clevelandii_SDSU20782_	X								
C_clokeyi_UCR164170_	X								
C_corollata_SDSU20775_	X								
C_crassisepala_SDSU20623_	X	X	X	X	X				
C_crinita_SDSU2082_	X								
C_cynoglossoides_SI87776_									X
C_decipens_SDSU20014_	X	X	X	X					
C_diffusa_MERL56799_							X	X	X
C_dumetorum_SDSU18694_		X							
C_echinella_SDSU19611_	X			X	X				
C_fendleri_SDSU20114_	X	X	X	X	X				
C_flaccida_SDSU19846_	X				X				
C_ganderi_SDSU20345_	X	X							
C_globulifera_CONC163475_						X	X	X	X
C_globulifera_SGO147985_						X	X	X	X
C_glomerata_SGO146941_							X	X	X
C_glomerulifera_CONC166867_									X
C_gnaphalioides_SGO146002_							X	X	
C_gracilis_UCR217631_	X	X	X	X	X				
C_hispida_CONC150914_							X		
C_incana_UCR227031_	X								
C_intermedia_SDSU20037_	X	X		X	X				
C_involucrata_SGO146942_						X	X	X	X
C_kelseyana_SDSU20630_				X	X				
C_kingii_SGO123832_								X	
C_leiocarpa_SDSU20759_	X								
C_linearis_SGO147688_						X	X	X	X
C_mariposae_SDSU20826_	X								
C_maritima_SDSU20050_	X	X		X					X
C_martirensis_SDSU18625_	X								
C_mexicana_SDSU20610_		X	X						
C_microstachys_SD16851_	X								
C_minima_SDSU20629_	X	X	X		X				
C_mohavensis_SDSU_	X								
C_muricata_SDSU20749_	X	X							
C_nemaclada_SDSU20774_	X								
C_nevadensis_SDSU20393_	X	X	X	X	X				
C_nevadensisR_SDSU20766_	X		X	X	X				
C_oxygona_RSA685321_	X								
C_peruviana_SGO140959_						X			
C_phaceloides_SGO146206_						X			X
C_pterocarya_SDSU20355_	X	X	X	X	X				
C_recurvata_UCR225245_	X			X	X				

C_scoparia_UCR211150_	X				X	X					
C_simulans_SDSU20390_	X				X	X					
C_sparsiflora_UCR184326_	X										
C_subamplexicaule_SGO129437_								X			
C_texana_SDSU20611_			X								
C_torreyana_	X				X	X					
C_utahensis_SDSU20348_	X	X			X						
C_watsonii_UCR226737_	X				X	X					
C_wigginsii_SDSU20082_	X										
Cyno_gran_MGS_	X					X					
Dasynotus_daub_SDSU20343_					X	X					
E_micrantha_	X	X	X		X	X					
G_simulis_SDSU20605_	X										
J_angustifolia_RSA731212_	X	X	X		X	X					
J_racemosa_SDSU18710_	X	X	X		X	X					
Microula_tibetica_GH00466293_											
O_flavoculata_	X				X	X					
O_setosissima_		X	X		X	X					
Pec_penicillata_	X	X			X	X					
O_virgata_SDSU20117_						X					
Plagio_fulvus_	X					X					X
Plagio_greenei_	X										
Plagio_hispidus_JEPS87508_	X					X					
Plagio_jonesii_UCR215416_	X					X					

Notes: CA= subtropical dry forest and subtropical mountain system, SBWh= subtropical desert, SBSh= subtropical steppe, TeBWk= temperate desert, TeM= temperate mountain system. TM= tropical mountain system, TBWh= tropical desert, SBShCs= subtropical steppe and subtropical dry forest, SM= subtropical mountain system. A-I: corresponding regions in BioGeoBEARS (Matzke 2012, 2013).

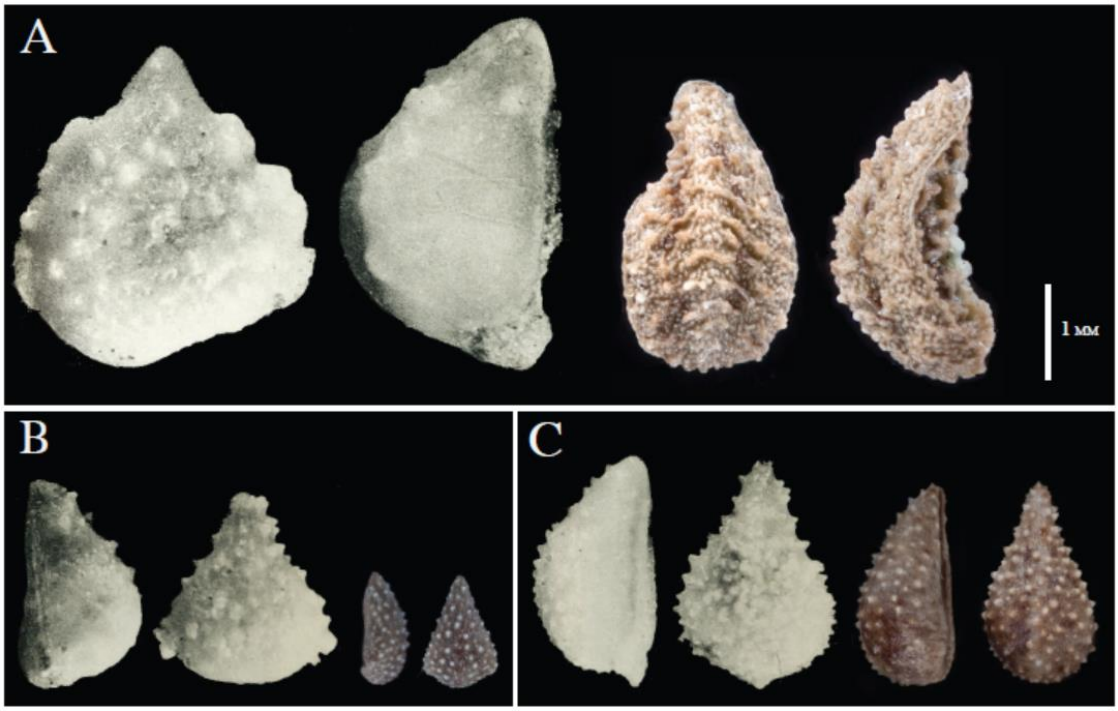

Figure 4. Comparison of Fossil Amsinckiinae used for calibration points and extant taxa. A= *Cryptantha chaneyi* (left) and *Oreocarya flavoculata* (right). B= *C. auriculata* (left) and *C. albida* (right). C= *C. coroniformis* (left) and *C. crassisepala* (right). All photos to scale, bars are 1mm.

Oreocarya, does have a large size and a triangular areola at the base of the attachment scar which then narrows into a groove that does not reach the apex of the nutlet body (Segal 1966; Figure 4A). This has been observed as a characteristic for the genus *Oreocarya* (Simpson and Hasenstab 2009) and therefore *C. chaneyi* was used to root the crown node of that clade. *Cryptantha auriculata* was used to root the base of the lineage containing *C. albida* (Kunth) I.M. Johnston, as it has similar morphological characters to *C. albida* with its triangular shaped nutlet (Segal 1966; Figure 4B). Lastly, *C. coroniformis* was used to root the crown node of the clade that contained the extant species *C. crassisepala* (Torrey & A. Gray) Greene and *C. minima* Rydberg as supported by several morphological similarities noted by Segal (1966; Figure 4C). Similarities include heteromorphism with regard to nutlet sculpturing, with one nutlet more or less smooth and the other(s) rough. All three fossil nutlets were all found in the Ogallala formation in Kansas, USA, in Ash Hollow Rock. Boellstorff (1976, 1978) dated this formation to be from the Hemphillian period (10.3-4.9 million years ago). BEAST (version 1.8.0, Drummond et al. 2012) runs conducted used a lognormal distribution with a mean of 10.3 million years ago (Ma), log standard deviation of 0.69, and an offset of 4.9 Ma. All analyses were run on the full dataset for 250 million generations.

RESULTS

SEQUENCE MATRICES

Genome skimming resulted in 81 individual read pools. *Oreocarya flavoculata* A. Nelson had the largest read pool of 7,593,640 reads. *Microula tibetica* resulted in the smallest read pool of just 820,347 reads. Although the latter read pool had significantly fewer reads, the plastome, complete cistron, and mitochondrial genes were all successfully recovered. *De novo* assembly of *Cryptantha barbigera* resulted in a 125,000 bp contig that was further used as a reference for assembly of all other cpDNA. After editing, an alignment of 119,580 bp was used for phylogenetic inference. A total of 14,728 variable and 6,964 parsimony informative characters were found. A complete cistron (5,638 bp) was recovered for all taxa. Non-coding regions contained most of the variability; however, coding regions did contribute to the total of 498 variable characters, of which 304 were parsimony informative. Lastly, the mitochondria assembly resulted in the recovery of 38 genes. Of those 38 genes, 23 of them were complete in all taxa and used for phylogenetic inference. These genes ranged from 100 bp to over 1,000 bp in length. Concatenation of the 23 genes resulted in a 9,685 bp alignment with 1,888 variable, and 1,038 parsimony informative characters.

PHYLOGENETIC ANALYSIS

Maximum likelihood (ML) and Bayesian inference (BI) of the chloroplast (cpDNA) resulted in trees with the exact same topology (Figure 5; Figure 6). In both analyses, three separate monophyletic groups of *Cryptantha* taxa are recovered. One monophyletic group consisting of *C. maritima* (Greene) Greene, *C. martirensis* M.G. Simpson & Rebman, *C. clokeyi* I.M. Johnston, and the South American species *C. subamplexicaulis* (Philippi) I.M. Johnston (referred to as the Maritimae Clade) is recovered with strong support (BS=100, PP=1). A second clade containing *C. albida*, *C. mexicana* I.M. Johnston, *C. texana* Greene, and the South American species *C. hispida* (Philippi) Reiche is found with strong support (BS=100, PP=1) and as sister to the genus *Johnstonella* (BS =100, PP=1). This group will be

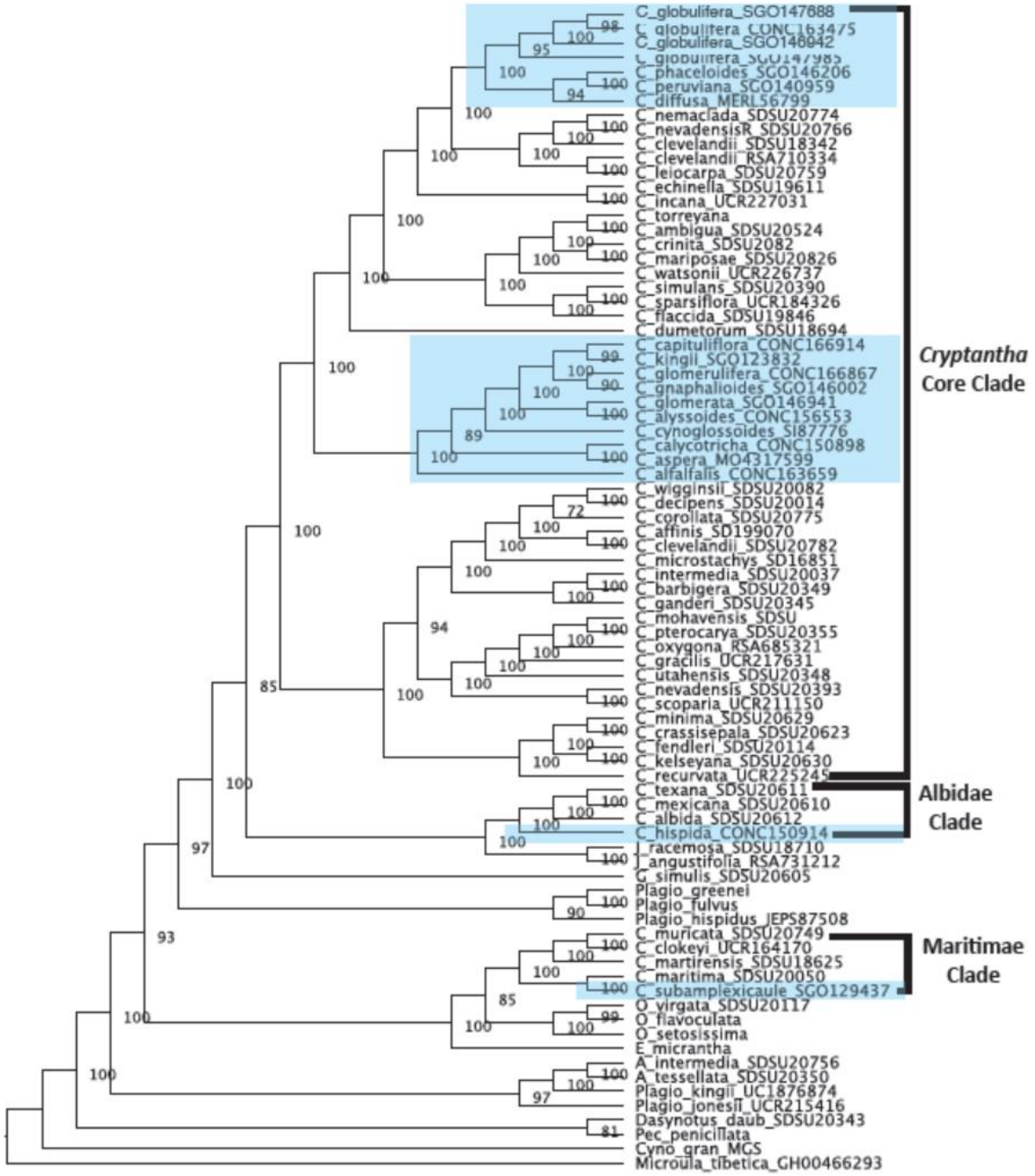

Figure 5. Maximum likelihood tree of the chloroplast (cpDNA). Major clades are identified and South American species are highlighted in blue. A=*Amsinckia*, C=*Cryptantha*, E=*Eremocarya* , G= *Greeneocharis*, J= *Johnstonella*, O=*Oreocarya.*

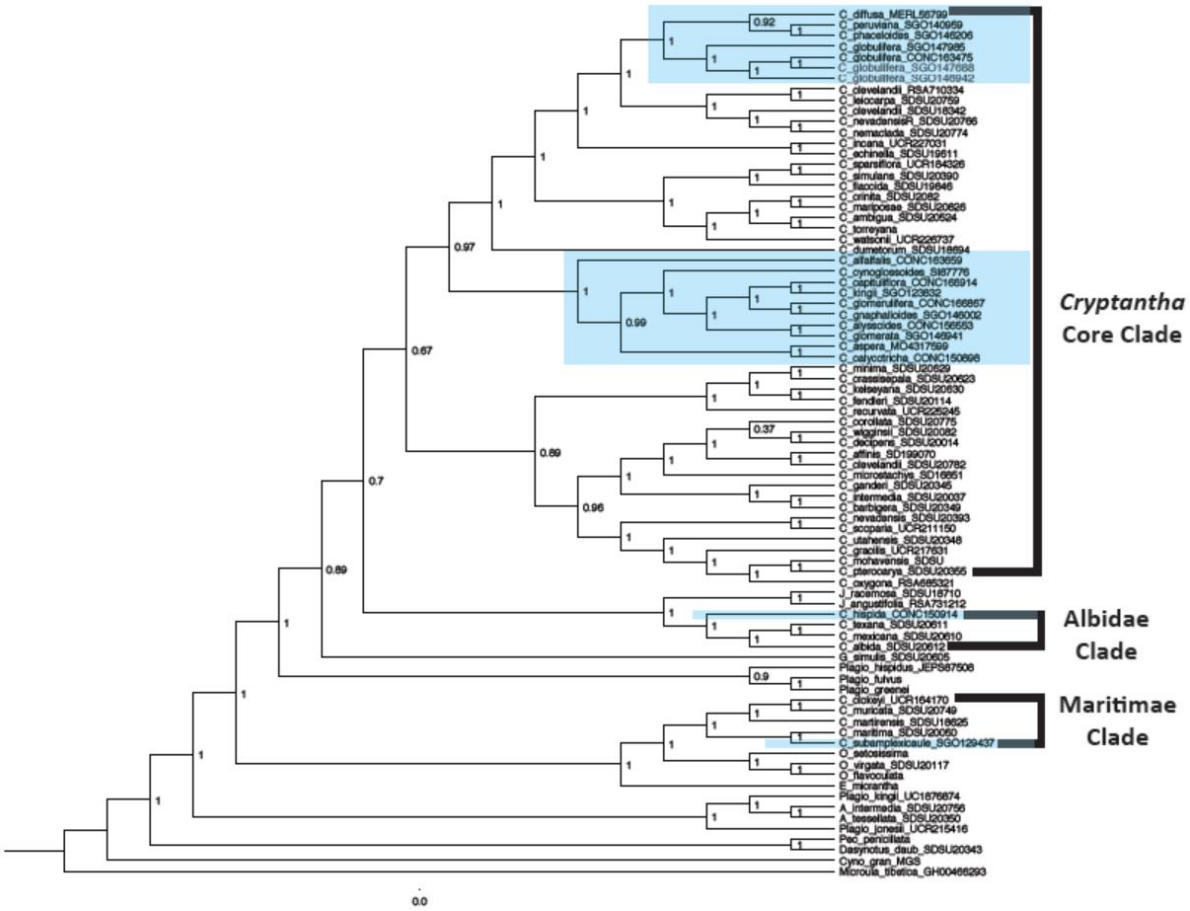

Figure 6. Maximum clade credibility tree using Bayesian Inference of the chloroplast (cpDNA). Major clades are identified and South American species are highlighted in blue. A=*Amsinckia,* C=*Cryptantha,* E=*Eremocarya* , G= *Greeneocharis,* J= *Johnstonella,* O=*Oreocarya.*

further referred to as the Albidae Clade. The Albidae Clade and *Johnstonella* itself, is recovered sister to the *Cryptantha* Core Clade with moderate support (BS=85, PP=0.7). The remaining *Cryptantha* taxa sampled form a well-supported clade (BB=100, PP=0.67). Within this *Cryptantha* Core Clade, two monophyletic groups of South America taxa are found, both strongly supported (BB=100, PP=1).

Both the ML and BI analyses of the cistron (nrDNA) resulted in the exact same topologies to one another (Figure 7; Figure 8). The Maritimae Clade is recovered as monophyletic (BS=58, PP=0.97), but differs from the cpDNA analysis in being sister to the *Cryptantha* Core Clade with weak support (BS=35, PP=0.64).The Albidae Clade is recovered as monophyletic with strong support (BB=100, PP=1); however, *C. hispida* falls out with the two representatives of the genus *Johnstonella*, as opposed to the other

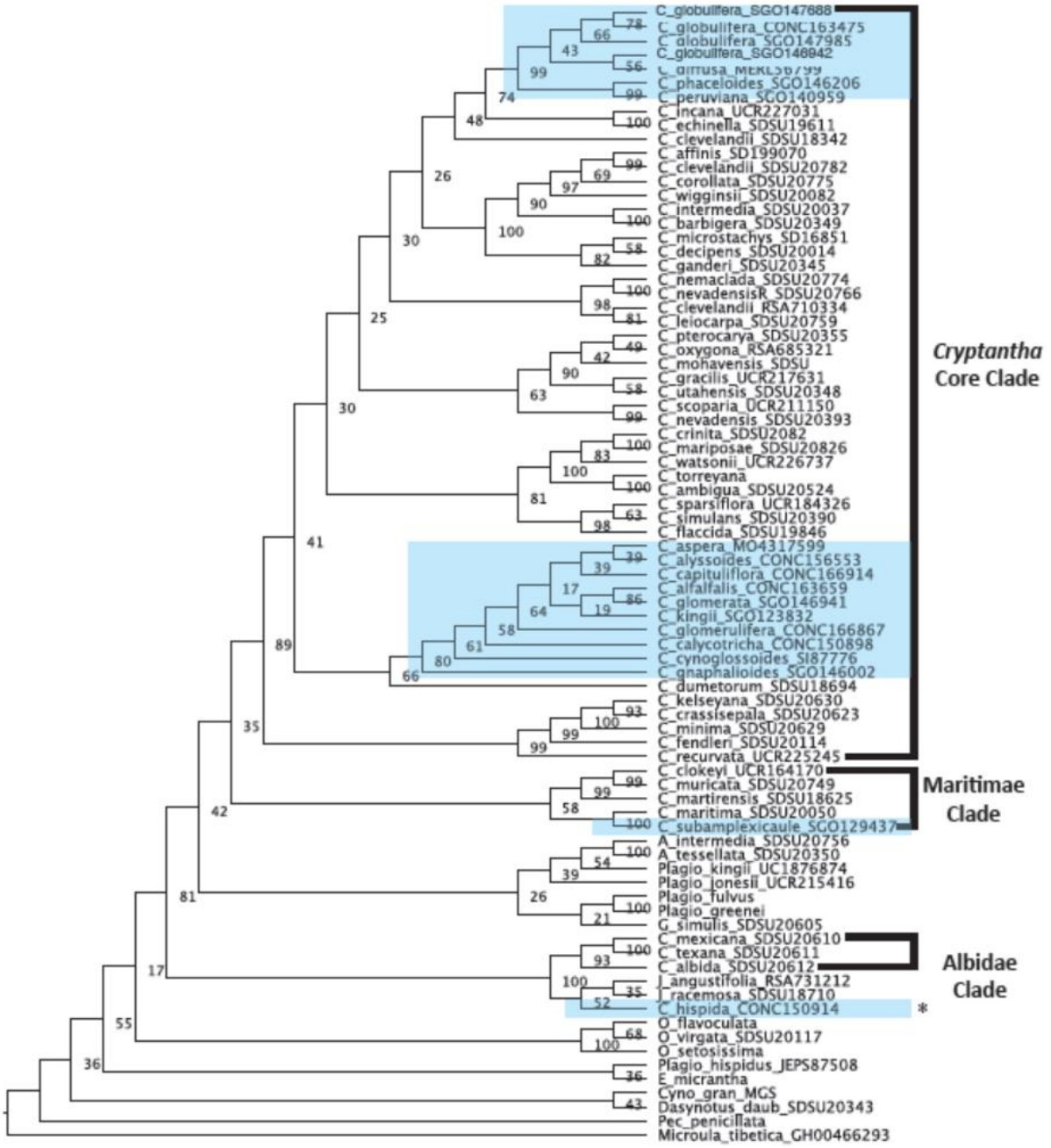

Figure 7. Maximum likelihood tree of the ribosomal cistron (nrDNA). Major clades are identified and South American species are highlighted in blue. * Indicates *C. hispida* position with *Johnstonella.* A=*Amsinckia,* C=*Cryptantha,* E=*Eremocarya,* G= *Greeneocharis,* J= *Johnstonella,* O=*Oreocarya.*

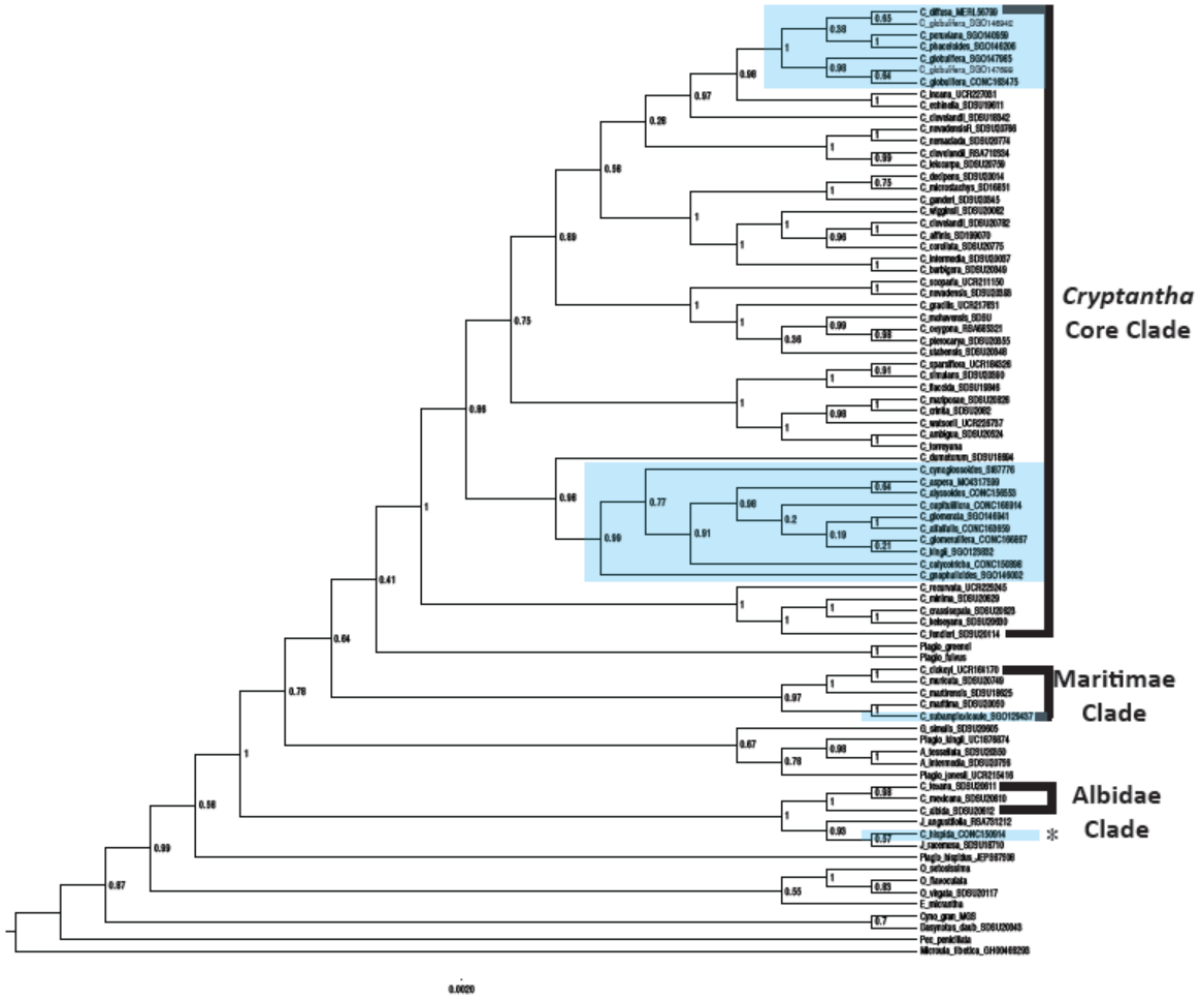

Figure 8. Maximum clade credibility tree using Bayesian Inference of the ribosomal cistron (nrDNA). Major clades are identified and South American species are highlighted in blue. * Indicates *C. hispida* position with *Johnstonella.* A=*Amsinckia,* C=*Cryptantha,* E=*Eremocarya* , G= *Greeneocharis,* J= *Johnstonella,* O=*Oreocarya.*

Cryptantha taxa (indicated with an * in Figure 7 and Figure 8). The *Cryptantha* Core Clade is again resolved as monophyletic with strong support (BB=89, PP=1). Both South American clades are recovered as monophyletic, however; *C. dumetorum* (A. Gray) Greene is found as sister to one South American clade, and *C. incana* Greene and *C. echinella* Greene are together found sister to the other South American clade.

Mitochondrial DNA (mtDNA) ML and BI analyses did not return trees with the same topology (Figure 9, Figure 10). In both trees, all three major clades from the previous analyses are recovered as monophyletic: the Maritimae Clade and the Albidae Clade with strong support (BS=80, PP=0.98; BS= 100, PP=1, respectively) and the *Cryptantha* Core

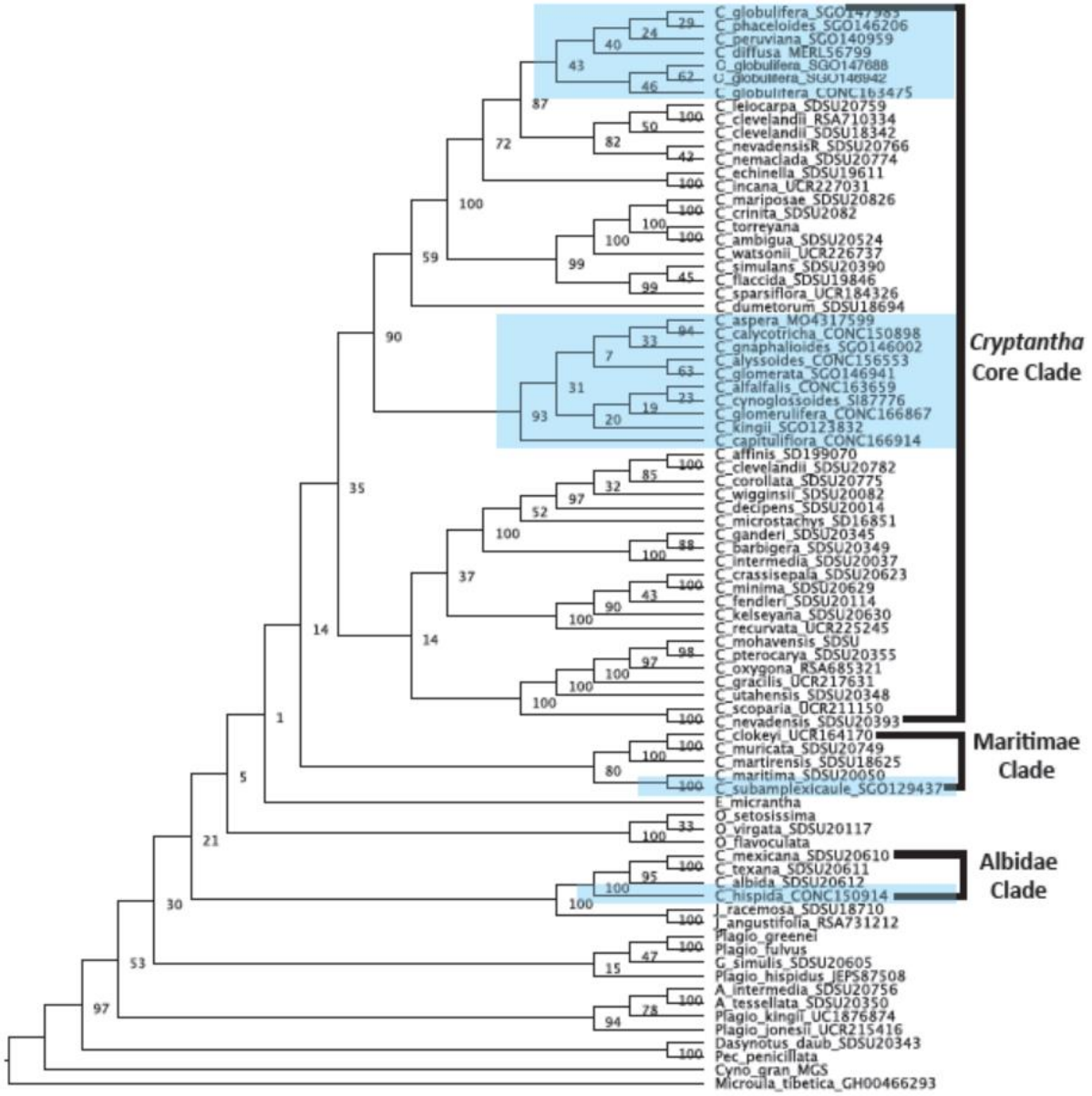

Figure 9. Maximum likelihood tree of 23 concatenated mitochondrial genes (mtDNA). Major clades are identified and South American species are highlighted in blue. A=*Amsinckia*, C=*Cryptantha*, E=*Eremocarya* , G= *Greeneocharis*, J= *Johnstonella*, O=*Oreocarya.*

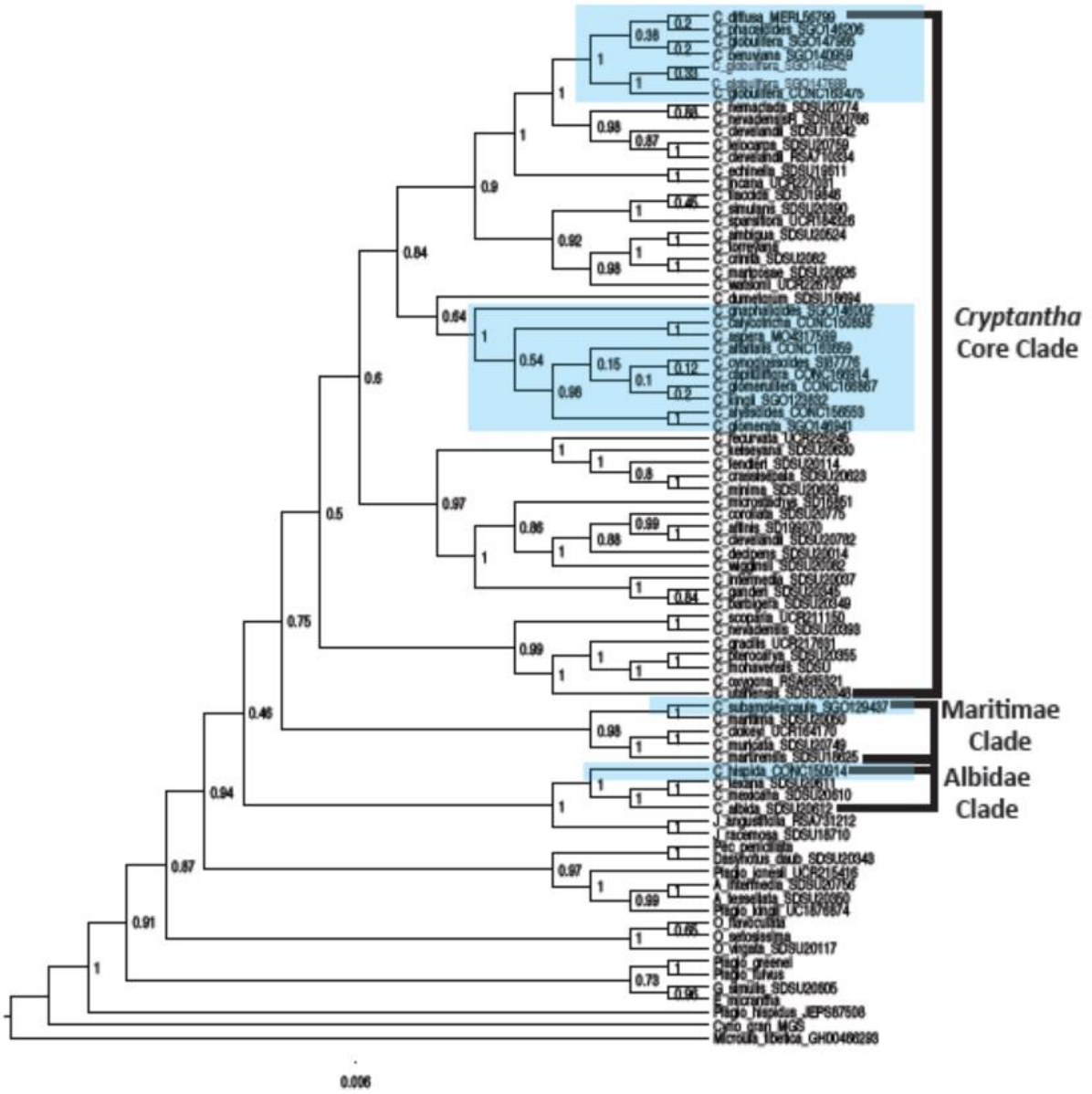

Figure 10. maximum clade credibility tree using Bayesian Inference of 23 concatenated mitochondrial genes (mtDNA). Major clades are identified and South American species are highlighted in blue. A=*Amsinckia*, C=*Cryptantha*, E=*Eremocarya* , G= *Greeneocharis*, J= *Johnstonella*, O=*Oreocarya.*

Clade with weak support (BS=35, PP=0.5). The major difference between the ML and BI analyses is the placement of the other genera in relation to these major clades. The *Cryptantha* Core Clade and the Maritimae Clade are recovered as sister in both analyses with weak support (BS=14, PP=0.75), but the placement of the Albidae Clade is different in these two trees.

Overall, the mtDNA tree provided poor support (least amount of supported nodes) for the relationships of these taxa.

Species tree estimation using ML concatenation of the three regions resulted in the tree with the greatest number of well-supported nodes (Figure 11). All but one node is strongly supported with a bootstrap of 80 or better. The same three major clades are recovered as in the gene trees. However, in the ML concatenated species tree, the placement of these three clades in relation to one another and in relation to other genera is resolved with high support. The Albidae Clade is placed sister to the *Cryptantha* Core Clade (BS=89), while the Maritimae Clade is placed sister to *Oreocarya* and *Eremocarya*. These relationships were also recovered in both the ML and BI cpDNA analysis. Concatenation using BI resulted in a tree with the exact same topology as the ML tree (not shown here).

Species tree estimates using *BEAST (version 1.8.0, Drummond et al. 2012) for a multi-species coalescent approach were unable to converge after 500 million generations. Therefore *BEAST analyses were run on a reduced taxa dataset of only 50 taxa (Figure 12). The tree topology of the *BEAST tree recovered the same major clades with poor support for the *Cryptantha* Core Clade and moderate support for both the Maritimae Clade and Albidae Clade. Species tree estimates using STAR (Liu et al. 2009) produced a completely supported phylogeny (BS=100 for all nodes; Figure 13). To compare species tree estimates of STAR (Liu et al. 2009) to *BEAST (Drummond et al. 2012), both the full dataset and the reduced taxa dataset were run in STAR (Liu et al. 2009). The phylogeny of the reduced taxa datset of STAR (Liu et al. 2009), was also completely supported (BS=100 for all nodes). ASTRAL (Mirarab et al. 2014) recovered the same three major clades; however, the placement of them in relation to the other genera is incongruent with both the ML and STAR (Liu et al. 2009) topologies (Figure 14).

CHARACTER EVOLUTION

Using the maximum likelihood (ML) concatenated tree, character evolution using ML and the MK1 model in Mesquite (Maddison and Maddison 2010) resulted in the reconstruction of six traits that are diagnostic for species identification in *Cryptantha*. For nutlet number per fruit, there was equal likelihood for any of the states to be ancestral. This was true for all three major clades (Figure 15). Analysis of fruit heteromorphism, however,

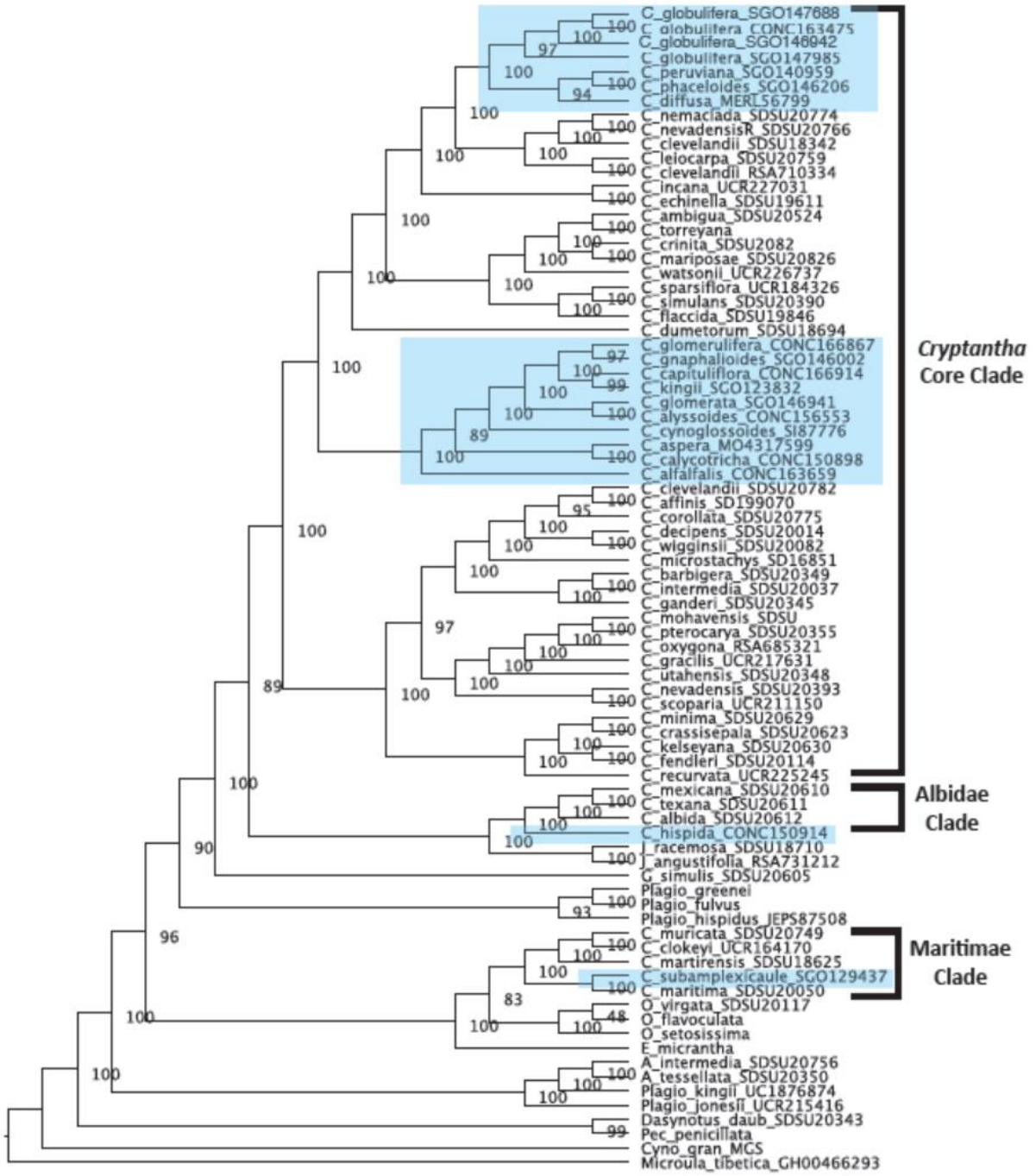

Figure 11. Maximum likelihood tree of concatenated cpDNA (chloroplast), nrDNA (ciston), and mtDNA (mitochondrial) regions. Major clades are identified and South American species are highlighted in blue. A=*Amsinckia*, C=*Cryptantha*, E=*Eremocarya* , G= *Greeneocharis*, J= *Johnstonella*, O=*Oreocarya.*

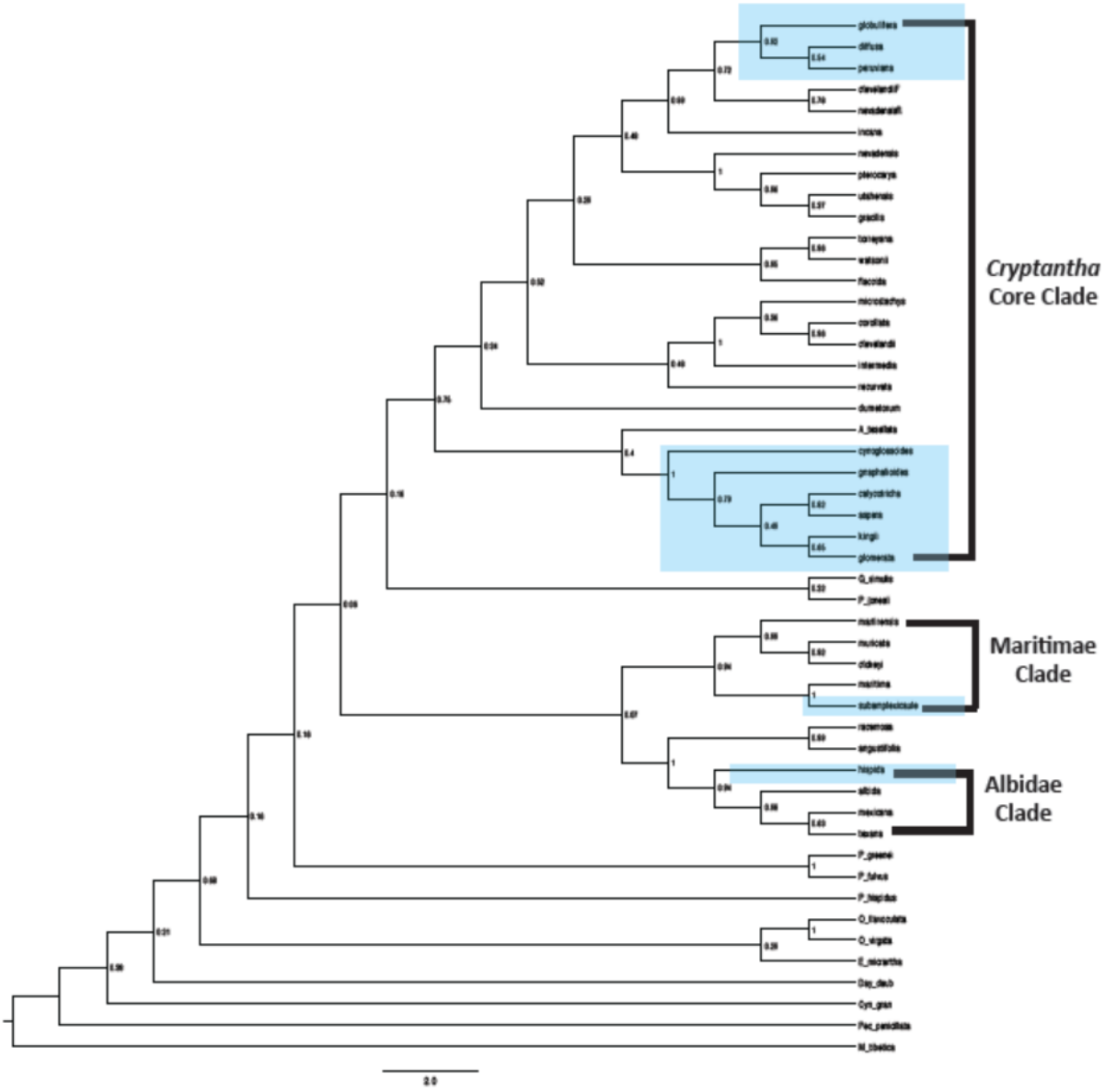

Figure 12. Muli-species coalescent tree, as inferred with *BEAST of the reduced (50) taxa dataset. Major clades are identified and South American species are highlighted in blue. A=*Amsinckia,* E=*Eremocarya,* G= *Greeneocharis,* M= *Microula,* O=*Oreocarya,* P= *Plagiobothrys.*

strongly supported homomorphic nutlets as the ancestral state for each of the three major clades. Within the *Cryptantha* Core Clade, heteromorphism evolved a minimum of seven times (Figure 16). The Albidae Clade was strongly supported as ancestrally homomorphic, and within the Maritimae Clade two species, *C. subamplexicaulis* and *C. maritima,* are either homomorphic or heteromorphic. With regard to nutlet sculpturing, rough nutlets are strongly supported as ancestral for all three major clades; smooth nutlets evolved as many as nine

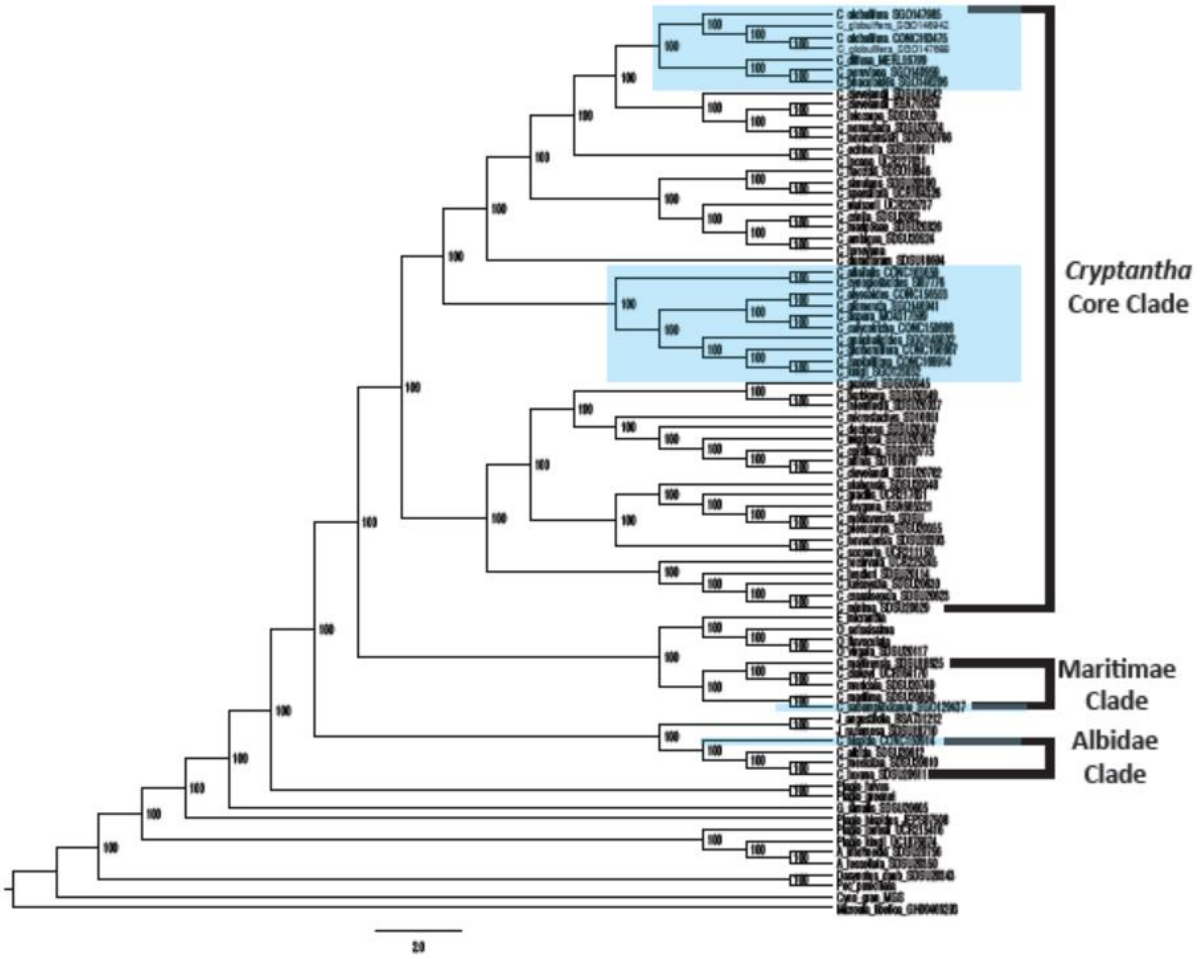

Figure 13. Species tree estimated using STAR of the full dataset (all 81 taxa). Major clades are identified and South American species are highlighted in blue. A=*Amsinckia*, C=*Cryptantha*, E=*Eremocarya* , G= *Greeneocharis*, J= *Johnstonella*, O=*Oreocarya.*

times in the *Cryptantha* Core Clade (Figure 17). Both South American clades nested in the *Cryptantha* Core Clade recover rough as the ancestral condition with strong support. For plant duration, annual is resolved as ancestral for all three major clades (Figure 18). Perennial plant duration is found to have evolved at least once in the South American *Eucryptantha/Geocarya* clade. Ancestral reconstruction for cleistogamy recovered chasmogamy as the ancestral state, with cleistogamy evolving once in the South American *Eucryptantha/Geocarya* clade (Figure 19). This clade consists of Johnston's (1927) sections *Eucryptantha* and *Geocarya*. Within this clade, section *Geocarya* is recognized as having cleistogenes, specialized fruits born at the base of the plant. These cleistogenes have evolved at least three times within this clade. One reversal to chasmogamy, in *C. gnaphalioides*

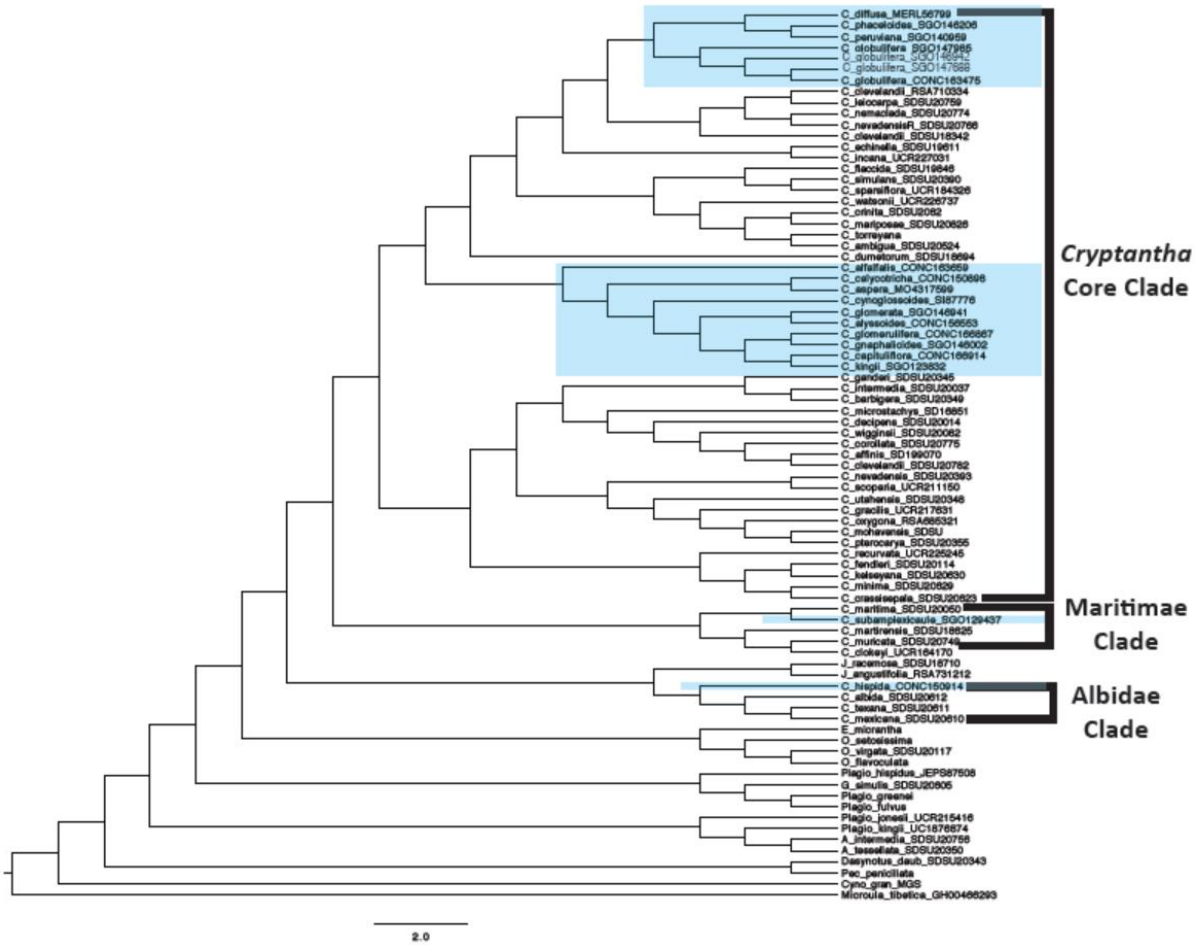

Figure 14. Species tree estimated using ASTRAL of the full dataset (all 81 taxa). Major clades are identified and South American species are highlighted in blue. A=*Amsinckia*, C=*Cryptantha*, E=*Eremocarya* , G= *Greeneocharis*, J= *Johnstonella*, O=*Oreocarya.*

(A.DC.) Reiche, is recovered. Lastly, analysis of trichome vesiture, specifically the orientation of the trichome to the stem, recovered spreading trichomes as ancestral with appressed trichomes evolving a minimum of six times (Figure 20).

BIOGEOGRAPHIC INFERENCE

The statistical analysis in BIOGEOBEARS (Matzke 2012, 2013) using the log likelihood score returned the BAYAREALIKE model as the best fit for the data (Table 6). The BAYAREALIKE model excludes vicariance, only allowing complete sympatric speciation to occur. Surprisingly, is that the BAYAREALIKE+J model returned a lower log likelihood score. The "J" function allows for jump dispersal to occur, which for *Cryptantha* should be considered. This lower log likelihood may have originated from incorrect starting

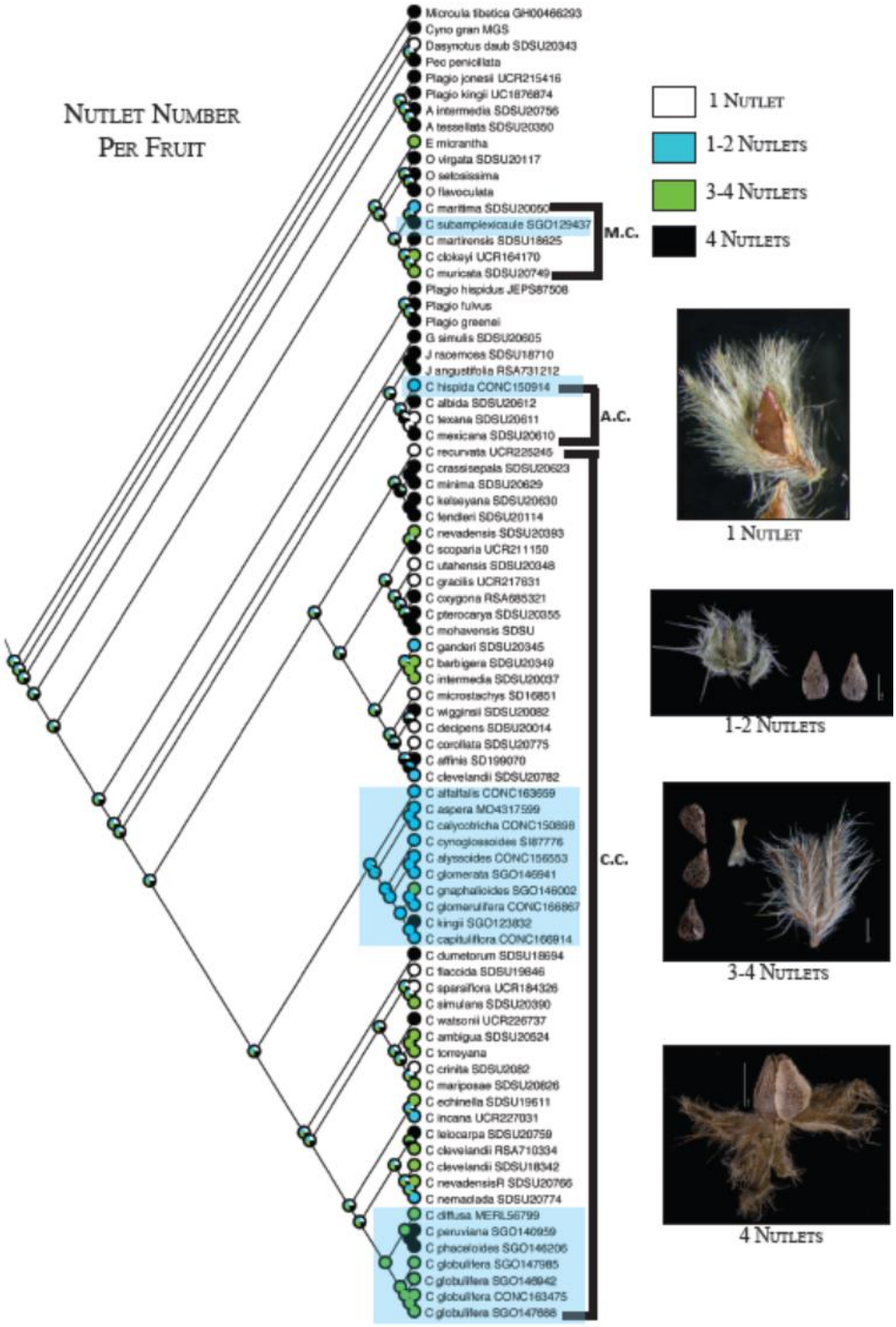

Figure 15. Character evolution of nutlet number per fruit, maximum likelihood tree shown. White= 1 nutlet/fruit, blue= 1-2 nutlets/fruit, green= 3-4 nutlets/fruit, black= 4 nutlets/fruit. Major clades are identified and South American species are highlighted in blue. A.C.= Albidae Clade, C.C.= *Cryptantha* Core Clade, M.C.= Maritimae Clade.

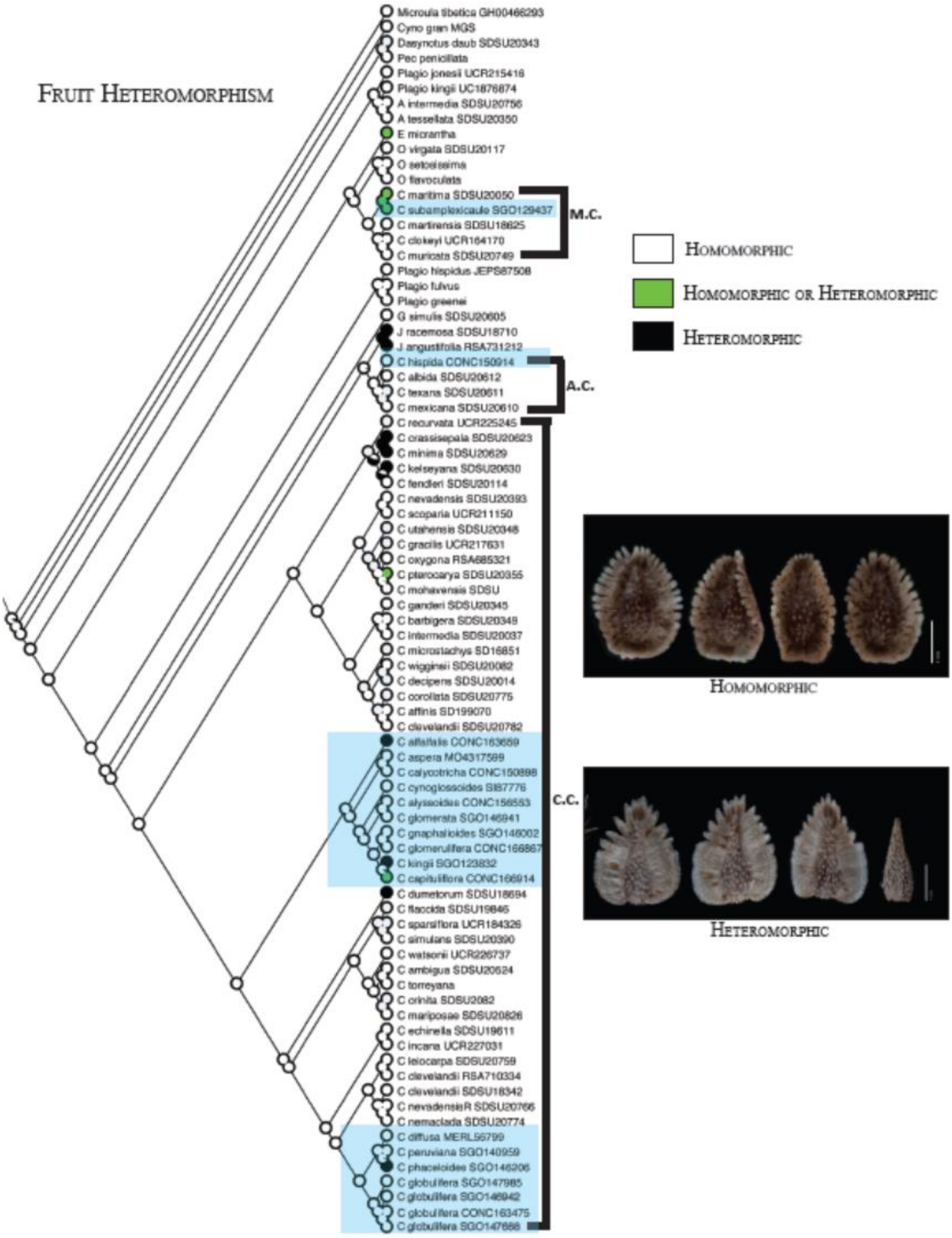

Figure 16. Character evolution of fruit heteromorphism, maximum likelihood tree shown. White= homomorphic, green= homomorphic or heteromorphic, black= heteromorphic. Major clades are identified and South American species are highlighted in blue. A.C.= Albidae Clade, C.C.= *Cryptantha* Core Clade, M.C.= Maritimae Clade.

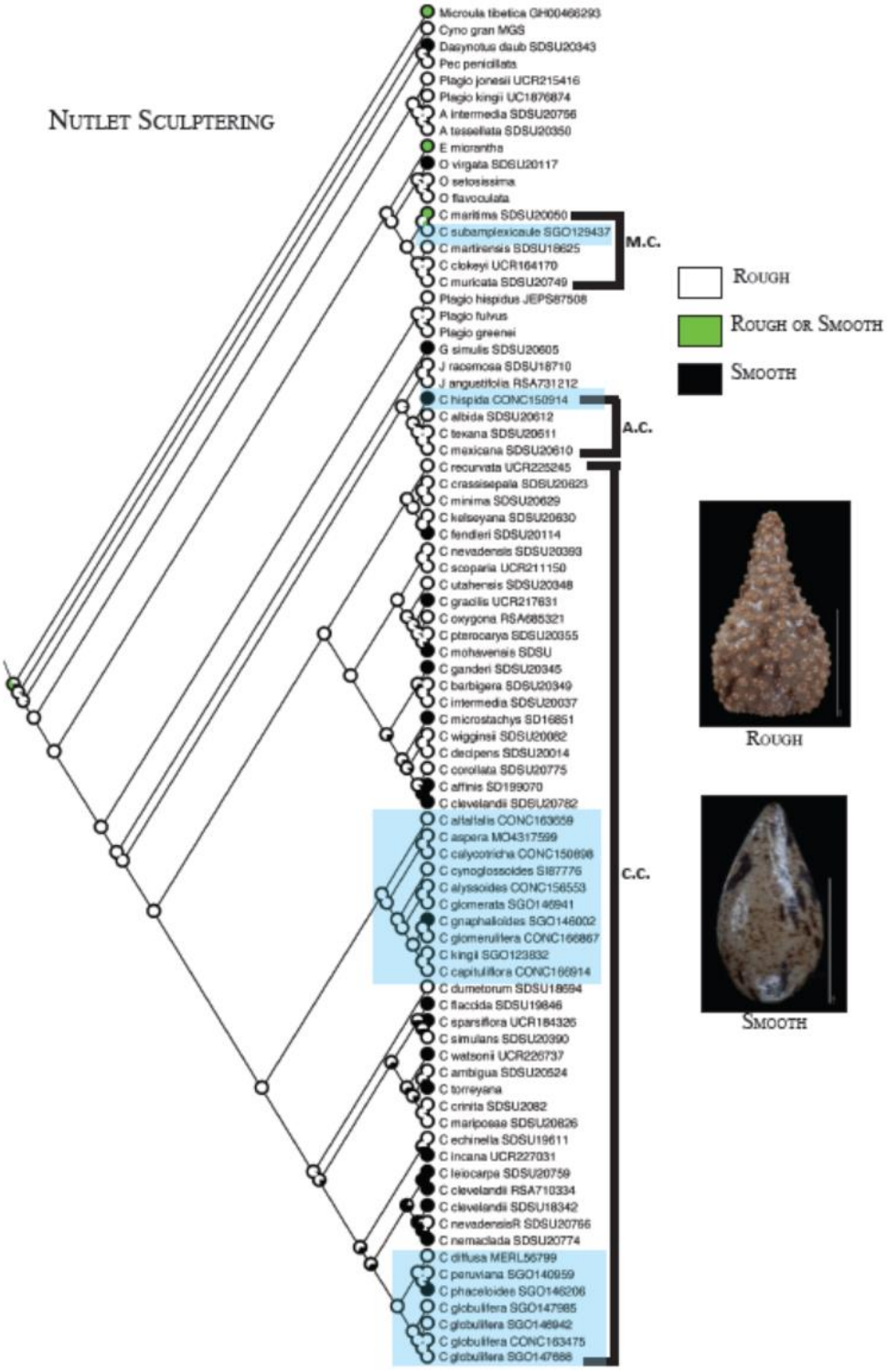

Figure 17. Character evolution of nutlet sculpturing, maximum likelihood tree shown. White= rough nutlets, green= rough or smooth nutlets, black= smooth nutlets. Major clades are identified and South American species are highlighted in blue. A.C.= Albidae Clade, C.C.= *Cryptantha* Core Clade, M.C.= Maritimae Clade.

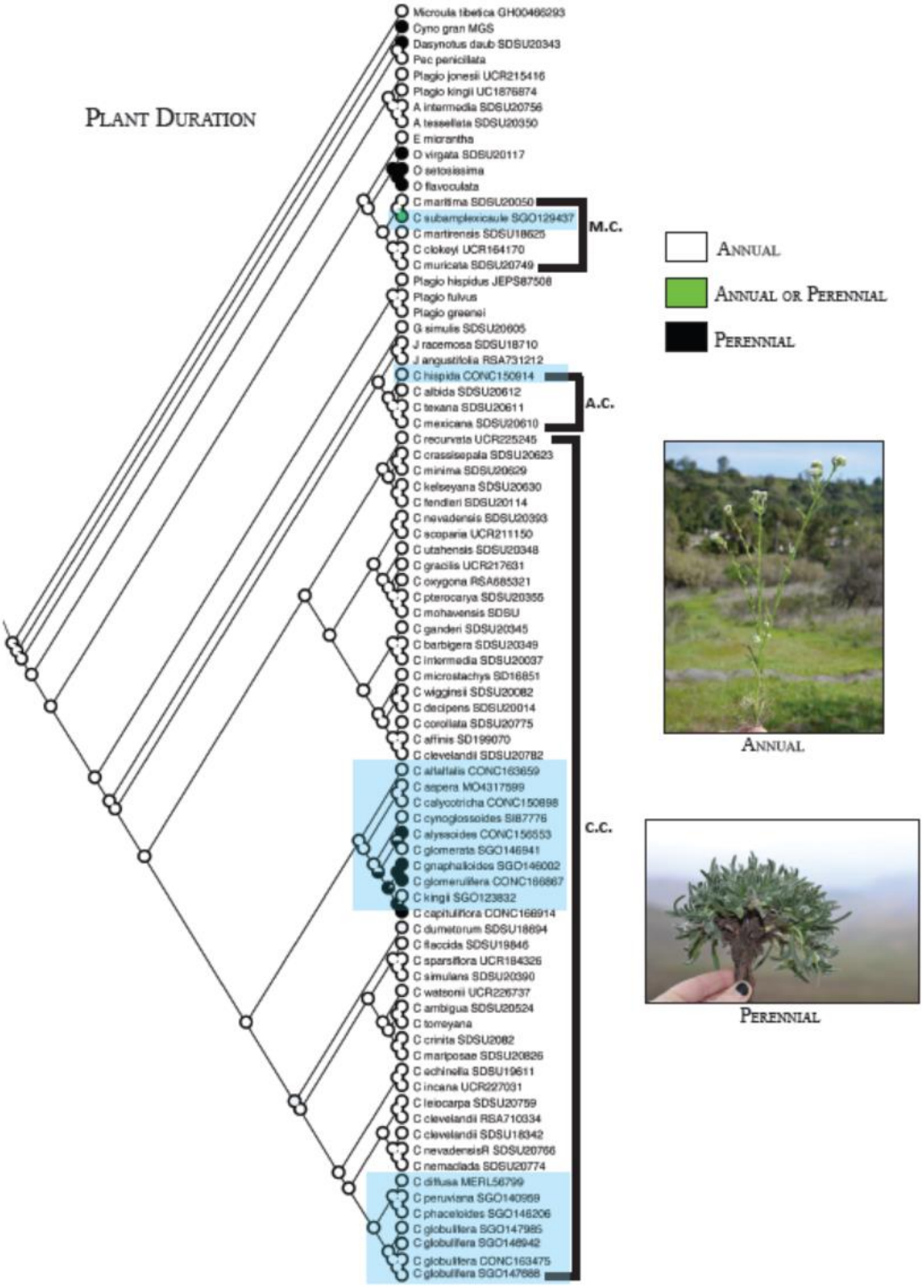

Figure 18. Character evolution of plant duration, maximum likelihood tree shown. White= annual, green= annual or perennial, black= perennial. Major clades are identified and South American species are highlighted in blue. A.C.= Albidae Clade, C.C.= *Cryptantha* Core Clade, M.C.= Maritimae Clade.

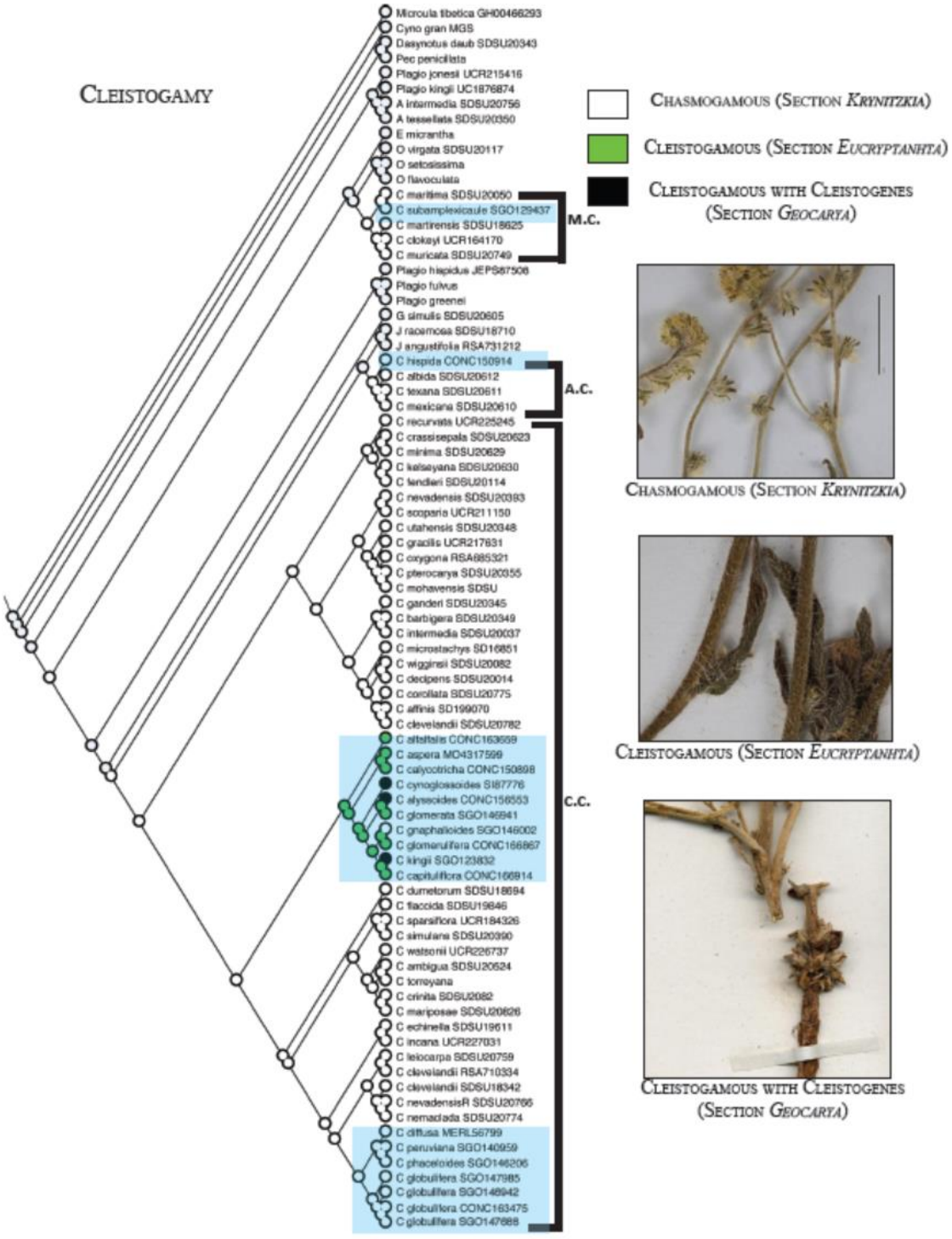

Figure 19. Character evolution of cleistogamy, maximum likelihood tree shown. White= chasmogamous (section *Krynitzkia*), green= cleisogamous (section *Cryptantha*), black= cleisogamous with cleistogenes (section *Geocarya*). Major clades are identified and South American species are highlighted in blue. A.C.= Albidae Clade, C.C.= *Cryptantha* Core Clade, M.C.= Maritimae Clade.

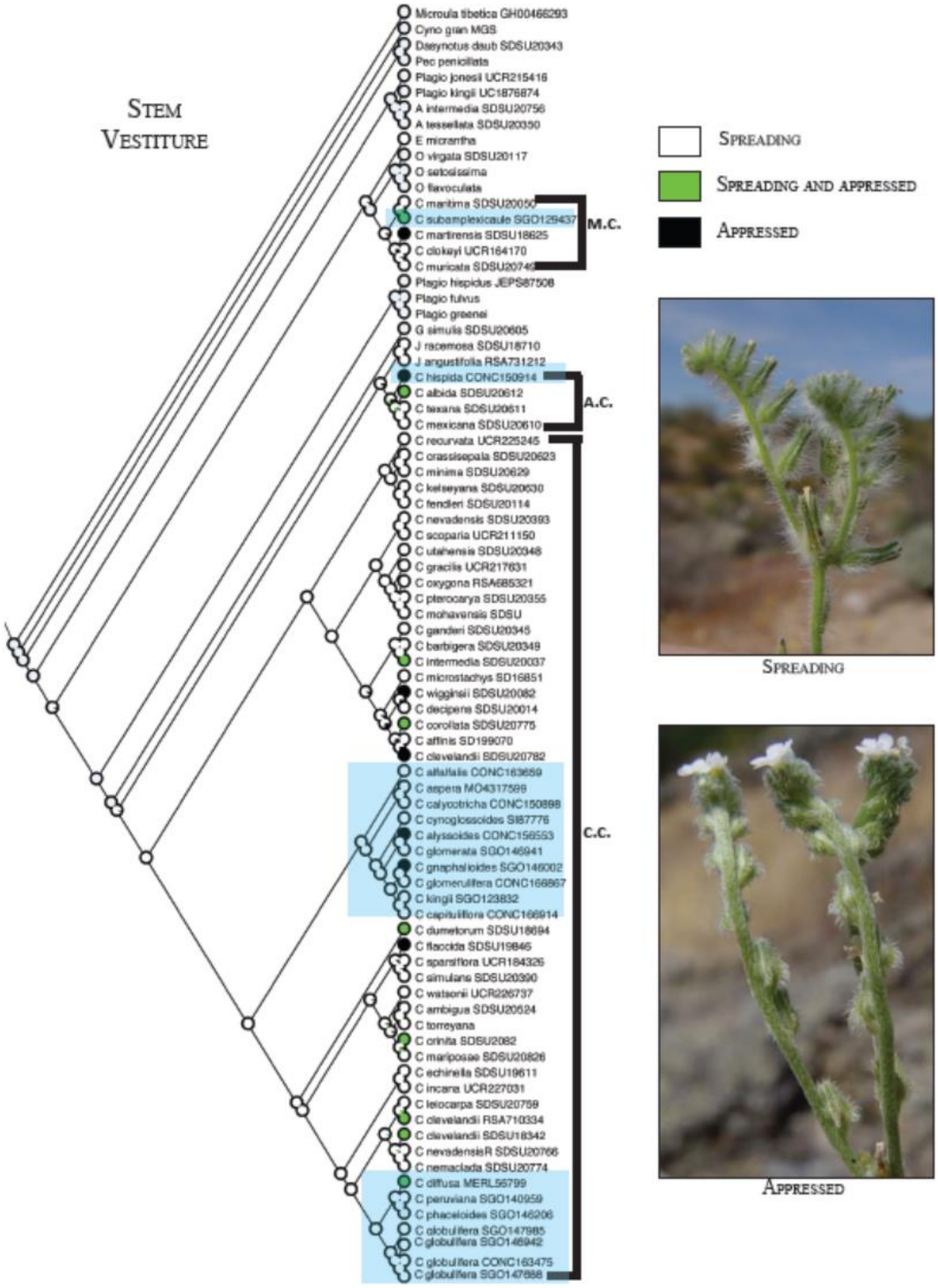

Figure 20. Character evolution of trichome vestiture, maximum likelihood tree shown. White= only spreading trichomes, green= spreading and appressed trichomes, black= only appressed trichomes. Major clades are identified and South American species are highlighted in blue. A.C.= Albidae Clade, C.C.= *Cryptantha* Core Clade, M.C.= Maritimae Clade.

Table 6. Log Likelihood Scores for Each Model of Biogeographic Dispersal Run in BioGeoBEARS (Matzke 2012, 2013)

Model	LnL
DEC	-530.1110
DEC+J	-522.8270
DIVALIKE	-621.8023
DIVALIKE+J	-604.4419
BAYAREALIKE	-446.7581
BAYAREALIKE+J	-655.7634

Note: The model BAYAREALIKE had the highest log likelihood score and was therefore chosen as the model that explained the data best.

values. Both analysis, with and without "J", were similar with regard to major dispersal events.

A minimum of four unidirectional intercontinental dispersals are recovered. All dispersal events resulted from a Mediterranean North America ancestor dispersing into the Mediterranean South America region (Figure 21; Figure 22). Within North America, one dispersal into the temperate mountain system and multiple dispersals into the desert regions are recovered. The Albidae Clade dispersed from the Mediterranean North America region to many desert regions, including the tropical desert region (the Atacama Desert) of South America (*C. hispida*). There is strong support for a Mediterranean North America ancestry of the Maritimae Clade, with most of the species that compose this clade still found in the Mediterranean Region of North America. In this same clade, one dispersal to the South America tropical desert (the Atacama Desert) is recovered (*C. subamplexicaulis*). Both dispersals from North to South America in the *Cryptantha* Core Clade had Mediterranean North America ancestors. In the first South America clade, the ancestor dispersed to the Mediterranean South America region with a later dispersal to the high elevation areas of the Andes. Also in this clade, one dispersal back to the Mediterranean region of South America is recovered by *C. gnaphalioides*. The second South America clade had an ancestor that dispersed from Mediterranean North America to Mediterranean South America, with a later dispersal to the tropical Andes (*C. peruviana* I.M. Johnston).

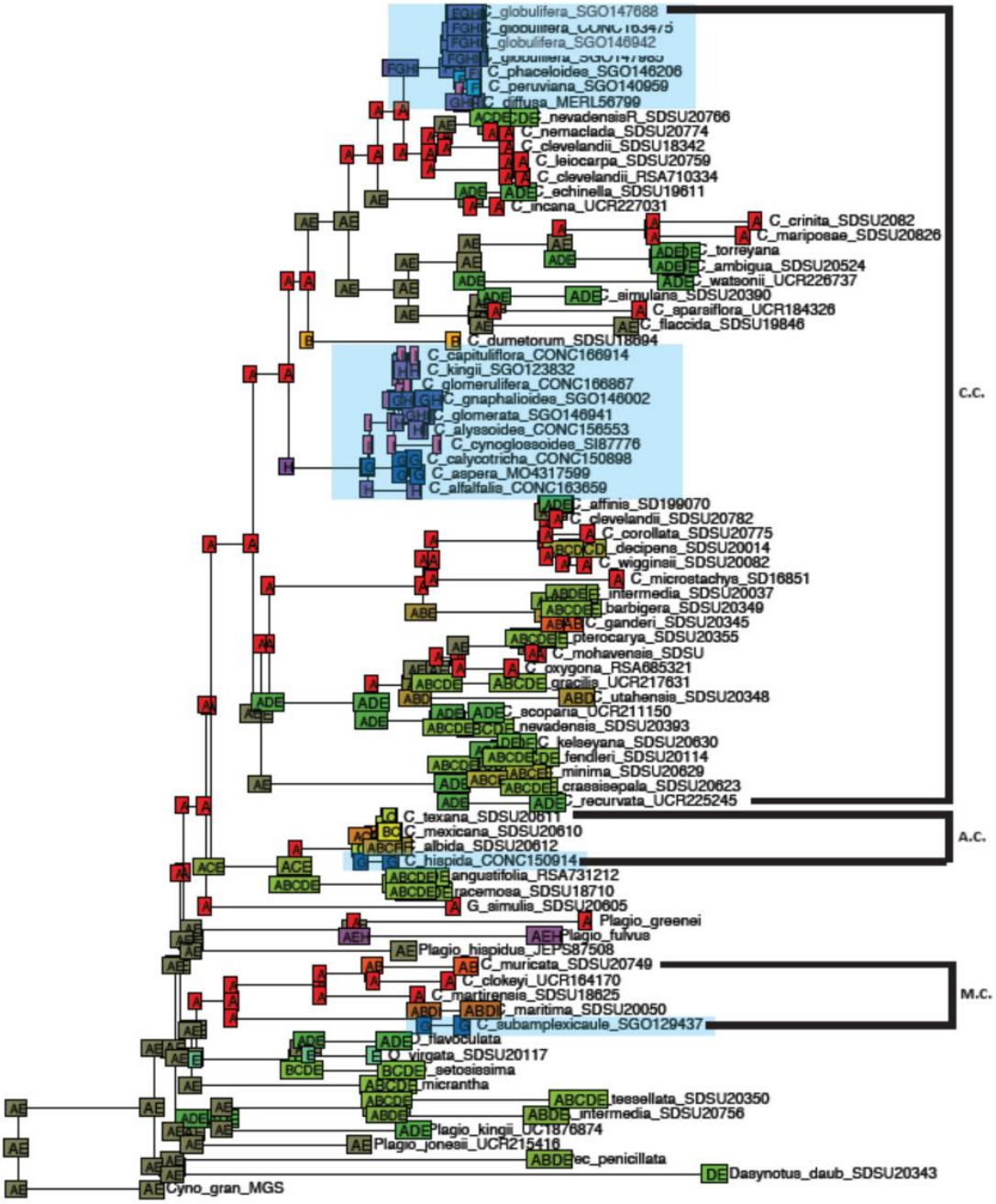

Figure 21. BioGeoBEARS graphical output, showing the most likely ancestral range for *Cryptantha*. A (red) = North America subtropical dry forest and mountain system, B (orange) = North America subtropical desert, C (yellow) = North America subtropical steppe, D (light green)= North America temperate desert, E (green) = North America temperate mountain system, F (light blue)= South America tropical mountain system, G (blue) = South America tropical desert, H (purple)= South America subtropical steppe and dry forest, I (pink)= South America temperate oceanic forest. Major clades are identified and South American species are highlighted in blue. A.C.= Albidae Clade, C.C.= *Cryptantha* Core Clade, M.C.= Maritimae Clade.

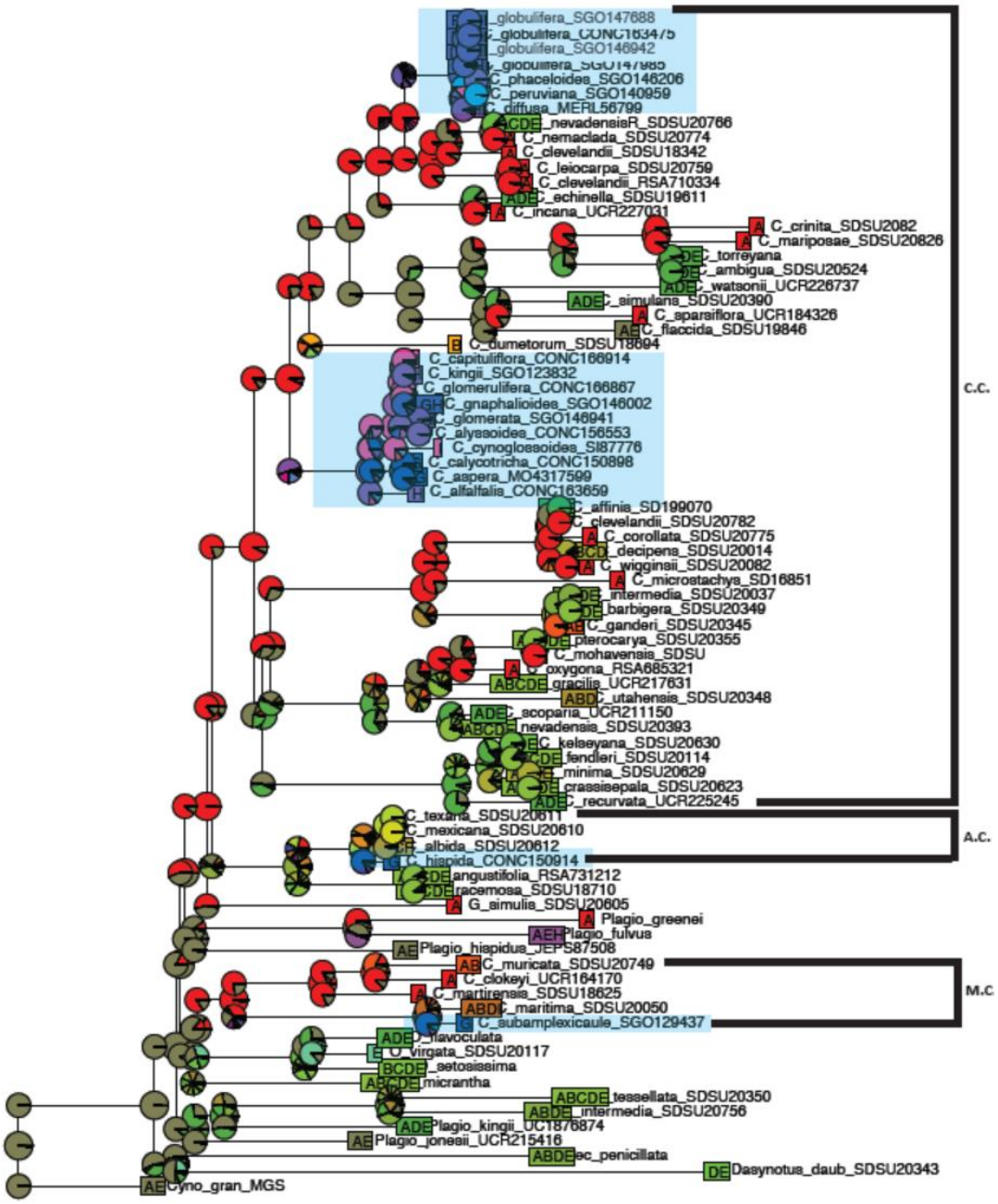

Figure 22. BioGeoBEARS graphical output, showing the most likely ancestral range for *Cryptantha* in pie graph form. A (red) = North America subtropical dry forest and mountain system, B (orange) = North America subtropical desert, C (yellow) = North America subtropical steppe, D (light green)= North America temperate desert, E (green) = North America temperate mountain system, F (light blue)= South America tropical mountain system, G (blue) = South America tropical desert, H (purple)= South America subtropical steppe and dry forest, I (pink)= South America temperate oceanic forest. Major clades are identified and South American species are highlighted in blue. A.C.= Albidae Clade, C.C.= *Cryptantha* Core Clade, M.C.= Maritimae Clade.

DIVERGENCE TIME ESTIMATION

Divergence time estimates using the ITS rate of substitution or fossil dates as calibration returned dates of divergence that were two orders of magnitude different. Because the analysis using the ITS rate of substitution returned a date of divergence for the stem node of the Amsinckiinae that did not correspond to the accepted date of divergence of the Angiosperms, it was not considered for further discussion. Although, the analysis using fossil calibration returned very large confidence intervals for the stem node of the Amsinckiinae, the 95% confidence intervals included the Angiosperm divergence date of about 130 Ma. Within the *Cryptantha* Core Clade, two dispersals from North to South America occurred. The first clade diversified at about 23Ma, and the second clade much more recently at around 4 Ma (Figure 23). The South America species *C. hispida*, which is nested in the Albidae Clade, originated around 17 Ma from other North America species in this clade. *Cryptantha subamplexicaulis* from South America, in the Maritimae Clade, also originated around 17 Ma.

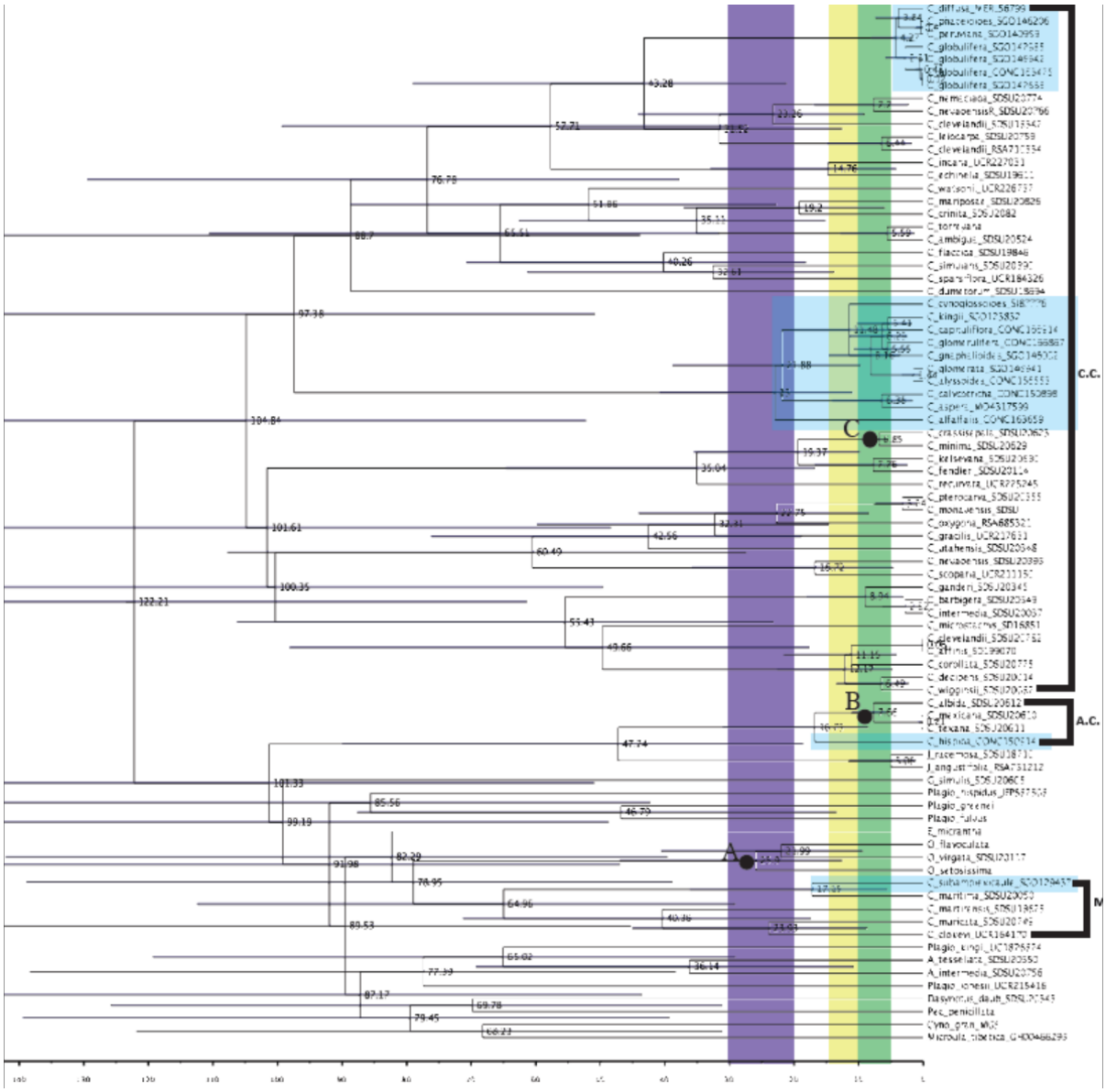

Figure 23. Bayesian tree showing dates of clade diversification. South America taxa are highlighted in blue. In purple is the appoximate timing of the first uprise of the Andes, in green is the appoximate timing of the second pulse of that uprise. In yellow is the appoximate timing of the hyperridity of the Atacama Desert. Calibrated nodes are indicated with black circles (A, B, and C). A= *Cryptantha chaneyi*, B= *C. auriculata*, C= *C. coroniformis.* Major clades are identified: A.C. = Albidae Clade, C.C. = *Cryptantha* Core Clade, M.C. = Maritimae Clade.

DISCUSSION

PHYLOGENETIC ANALYSIS

Genome skimming methods successfully recovered nearly complete sequence data from the three major regions of the plant genome for all taxa studied. However, trees obtained using each of the separate genomes differed. Possible reasons for the incongruence between these genomes may be related to how they are inherited. Both the chloroplast and mitochondria are uniparentally inherited, possibly confounding results by tracing evolution from only one line of descent (Rieseberg and Soltis 1991; Rieseberg and Wendel 1993). Problems have also been noted with regard to using the ITS regions of the cistron (nrDNA) for phylogenetic analyses (Alvarez and Wendel 2003). Although the cistron is part of the nuclear genome and is therefore biparently inherited, many plant genomes are found with several different copies of ITS sequences (Alvarez and Wendell 2003). These multiple copies are perhaps due to incomplete homogenization, making paralog sequence relationships potentially misleading for phylogenetic analysis (Alvarez and Wendell 2003). For this analysis, positions of the ITS that may have been subject to incomplete homogenization were removed using a strict 75% matching consensus sequence requirement and removing any base pair positions with ambiguity codes.

In all analyses *Cryptantha* is recovered as non-monophyletic. Although there is discordance between the three regions (cpDNA, mtDNA, nrDNA) on the placement of these clades, all analyses recover three well-supported monophyletic groups of *Cryptantha* taxa. The clades recovered include the Maritimae Clade (compatible with Hasenstab-Lehman and Simpson's 2012 *Cryptantha* s.s. 2, but with additional taxa added and two unexamined in this analysis), consisting of North American *C. maritima*, *C. martirensis*, *C. muricata*, and *C. clokeyi*, plus the South American species *C. subamplexicaulis*. A second group, the Albidae Clade, includes the North American *C. texana*, C. *mexicana*, and *C. albida*, plus the South American species *C. hispida*; none of these taxa were examined by Hasenstab-Lehman and Simpson (2012). Lastly, a clade of the remaining *Cryptantha* species is recovered in all

analyses. This *Cryptantha* Core Clade is largely compatible with *Cryptantha* s.s. 1 of Hasenstab-Lehman and Simpson (2012), but with a large addition of samples and some samples omitted.

The ML concatenated tree provides the strongest support for the placement of the three major clades in relation to one another and to other genera. Although, the STAR tree did recovered 100 BS for all nodes, this is likely due to only three gene trees as input. Although the STAR tree may represent the species tree, for this analysis, the ML concatenated tree is accepted. In the ML concatenated tree all nodes except one have a bootstrap support of greater than 80 (Figure 11). Placement of the *Cryptantha* Core Clade and Maritimae Clade differs from that found by Hasenstab-Lehman and Simpson (2012), although some similarities are noted. The Maritimae Clade is similar to the *Cryptantha* s.s. 2 clade and the *Cryptantha* Core Clade in this study is compatible to the *Cryptantha* s.s. 1 clade in Hasenstab-Lehman and Simpson (2012). Unlike Hasenstab-Lehman and Simpson (2012), the addition of more taxa and significantly more data allows for resolution of the relationships of these clades. The Albidae Clade, along with all examined species of the genus *Johnstonella*, is recovered as sister to the *Cryptantha* Core Clade with fairly strong support (BS=89). *Greeneocharis* is found as sister to these two sister groups, followed by a clade of *Plagiobothrys*. The Maritimae Clade forms a well-supported group sister to *Oreocarya* and *Eremocarya*. Most differences from Hasenstab-Lehman and Simpson (2012) are with regard to the placement of other Amsinckiinae genera in relation to the three major clades recovered. Although concatenation has its caveats, especially in a dataset where the cpDNA dataset is more than ten times greater in length than the other two regions, the concatenated analyses results in the tree with the greatest number of well-supported nodes in this study.

To assess species relationships under a multi-species coalescent model, *BEAST (Drummond et al. 2012) was used with a reduced, 50 taxa dataset (Figure 12). This analysis with the reduced dataset resulted in relationships that were not congruent with those found in the ML concatenation, although the three major clades are recovered. Because of the large number of taxa used in the study and the large number of base-pairs obtained, summary statistic coalescent programs such as STAR (Liu et al. 2009) may be more appropriate than the multi-species coalescent approach. The relationships of the three major clades recovered

from the STAR analysis differ in placement from the ML concatenation. In the STAR tree, the Maritimae Clade plus *Eremocarya* and *Oreocarya* is sister the *Cryptantha* Core Clade, while in the ML tree the Albidae Clade plus *Johnstonella* is recovered as sister to the *Crypantha* Core Clade. However, given only three gene trees were used, this STAR species tree estimate may not be accurate for species tree inference. Simulations show that summary statistic coalescence methods require many gene trees (more than three) to accurately recover the true species tree (Mirarab et al. 2014).

To test the accuracy of the STAR species tree analysis, *BEAST (Drummond et al. 2012) was run with a subset of 50 taxa, as discussed above. The same subset of taxa was then run in STAR and compared. Again the same three major clades of *Cryptantha* taxa are recovered, but the placement of them in relation to other genera differs from other analyses. The program ASTRAL (Mirarab et al. 2014) has been shown through simulation studies to recover the true species tree more often than STAR and was therefore used to provide species tree estimates as well (Mirarab et al. 2014). ASTRAL recovered all three major clades, but relationships between them and the other genera differed from both the *BEAST and STAR trees (Figure 14). With just three gene trees (cpDNA, nrDNA, mtDNA), summary statistic coalescent methods have access to limited information in accurately resolving the species tree and concatenation methods are preferred (Mirarab et al. 2014).

Even though the placement of the three major *Cryptantha* clades relative to one another and to other genera is unclear, interrelationships within these clades are well-supported and largely congruent. Recovery of the Albidae Clade as sister to *Johnsontella* is not a surprising find. *Cryptantha albida* and *C. mexicana* both share morphological features with the genus *Johnsontella*. The former have whitish tubercles and nutlets that are triangular in shape, similar to species of *Johnstonella* (Hasenstab-Lehman and Simpson 2012). *Cryptantha texana* of the Albidae Clade also has nutlets that bear similarities to the odd nutlet of *Johnstonella angustifolia* (I.M. Johnst.) Hasenstab & M.G. Simpson. *Cryptantha hispida* however, shares no known morphological similarities to the other members of this clade. Species of the Maritimae Clade were a little more surprising. Hasenstab-Lehman and Simpson (2012) also recovered a clade (which they termed *Cryptantha* s.s. 2) including *C. maritima* along with *C. chaetocalyx* (Philippi) I.M. Johnston, *C. grandulosa* (Ruiz & Pavon) I.M. Johnston, and a South American species of *C. maritima*

Unfortunately, the latter three samples did not pass quality control for library prep in this study, and were not included. However, for further discussion, these species are assumed to nest within the Maritimae Clade. The placement of the North American species *C. clokeyi*, *C. martirensis*, and *C. muricata* in this clade is unexpected. All three of these species are muricate in nutlet sculpturing, unlike the other species found in this clade, which are tuberculate or (in some *C. maritima*) even smooth. Overall, the Maritimae Clade contains taxa from both North and South America with varying morphological similarity. Because of this, no uniting non-molecular apomorphy is currently known, indicating that this is a group warranting additional study.

The (non-monotypic) taxonomic series of Johnston (1925), although based on diagnostic morphological characters, do not form monophyletic groups as inferred from these analyses. Many of his series, such as *Barbigerae*, are scattered throughout different clades with members of other series in the trees derived here. Multiple clades containing taxa of *Barbigerae* and *Leiocarpae* are found throughout the tree. However, some of his series form near monophyletic groups in combination. For example, series *Pterocaryae*, *Graciles*, and *Mohavenses* together form a clade when one taxa of *Mohavenses* (*C. watsonii* (A. Gray) Greene) is omitted (the latter placed in a clade with members from series *Ambiguae* and *Flaccidae*). Johnston's series were described using only morphologically characteristics, and, as suggested by the results from the character evolution analysis, many of these traits are evolutionary plastic.

Johnston's (1927) sections agree more with the molecular phylogenetic analyses presented here. One clade of South America taxa contains all species that are considered cleistogamous (sections *Eucryptantha* and *Geocarya*) One exception is the species *C. gnaphalioides*, which belongs to Johnston's section *Krynitzkia* and is found within this South American clade. Interestingly, however, this species has a perennial duration, like many *Eucryptantha* and *Geocarya* species. This taxon warrants additional sampling in future studies to verify its position within this clade. Within this *Eucryptantha-Geocarya* clade, however, neither of these two sections as defined by Johnston is monophyletic.

CHARACTER EVOLUTION

Ancestral state reconstruction for nutlet number per fruit showed no strong pattern. This characteristic is hard to identify, as many species have more nutlets then recorded here, but they do not mature all the way. Hasenstab-Lehman and Simpson (2012) inferred four nutlets per fruit as ancestral. Results here do not conflict with this result. The family Boraginaceae, subfamily Boraginoideae is delimited as having four-lobed ovaries. At maturity, each lobe typically develops into one unit fruit (the nutlet), containing a single seed. Many species in the complex consistently produce fruits with a reduced nutlet number used to define taxa (Hasenstab-Lehman and Simpson 2012). One find that corroborates Hasenstab-Lehman and Simpson (2012) is that a reduced (one-two) nutlet number is again found for the South American clade of cleistogamous taxa. This reduction in nutlet number may correlate with the evolution of this specialized self-pollinating mechanism, but further studies to test this hypothesis are needed.

Fruit heteromorphism evolved a minimum of six times within *Cryptantha* taxa, with homomorphic fruits as ancestral. Heteromorphic nutlets may be adaptive as a dispersal device. Generally, the larger nutlet remains firmly attached to the fruit gynobase, and the three smaller nutlets detach easily. This may provide a mechanism in which some propagules remain close to the parent, whereas other propagules are capable of dispersal to greater distances. One surprising find is that the Albidae Clade is not heteromorphic. The genus *Johnstonella*, which this clade is sister to, is characterized by most of its species having heteromorphic fruits; this implies that heteromorphism is a trait distinctive of *Johnstonella*, or of some subset of *Johnstonella*. The fact that the *Cryptantha* species found in the Albidae Clade are not heteromorphic suggests that this clade may be morphologically distinct.

Character analysis of nutlet sculpturing recovered rough nutlets as ancestral for all three major *Cryptantha* clades. The possible adaptive value of going from rough to smooth nutlets is unknown; however, rough nutlets may aid in dispersal by attaching to the outer surfaces of animals (Grau 1983). Rough nutlets as an ancestral feature agrees with the results of Hasenstab-Lehman and Simpson (2012); however, previous researchers believed that smooth nutlets were the ancestral condition (Payson 1927). Results found here and from Hasenstab-Lehman and Simpson (2012) argue that these earlier conjectures are unsupported.

Plant duration is recovered as being ancestrally annual for all major *Cryptantha* clades. A perennial duration is shown to have evolved once in the South American *Euryptantha/Geocarya* clade. The advantage of perennial plant duration may correlate with a high elevation habitat; however, more samples from South America would be needed to test this hypothesis. Early conjectures by Johnston (1925) and Higgins (1971) suggested that a perennial duration, which is found in all *Oreocarya*, was the ancestral condition for this complex. Results found here, however, agree with Hasenstab-Lehman and Simpson (2012) that an annual duration is ancestral for the complex.

Cleistogamy, a specialized type of self-pollination, evolved once in *Cryptantha*. The South American clade of taxa from sections *Eucryptantha* and *Geocarya* is distinguished in having cleistogamous flowers in either the middle and lower regions of inflorescence units of the plant (section *Eucryptantha*) or near the base of the plant with modified nutlets, the cleistogenes (section *Geocarya*). These specialized cleistogenes evolved a minimum of three times, as evidenced from this study. However, maximum likelihood reconstruction strongly supports normal cleistogamy evolving before clestogenes. A possible advantage of cleistogamy is the ability to produce offspring without the presence of pollinators. The South America clade characterized by cleistogamy corresponds with the first dispersal into South America. One possible explanation of this pattern is that the novel environment that the ancestor of this clade encountered lacked pollinators (at least initially), setting up the selective pressure for self-pollination.

Lastly, analysis of stem vestiture indicated that a spreading trichome orientation is ancestral, with many clades or species subsequently evolving appressed or both spreading and appressed trichomes. Spreading trichomes may aid in dispersal, as these trichomes are typically quite stout (hirsute to hispid), enabling whole plant segments to attach to a passing animal. The evolutionary advantage (if any) to having only appressed trichomes is unknown and seen in very few taxa.

BIOGEOGRAPHICAL INFERENCE AND DIVERGENCE TIME ESTIMATION

Four unidirectional dispersals of *Cryptantha* taxa from North to South American were recovered. This pattern of unidirectional dispersal from North to South agrees with studies of

other plant taxa that are amphitropically distrubuted (Moore et al. 2006). In the *Cryptantha* Core Clade, the first South America clade diversified around 23 Ma. This correlates with the average node age of the clade containing species in sections *Geocarya* and *Eucryptantha.* Within this clade there are multiple dispersals in the high elevation areas of the Andes and back. The first uplift of the Andes occurred around 20-30 Ma (Heibel and Renner 2012), resulting in the establishment of new topographic niches. Thus, dispersals into the newly uplifted Andes could be a potential causative factor in the diversification of this clade. Heibel and Renner (2012) proposed that the Mediterranean region of Chile acted as a refuge for species not able to adapt to harsh environments such as high elevation habitats or the hyper-aridity of the Atacama Desert. The one dispersal of *C. gnaphaloides* back to the Mediterranean South America region may provide additional support for this hypothesis.

The second South America clade of the *Cryptantha* Core Clade diversified around 4 Ma, roughly correlating with the second pulse of the Andean uplift (5-10 Ma; Heibel and Renner 2012). The ancestor of this group was also found in the Mediterranean South America region. Taxa belonging to this clade lack cleistogamic flowers and are more similar to the North American counterparts in section *Krynitzkia*. The common ancestor of two species, *C. peruviana* and *C. phaceloides*, was widespread in the Mediterranean region and tropical Andes, but since went subsequently extinct in Mediterranean South America and now extant taxa only occur in the tropical part of the Andes (Figure 21; Figure 22).

The Albidae Clade diversified around 17 Ma. Species within these more recent diversifications tend to be much more similar to their North American counterparts. Within the Albidae Clade, *C. albida* occurs in both North and South America, and thus the South American populations of this species may be indicative of a very recent dispersal event. The distribution of the North American species in tropical and subtropical deserts of North America may have pre-adapted these South American species for life in one of the driest region of the world, the Atacama Desert.

The diversification of the Maritimae Clade also occurred around 17 Ma. The common ancestor of this clade was found in the Mediterranean South America region. *Cryptantha maritima,* like *C. albida* in the Albidae Clade, occurs in the subtropical mountain regions of both North and South America. Other South America species in this clade occur in the Atacama region of Chile (*C. chaetocalyx*, not examined here, and *C. subamplexicaulis*), and

Peru (*C. granulosa*, not examined here). Although intriguing patterns of dispersal within South America emerge, these results should be considered preliminary. The addition of South American taxa that occur in the Atacama Desert in future analyses will contribute greatly to a better understanding of the history of this group, including evaluations of the hypothesis that these diversification events correlate with the hyperaridity of the Atacama Desert (10-15 Ma; Heibel and Renner 2012)

The program BAYAREALIKE in BioGeoBEARS (Matzke 2012, 2013) models a cladogenesis event that copies the ancestral range exactly for the daughter ranges, meaning that at species events, the range is unchanged. Although this model was chosen as the best fit model for the data, results should be interpreted with caution. It is surprising that the model chosen was without the founder event speciation ("J") option, as these analyses show that a founder event was likely an important part of the evolution of the South America taxa.

These biogeographic analyses should be viewed as approximate because of incomplete sampling of taxa or limited information available on species ranges. Species ranges for South America taxa were determined using data collected from visits to herbaria and from collections found online. Sometimes only one or two specimens were used to determine species ranges. These limitations may have had a significant effect on the results. As noted above, additional South America taxa should be added to include a more complete range of species occurrences. Two separate trips to South America were made in 2014 to collect South American *Cryptantha* taxa both from herbaria (MERL, SGO, SI) and from personal field collections. Although these collections significantly increased the sampling of South America species used in this study, some areas, such as the Atacama, where numerous species occur, have received little to no rain in recent years; other areas are very difficult to access. Another potential problem in the biogeographic analysis is misidentification of South American species in herbarium collections that were used to assess geographic ranges. All efforts went in to ensure correct identification for this study; however, even experts are in disagreement on identity of some South America taxa.

CONCLUSIONS

In conclusion, the genus *Cryptantha* is confirmed to be non-monophyletic, requiring changes to the current nomenclature. This study strongly supports the existence of three major *Cryptantha* clades, termed here the *Cryptantha* Core Clade, the Maritimae Clade, and the Albidae Clade. The fomer two clades largely correspond with, respectively, the *Cryptantha* s.s. 1 and s.s. 2 groups of Hasenstab-Lehman and Simpson (2012). However, the Albidae Clade is a new discovery of this study. The placement of these clades within the Amsinckiinae, however, varies in different analyses. Future nomenclatural changes, including the naming of one or more genera, will likely be needed.

Character analysis based on these phylogenetic studies indicates that the ancestral condition for *Cryptantha* was: 1) one to four nutlets per fruit; 2) nutlet homomorphism; 3) nutlets rough; 4) plants annual in duration; 5) flowers chasmogamous; and 6) stem trichomes spreading. The possible adaptive significance of these features is not always clear. However, it is likely that nutlet heteromorphism is related to more effective as a function of propagule dispersal. Cleistogamy (and its more specialized manifestation, cleistogenes), which occurs only in South American species, may function as a mechanism ensuring seed set in the absence of pollinators when these taxa were dispersed to a novel environment.

Four unidirectional dispersals from North to South America were recovered in the biogeographic analysis. Each of the three major *Cryptantha* clades contains at least one South America taxon, with strong support that dispersal has occurred unidirectionally from North to South America more frequently than previously thought. How these plants are dispersing to South America is still unknown. No known observations of birds feeding on or near plants have been documented. Migratory birds flying south, perhaps in a single uninterrupted flight, are still the best hypothesis to explain this pattern. Currently there are no known fossils of *Cryptantha* plant, nutlets, or pollen in the tropics, indicating that these species never occurred or could not establish there, supporting the hypothesis that the

amphitropical distribution is in fact caused by dispersal events, not a widespread population with subsequent extinction of species in the tropics.

Although the three major *Cryptantha* clades are consistently recovered, their placement in relation to one another and to other included genera is still preliminary. Future work must include additional representatives of all taxa in the Amsinckiinae to acquire strong support for these relationships in order to carry out complete taxonomic revisions. This study is a crucial first step in determining the sampling for these future studies. It also provides supported hypotheses for the dispersal patters of amphitropically distributed plants. Understanding the timing, direction, and frequency of dispersal between North and South America in *Cryptantha* gives insight to the origin of the great biodiversity of these regions and informs future studies on other species that share this distribution.

ACKNOWLEDGEMENTS

I would like to thank all Herbaria and staff that provided material for this study; Universidad de Concepcion, Chile (CONC), Gray Herbarium (GH), Instituto Argentino de Investigaciones de las Zonas Áridas, Argentina (MERL), Missouri Botanical Garden (MO), Rancho Santa Ana Botanical Garden (RSA), San Diego Natural History Museum (SD), San Diego State University (SDSU), Museo Nacional de Historia Natural, in Chile (SGO), Institutio de Botanica Darwinion, Argentina (SI), University of California (UC, JEPS), and University of Riverside (UCR).

I also thank the South American botanists; Gina Arancio, Victor Ardilles, Robeto Kiesling, Melica Munoz, Gloria Rojas, and Rosita Scherson who supported this project by collecting plants in the field, providing information on localities, and access to the herbarium collections.

To my advisor, Dr. Michael Simpson, thank you for your support, for sharing your enthusiasm and interest in Cryptantha with me, and for two memorable trips to South America.

I would like to thank my committee members, Dr. Marshal Hedin for his guidance in phylogenetic programs, and Dr. Robert Edwards, for his computer programming insights. I also thank the rest of the Evolutionary Biology staff: Dr. Tod Reeder for his assistance with phylogenetic programs, Dr. Annalisa Berta for her discussions on fossil dating, Dr. Kevin Burns for his support, Dr. Elizabeth Waters for her insights of molecular evolution, and Dr. John O' Leary in the Geography Department for assistance in finding the Global Ecological Zones to use in the biogeographic analyses.

I would also like to thank the evolutionary biology graduate students for their support, help, and advice. In the Simpson lab I thank Amanda Everett, Andy Siekkinen, and Nuri Pierce for their support and memories. I thank the "Crypts": Matt Guilliams, Kristen Hasenstab-Lehman, and Lee Ripma, for allowing me to tag along and for teaching me what it

takes to be successful; to my undergraduate research assistant, Regina Dowdy, thank you for helping with lab work, field work, and keeping me sane.

To my friends and family, thank you for your support and love these past years.

Lastly, I want to thank my funding sources for this study: the American Society of Plant Taxonomists, California Native Plant Society, Joshua Tree National Park, and San Diego State University Travel Grants.

REFERENCES

Aikaike, H. 1974. A new look at statistical model identification. *IEEE Transactions on Automatic Control* 19: 716-723.

Alvarez, I., and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417-434.

APGIII. 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.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L.Wheeler. 2005. GenBank. *Nucleic Acids Research* 33: D34-D38.

Brand, A. 1931. Borraginaceae-Borraginoideae-Cryptantheae. p. 22 in *Das Pflanzenreich 97*, ed. A. Engler. Leipzig: Verlag von Wilhelm Engelmann.

Boellstorff, J. 1976. The succession of late Cenozoic volcanic ashes of the Great Plains: a progress report. Pp. 37-71 in *24th annual meeting of the midwestern friends of the pleistocene guidebook,* ed. C. K. Bayne. Wichita: Kansas Geological Survey.

Boellstorff, J. 1978. Chronology of some late Cenozoic deposits from the central United States and the Ice Ages. *Transactions of the Nebraska Academy of Science* 6: 35-49.

Cohen, J. I. 2013. A phylogenetic analysis of morphological and molecular characters of Boraginaceae: evolutionary relationships, taxonomy, and patterns of character evolution. *Cladistics* 30: 1-31.

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

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

Elias, M. K. 1942. Tertiary prairie grasses and other herbs from the high plains. *Geological Society of America* 41: 176.

Forestry Department of the Food and Agriculture Organization of the United Nations. 2001. *Global ecological zoning for the global forest resources assessment*. Rome: FRA.

Grau, J. 1983. Life form, reproductive biology and distribution of the Californian/Chilean genus Cryptantha. Pp. 231-240 in *Sonderba¨nde des Naturwissenshaftlichen Vereins in Hamburg 7*, ed. K. Kubitski. Hamburg: P. Pary.

Hasenstab-Lehman, K. E., and M. G. Simpson. 2012. Cat's Eyes and Popcorn Flowers: phylogenetic systematics of the genus Cryptantha s. l. (Boraginaceae). *Systematic Botany* 37: 738-757.

Heibel, C., and S. S. Renner. 2012. Distrubution models and a dated phylogeny for Chilean Oxalis species reveal occupation of new habitats by different lineages, not rapid adaptive radiation. *Systematic Biology* 61(5): 823-834.

Higgins, L. C. 1971. A revision of Cryptantha subgenus Oreocarya. *Brigham Young University Science Bulletin Biological Series* 8:1-62.

Johnston, I. M. 1925. Studies in the Boraginaceae IV. The North American species of Cryptantha. *Contributions from the Gray Herbarium of Harvard University* 74: 1–114.

Johnston, I. M. 1927. Studies in the Boraginaceae VI. A revision of the South American Boraginoideae. *Contributions from the Gray Herbarium of Harvard University* 78: 1–118.

Johnston, I. M. 1961. Notes on some Texas borages. *Wrightia* 2: 158-162.

Kartesz, J. T. 2014. BONAP the biota of North America program. v. 1.0. Chapel Hill: Biota of North America Program.

Kay, K. M., J. B. Whittall, and S. A. Hodges. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology* 6: 36.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies, and A. Drummond. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12): 1647-1649.

Landis, M., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck. 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology* 62(6): 789-804.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29(6): 1695-1701.

Liu, L., L. Yu, D. K. Pearl, and S. V. Edwards. 2009. Estimating species phylogenies using coalescence times among sequences. *Systematic Biology* 58: 468-477.

Mabberley, D. J. 2008. *Mabberley's plant-book: a portable dictionary of the higher plants, their classification and uses*. Ed. 3. Cambridge: Cambridge University Press.

Maddison, W. P., and D. R. Maddison. 2010. Mesquite: a modular system for evolutionary analysis, v. 2.73. Vancouver: Mesquite Software, Inc.

Matzke, N. J. 2012. Founder-event speciation in BioGeoBEARS package dramatically improves likelihoods and alters parameter inference in Dispersal-Extinction-Cladogenesis (DEC) analyses. *Frontiers of Biogeography* 4(suppl. 1): 210.

Matzke, N. J. 2013. *BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R scripts*. Berkeley: University of California, Berkeley.

Mirarab, S., R. Reaz, M. S. Bayzid, T. Zimmerman, M. S. Swenson, and T. Warnow. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *ECCB* 30: 541-548.

Misawa, K., and K. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on a fast Fourier transformation. *Nucleic Acids Research* 30: 3059-3066.

Moore, M. J., A. Tye, and R. K. Jansen. 2006. Patterns of long-distance dispersal in Tiquilia subg. Tiquilia (Boraginaceae): implications for the origins of amphitropical disjuncts and Gala´pagos Islands endemics. *American Journal of Botany* 93: 1163-1177.

Nazaire, M., and L. Hufford. 2012. A broad phylogenetic analysis of Boraginaceae: implications for the relationships of Mertensia. *Systematic Botany* 37: 758-783.

Payson, E. B. 1927. A monograph of the section Oreocarya of Cryptantha. *Annals of the Missouri Botanical Garden* 14: 211-358.

Rambaut, A. 2014. FigTree, v. 1.4.2. Edinburgh: University of Edinburgh.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer, v. 1.6. Oxford: University of Oxford.

Raven, P. H. 1963. Amphitropical relationships in the floras of North and South America. *The Quarterly Review of Biology* 38: 151-177.

Raven, P. H., and D. I., Axelrod. 1974. Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden* 61: 539-673.

Rieseberg, L. H., and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Trends in Plant Evolution* 5: 65-84.

Rieseberg, L. H., and J. F. Wendel. 1993. Introgression and its consequences in plants. Pp. 70-109 in *Hybrid zones and the evolutionary process*, ed. R. Harrison. Oxford: Oxford University Press.

Ripma, L., M. G. Simpson, and K. Hasenstab-Lehman. 2014. Geneious! Simplified genome skimming methods for phylogenetic systematic studies: a case study in Oreocarya (Boraginaceae). *Applications in Plant Sciences* 2(12): 1-12.

Ree, R. H., and S. A. Smith. 2008. Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4-14.

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

Schmieder R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863-864.

Segal, R. 1964. Nomenclatural changes in fossil species of Cryptantha. *Transactions of the Kansas Academy of Science* 67: 203.

Segal R. 1966. Taxonomic study of the fossil species of the genus Cryptantha (Boraginaceae). *The Southwestern Naturalist* 11(2): 205-210.

Simpson, M. G., and K. E. Hasenstab. 2009. Cryptantha of southern California. *Crossosoma* 35: 1-59.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57: 758-771.

Straub, C. K., M. Fishbein, T. Livshultz, Z. Foster, M. Parks, K. Weitemier, and R. Cronn. 2011. Building a model: developing genomic resources for common milkweed (Asclepias syriaca) with low coverage genome sequencing. *BMC Genomics* 12: 211.

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

Weigend, M., F. Luebert, F. Selvi, G. Brokamp, and H. Hilger. 2013. Multiple origins for Hound's tonges (Cynoglossum L.) and Navel seeds (Omphalodes Mill.): the phylogeny of the borage family (Boraginaceae s.str). *Molecular Phylogenetics and Evolution* 68: 604-618.