Copyright

by

Michael James Moore

2005

The Dissertation Committee for Michael James Moore certifies that this is the approved version of the following dissertation:

Molecular Systematics of *Tiquilia* (Boraginaceae): Age, Origin, Dispersal History, and Gypsophilic Evolution

Committee:

____
Robert K. Jansen, Supervisor

____
Donald A. Levin, Co-Supervisor

____
C. Randal Linder

____
Alan M. Lloyd

____
Ulrich Mueller

____
Beryl B. Simpson

Molecular Systematics of *Tiquilia* (Boraginaceae): Age, Origin, Dispersal History, and Gypsophilic Evolution

by

Michael James Moore, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2005

Acknowledgments

First, I want to thank my adviser, Bob Jansen, for initially suggesting gypsophily as a potential dissertation topic. I also thank him for his support in all of its various forms, but especially for his willingness to help at a moment's notice. He has been an excellent adviser. I also want to thank my other committee members, Beryl Simpson, Randy Linder, Ulrich Mueller, and Alan Lloyd, for their support and direction. I also thank Billie Turner for his help and his constant willingness to discuss *Tiquilia*, and Jack Neff for informing me about *Tiquilia* bees.

I thank Alan Tye and the staff of the Charles Darwin Research Station for helping me obtain a collecting permit in the Galápagos Islands, and for sending me leaf material and voucher specimens of Galápagos *Tiquilia*. I also thank Dr. Tye for helping me sort out the confusing nomenclature of Galápagos *Tiquilia*. I also want to thank Asunción Cano and the staff at USM in Lima for their tremendous help in obtaining a collection permit in Perú and for facilitating my work while I was in the country. I also thank Fátima Cáceres for her invaluable help in collecting *Tiquilia* around Arequipa, and Mónica Arakaki for putting me in touch with Dr. Cáceres. I am also indebted to Blanca León and Martín Timaná for their wonderful help in preparing my Peruvian collecting permit, and for suggesting Dr. Cano as a contact. I also thank Helen Cortés-Burns, who singlehandedly made it possible for me to give a seminar in Spanish. I am also grateful to all the others who provided leaf material or DNA, including Dick Olmstead, Marc Gottschling, José Panero, Tim Chumley, and Barbara Goodson. I also want to thank

those who accompanied me on my various collecting trips to Argentina and to México, including Tim Chumley, Barbara Goodson, Geoff Denny, and Sarah Taylor. I also thank Tom Wendt and the staff of Plant Resource Center for their help.

I am also indebted to all the members of the plant systematics group here at UT, including all the graduate students past and present who have helped me along the way. I am particularly grateful to Andy Alverson for his help in the lab and on the computer over the past year, without which I could not have finished. I also thank all the other fellow members of the Jansen, Theriot, Simpson, and Linder labs through the years, including (in no particular order) Jan Barber, Les Goertzen, Brian Vanden Heuvel, Andrea Weeks, Heidi Meudt, Jennifer Tate, Mario McDermott, Tim Chumley, Mary Guisinger, Barbara Goodson, Romey Haberle, Elizabeth Ruck, Ruth Timme, Josh McDill, Sarah Taylor, Anneke Padolina, Joanne Birch, Leanne Kolnick, and Cynthia Peñaflor.

I am also deeply indebted to my parents for always providing an environment in which I could succeed and for encouraging me every step of the way. Finally, and most importantly, I especially thank my wife, Melissa: without her love, encouragement, understanding, and selflessness, none of this would have been possible.

This work was funded in part through much-appreciated grants from the National Science Foundation (through a Dissertation Improvement Grant), the American Society of Plant Taxonomists, the Botanical Society of America, Sigma Xi, and the Plant Biology Graduate Program at the University of Texas.

Molecular Systematics of *Tiquilia* (Boraginaceae): Age, Origin, Dispersal History, and Gypsophilic Evolution

Publication No. _________

Michael James Moore, Ph.D.

The University of Texas at Austin, 2005

Supervisors: Robert K. Jansen and Donald A. Levin

Tiquilia Pers. is a genus of approximately 30 species of subshrubs that grow in all deserts of North and South America, as well as in the Galápagos Islands. The genus is divided into two subgenera, subg. *Tiquilia* and subg. *Eddya*, that differ in morphology, chromosome number, habitat, and geographic distribution. Three distinct evolutionary themes, including the origins of North American desert plant lineages, the origins of North/South American amphitropical disjunctions, and the evolution of obligate gypsophily in plants of the Chihuahuan Desert Region (CDR) are explored with a molecular phylogenetic approach using *Tiquilia* as a case study. DNA sequence data from *ndhF*, *matK*, *rps16*, ITS, and *waxy* were collected for 28 species of *Tiquilia* and three outgroups. Phylogenetic analyses strongly support the monophyly of both subgenera as well as seven major lineages within *Tiquilia*, and indicate that the genus, subgenera, and major lineages evolved in North America. Molecular dating analyses

suggest that *Tiquilia* arose in the Eocene, that both subgenera and all major lineages diverged in the Miocene, and that species diversity within each major lineage is of late Tertiary age. The phylogenetically isolated position of *Tiquilia* within the Boraginales in conjunction with these diversification dates supports the hypotheses of earlier researchers concerning the age and origins of the characteristic North American desert flora. Phylogeographic analyses of subg. *Tiquilia* require at least three separate long-distance dispersal events, all originating in North America, to explain the distribution of subg. *Tiquilia* in North and South America, contributing to a growing body of evidence that intercontinental dispersal has been more common than previously realized. The continental origins of the four Galápagos endemic taxa of subg. *Tiquilia* are unresolved. Phylogenetic analyses of subg. *Eddya* imply two origins of obligate gypsophily in the subgenus. The widespread obligate gypsophile *T. hispidissima* possesses the highest level of intraspecific sequence diversity in the genus, supporting the relatively great age of obligate gypsophily in this species and in the CDR. The pattern of geographic variation in this species correlates well with geographic variation in other CDR gypsophilic plant groups, suggesting the existence of broader phylogeographic patterns among CDR gypsophiles.

Table of Contents

List of Tables x
List of Figures xi
Chapter 1. Introduction 1
Chapter 2. Molecular evidence for the age and origin of the American desert plant genus *Tiquilia* (Boraginaceae) 6
- Introduction 6
- Materials and Methods 12
  - Taxon and marker selection—*Tiquilia* 12
  - Taxon selection—outgroups 14
  - DNA extraction, amplification, sequencing, and alignment 15
  - Phylogenetic analyses 18
  - Test of alternate topologies 19
  - Molecular-based dating 20
- Results 23
- Discussion 26
  - Phylogenetic relationships of *Tiquilia* 26
  - Molecular dating in the Boraginales 28
  - The role of aridity in the evolutionary history of *Tiquilia* 30
- Conclusions 39

Chapter 3. Patterns of long-distance dispersal in *Tiquilia* subg. *Tiquilia* (Boraginaceae): implications for the origins of amphitropical disjuncts and Galápagos Islands endemics 53
- Introduction 53
- Materials and Methods 58
  - Taxon and marker selection 58
  - DNA extraction, amplification, sequencing, and alignment 60
  - Phylogenetic analyses 61
  - Tests of topologies 64
  - Biogeographic analyses 67
- Results 67
  - Results of individual marker analyses (*rps16*, ITS, *waxy*) 68
  - Results of combined marker analyses and tests of topologies 71
  - Results of biogeographic analyses 72
- Discussion 73
  - Origin and history of dispersal in subg. *Tiquilia* 73
    - Dispersal in *Tiquilia nuttallii* 74
    - Dispersal in the *Tiquilia palmeri* clade 75
    - Dispersal in the *Tiquilia plicata* clade 78
  - Timing of dispersal events in subg. *Tiquilia* 84

Conclusions 86
Chapter 4. The origins and biogeography of obligate gypsophily in the Chihuahuan Desert plant group *Tiquilia* subg. *Eddya* (Boraginaceae) 102
Introduction 102
Materials and Methods 109
Taxon and marker selection 109
DNA extraction, amplification, sequencing, and alignment 110
Phylogenetic analyses 111
Test of alternate topologies 116
Results 117
Phylogenetic relationships and evolution of gypsophily in subg. *Eddya* 118
Relationships within the *Tiquilia gossypina* clade 119
Relationships within the *Tiquilia hispidissima* clade 123
Discussion 127
Phylogenetic relationships in subg. *Eddya* and origins of gypsophily 127
Relationships among the facultative gypsophiles within the *Tiquilia gossypina* clade 128
Obligate gypsophily in the *Tiquilia gossypina* clade 132
Relationships within the *Tiquilia hispidissima* clade 139
Origin and geographic pattern of gypsophily in the *Tiquilia hispidissima* clade 142
Age and geographic patterns of gypsophily in the Chihuahuan Desert Region 146
Conclusions 152
Appendix A. Preliminary phylogenetic analyses of *rps16*, ITS, and *waxy* data for all accessions of *Tiquilia* 178
Appendix B. Alignment of *ndhF* sequence data 184
Appendix C. Alignment of *matK* sequence data 220
Appendix D. Alignment of *rps16* intron sequence data 237
Appendix E. Alignment of ITS sequence data 290
Appendix F. Alignment of *waxy* sequence data 366
References 411
Vita 428

List of Tables

Table 2.1. List of accessions included in *ndhF* and 5-marker combined phylogenetic analyses. 41
Table 2.2. List of primers utilized in this study. 44
Table 2.3. Summary of data set characteristics. 46
Table 2.4. Crown ages of nodes derived from the r8s molecular dating analyses, along with standard deviations derived from parametric bootstrap analyses. 47
Table 3.1. List of accessions included in phylogenetic analyses. 89
Table 3.2. Summary of data set characteristics. 92
Table 4.1. List of accessions included in phylogenetic analyses. 154
Table 4.2. Summary of data set characteristics. 159

List of Figures

Fig. 2.1. Phylogram of the tree resulting from the maximum likelihood search of the 5-marker combined data set (lnL = -19287.127), with a reconstruction of dispersal history mapped onto the tree....................49
Fig. 2.2. Phylogram of the tree resulting from the maximum likelihood search of the *ndhF* data set (lnL = -11529.351), showing major lineages within *Tiquilia* and the Boraginales..50
Fig. 2.3. Non-rejected SH test topologies, reduced to the seven major lineages of *Tiquilia* and two outgroups...51
Fig. 2.4. Chronogram of the *ndhF* lnL tree derived from the penalized likelihood analysis, showing major lineages within *Tiquilia* and the Boraginales. 52
Fig. 3.1. Distribution of major species and species groups in *Tiquilia* subg. *Tiquilia*...93
Fig. 3.2. Phylogram of the tree resulting from the maximum likelihood search of the 5-marker combined data set (lnL = -19287.127), with a reconstruction of dispersal history mapped onto the tree....................94
Fig. 3.3. Phylogram of the tree resulting from the maximum likelihood search of the combined cpDNA data (lnL = -13079.277), with a reconstruction of dispersal history mapped onto the tree....................95
Fig. 3.4. Phylogram of the tree resulting from the maximum likelihood search of the combined nrDNA data (lnL = -7751.087), with a reconstruction of dispersal history mapped onto the tree. 96
Fig. 3.5. Phylogram of one of three MPTs (TL = 195, RCI = 0.834, RI = 0.962) derived from parsimony analysis of the *rps16* data set for subg. *Tiquilia*...97
Fig. 3.6. Phylogram of one of 43,658 MPTs (TL = 151, RCI = 0.899, RI = 0.976) derived from parsimony analysis of the ITS data set for the *T. plicata* clade...98
Fig. 3.7. Phylogram of one of 378 MPTs (TL = 135, RCI = 0.953, RI = 0.990) derived from parsimony analysis of the ITS data set for *T. nuttallii* and the *T. palmeri* clade...................................99
Fig. 3.8. Phylogram of one of 1512 MPTs (TL = 444, RCI = 0.677, RI = 0.890) derived from parsimony analysis of the *waxy* data set for subg. *Tiquilia*...100
Fig. 3.9. Most parsimonious reconstructions of dispersal history in the *T. plicata* clade. 101
Fig. 4.1. Map of parts of the southwestern United States and northern México, showing collection localities for accessions of the *T. hispidissima* clade that were included in phylogenetic analyses.........161

Fig. 4.2. Map of parts of southern New Mexico, west Texas, and northern México, showing collection localities for accessions of the *T. gossypin*a clade that were included in phylogenetic analyses............162
Fig. 4.3. Phylogram of one of three MPTs (TL = 261, RCI = 0.766, RI = 0.943) derived from parsimony analysis of the 56-accession *rps16* data set, showing subgenera of *Tiquilia* and the four major lineages of subg. *Eddya*...163
Fig. 4.4. Phylogram of one of 1176 MPTs (TL = 775, RCI = 0.593, RI = 0.902) derived from parsimony analysis of the 56-accession ITS data set, showing subgenera of *Tiquilia* and the four major lineages of subg. *Eddya*...164
Fig. 4.5. Phylogram of one of 213 MPTs (TL = 731, RCI = 0.639, RI = 0.907) derived from parsimony analysis of the 56-accession *waxy* data set, showing subgenera of *Tiquilia* and the four major lineages of subg. *Eddya*...165
Fig. 4.6. Phylogram of the tree resulting from the maximum likelihood search of the 5-marker combined data set (lnL = -19287.127), with a reconstruction of gypsophily in subg. *Eddya* mapped onto the tree...166
Fig. 4.7. Non-rejected SH test topologies for subg. *Eddya*, reduced to the seven major lineages of *Tiquilia* and two outgroups............................167
Fig. 4.8. Phylogram of the single MPT (TL = 37, RCI = 0.961, RI = 0.987) derived from a parsimony analysis of the *rps16* data set for the *T. gossypina* clade...168
Fig. 4.9. Phylogram of one of 3343 MPTs (TL = 209, RCI = 0.641, RI = 0.854) derived from a parsimony ratchet analysis of the full ITS data set for the *T. gossypina* clade. 169
Fig. 4.10. Phylogram of one of 7835 MPTs (TL = 203, RCI = 0.609, RI = 0.793) derived from a parsimony analysis of the reduced ITS data set for the *T. gossypina* clade...................................170
Fig. 4.11. Phylogram of one of 757 MPTs (TL = 183, RCI = 0.892, RI = 0.984) derived from a parsimony ratchet analysis of the *waxy* data set for the *T. gossypina* clade...171
Fig. 4.12. Illustration of *waxy* sequence polymorphism in directly sequenced PCR product of accessions of *T. mexicana* that are sympatric with *T. gossypina*..172
Fig. 4.13. Phylogram of the single MPT (TL = 37, RCI = 0.916, RI = 0.969) derived from a parsimony analysis of the *rps16* data set for the *T. hispidissima* clade, with population groups from the ITS analyses indicated for reference. 173
Fig. 4.14. Phylogram of one of 2715 MPTs (TL = 336, RCI = 0.598, RI = 0.885) derived from a parsimony ratchet analysis of the full ITS data set for the *T. hispidissima* clade.......................................174

Fig. 4.15. Phylogram of one of 2783 MPTs (TL = 326, RCI = 0.574, RI = 0.846) derived from a parsimony ratchet analysis of the reduced ITS data set for the *T. hispidissima* clade. 175
Fig. 4.16. Phylogram of one of 2090 MPTs (TL = 129, RCI = 0.869, RI = 0.958) derived from a parsimony ratchet analysis of the *waxy* data set for the *T. hispidissima* clade, with population groups from the ITS analyses indicated for reference. 176
Fig. 4.17. Phylogram of one of three MPTs (TL = 206, RCI = 0.686, RI = 0.736) derived from a parsimony analysis of combined *rps16*, ITS, and *waxy* data for a representative set of *T. hispidissima* accessions. 177
Fig. A1. Phylogram of one of 7 MPTs derived from parsimony analysis of the *rps16* data set for all outgroups (including *Pholisma*) and accessions of *Tiquilia*, showing major lineages and subgenera of *Tiquilia*. 179
Fig. A2. Phylogram of one of > 100,000 MPTs derived from parsimony analysis of the ITS data set for all accessions of *Tiquilia*, showing major lineages and subgenera of *Tiquilia*. 180
Fig. A3. Phylogram of one of > 100,000 MPTs derived from parsimony analysis of the waxy data set for all accessions of *Tiquilia*, showing major lineages and subgenera of *Tiquilia*. 182

Chapter 1. Introduction

Tiquilia Pers. is a genus of approximately 30 species of annual to perennial subshrubs that inhabit every desert in North and South America, as well as the Galápagos Islands. *Tiquilia* occupies an isolated position within subfamily Ehretioideae of Boraginaceae. It shares a bifid style with all members of this subfamily, but differs from all other Ehretioids by possessing nutlets for fruits instead of drupes, by inhabiting arid temperate and subtropical regions vs. semi-arid to wet tropical regions, and by its prostrate or subshrubby habit vs. shrubs to trees in other Ehretioideae (Richardson 1977). These differences have made it difficult to determine the placement of *Tiquilia* within the Ehretioideae, and they suggest a potentially old age for the genus.

The genus has been monographed by Richardson (1977), who divides *Tiquilia* into two subgenera that differ in morphology, chromosome number, habitat, and geographic distribution. Subgenus *Eddya* A.T. Richardson is centered in the Chihuahuan Desert Region of México and the United States and contains approximately ten species that are restricted to calcareous and gypseous substrates. Three of the ten species are obligate gypsophiles, while the remaining species are found on both limestone- and gypsum-derived substrates. In contrast, the approximately 20 species of subg. *Tiquilia* are found in loose sandy soils in the Great Basin, Mojave, and Sonoran Deserts of North America, the Monte Desert of Argentina, the coastal deserts of Perú and Chile, and the Galápagos Islands. Four of these species are endemic to the Sonoran and Mojave

Deserts, approximately 11 species are endemic to the coastal deserts of South America, four species are restricted to the Galápagos Islands. One species, *T. nuttallii* (Benth.) A.T. Richardson, is amphitropically disjunct; it grows throughout the Great Basin of western North America and in the Monte Desert of Argentina.

Tiquilia thus affords the opportunity to examine a number of plant evolutionary and biogeographic problems, including the origin of a desert plant group, the evolution of obligate gypsophily in the Chihuahuan Desert Region, and the origin and dispersal patterns of an amphitropically distributed plant group. This dissertation examines all of these issues using a molecular phylogenetic approach. Sequence data were collected for approximately 200 accessions representing 28 of the roughly 30 species of *Tiquilia* for three markers: the chloroplast *rps16* intron (Oxelman et al. 1997, Kelchner 2002), the nuclear ribosomal internal transcribed spacer (ITS; Baldwin et al. 1995), and part of the nuclear granule-bound starch synthase gene (GBSSI, or *waxy*; Mason-Gamer et al. 1998). These data were used to select 25 exemplar accessions, representing a cross-section of diversity in the genus, to sequence two additional chloroplast markers, the *ndhF* (Olmstead and Sweere 1994) and *matK* genes (including the 3' end of the *trnK* intron; Steele and Vilgalys 1994, Johnson and Soltis 1994, Kelchner 2002). All sequence data are listed in Appendices B-F. Sequence data and alignments will be deposited in public-access databases prior to publication of each chapter in scientific journals. Therefore GenBank numbers are not included in this dissertation.

Chapter 2 tests the hypotheses of Axelrod (1950, 1979a,b) concerning the origins of the arid-adapted floras of North America. Axelrod has shown from fossil and geologic

evidence that the desert habitats of North America are very young while the arid-adapted plants that inhabit them must be much older. He has suggested that the modern desert flora developed autochthonously in local dry pockets in western North America as the climate aridified during the Cenozoic. He posits that plants began adapting to drought in western North America as early as the Cretaceous, and subsequently have become increasingly drought tolerant as the Earth has cooled and dried since the end of the Eocene. Thus many of the desert-dwelling plant lineages of today have deep roots, despite the youth of the desert ecosystem. *Tiquilia* is well-suited to the study of this problem, because it is likely of North American ancestry, is widespread and common in desert habitats today, and is morphologically isolated within subfamily Ehretioideae, whose other members do not inhabit arid lands. To determine if *Tiquilia* follows the pattern of evolution implied by Axelrod, a well-supported molecular phylogeny of the genus and three outgroups was constructed using all 5 markers. This phylogeny was congruent with the *ndhF* phylogeny using the same accessions, and because a wide range of Boraginales *ndhF* sequences are available on GenBank, a phylogeny of Boraginales was developed for this marker that included all major lineages of *Tiquilia*. Divergence times for the various nodes in the tree were then estimated with the resulting phylogeny using a penalized likelihood approach. Although *Tiquilia* lacks a fossil record, a time calibration point near the base of the tree was selected from a recent, well-dated phylogeny of asterids (Bremer et al. 2004). From this analysis the probable age of the genus, the subgenera, and the major lineages could be roughly estimated, and any

correlation between speciation and important drying events during the Cenozoic could be determined.

Chapter 3 examines the origins of the amphitropical disjunction and the Galápagos endemics in subg. *Tiquilia*. Disjunction between the arid regions of North and South America is a common feature of the floras of these two continents, and it includes disjunctions involving individual species or closely related species to entire plant groups (Raven 1963). Several recent molecular phylogenies of these disjunct plant groups have suggested that multiple dispersals between North and South America might be common (Miller 2002b, Ickert-Bond and Wojciechowski 2004, Simpson et al. 2004, 2005), and that they probably have not occurred simultaneously, as would be expected under a long-distance dispersal hypothesis. Subgenus *Tiquilia* is well-suited to the study of American amphitropical disjunction because it possesses a number of endemic species in both North and South America, as well as one species that is itself disjunct between these regions. The South American taxa are unusual because they all appear to be tetraploid, and all but one of these species form a morphologically natural group that is characterized by blue flowers. The biogeography of subg. *Tiquilia* is further complicated by the presence of four endemic taxa in the Galápagos Islands. Sequence data were generated for all North American and most South American species, as well as all Galápagos taxa. Phylogenetic analysis of this data allowed the determination of the monophyly of the South American taxa and the Galápagos taxa, as well as the direction and number of long-distance dispersal events that are required to explain the current distribution of subg. *Tiquilia*.

Chapter 4 tests the hypotheses of earlier researchers concerning the age of the gypsophilic flora of the Chihuahuan Desert Region (CDR), and also examines the biogeographic patterns of gypsum endemism in this region. The CDR is home to an extensive gypsophilic flora. This flora includes a number of genera with multiple obligately gypsophilic taxa, as well as a number of obligately gypsophilic species that are distributed across large portions of the CDR in spite of the isolated nature of many gypsum deposits. For these reasons earlier researchers have suggested a relatively great age for plant gypsophily in the CDR (Johnston 1941, Turner and Powell 1979). *Tiquilia* subg. *Eddya* provides a good opportunity to test these hypotheses. Three of the ten species in the subgenus are obligate gypsophiles: one species [*T. hispidissima* (Torr. & A. Gray) A.T. Richardson] is found over most of the northern half of the CDR, while the remaining two are closely related narrow endemics of the Cuatro Ciénegas region of eastern Coahuila and northwestern Nuevo León. To examine the patterns of gypsophilic evolution among these species, populations of these three species and their hypothesized close relatives were sampled from across their ranges, and sequences were derived for *rps16*, ITS, and *waxy* for all accessions. From these data a rough idea of the amount of sequence variation could be determined for each of the gypsophiles, along with any potential correlation among sequence variation and geography for the widespread gypsophile *T. hispidissima*. This information was then used to help roughly date the possible emergence of obligately gypsophilic lineages in subg. *Eddya*. These data were also compared to the distributions of other gypsophilic lineages in the region in order to form a more general picture of the patterns of gypsum endemism in the CDR.

Chapter 2. Molecular evidence for the age and origin of the American desert plant genus *Tiquilia* (Boraginaceae)

Introduction

The warm deserts of the world are considered to be of a relatively young age, in most cases no older than the late Miocene/early Pliocene [7-5 Ma (million years ago); all dates follow Berggren et al. 1995, unless otherwise noted], and in some cases, as for example the North American deserts, as young as the early to late Quaternary (1.75 Ma to the present; Axelrod 1979a,b, Wolfe 1985, Graham 1999). Despite the youth of these ecosystems, each of the warm desert regions contains a large endemic flora with many unique arid-adapted morphotypes which suggests that such plants have evolved much earlier than the desert regions they inhabit (Axelrod 1958, 1979a,b). The deserts of southwestern North America furnish numerous examples of characteristic genera that likely possess relatively great antiquity, including *Fouquieria* HBK. (Fouquieriaceae), *Simmondsia* Nutt. (Simmondsiaceae), *Koeberlinia* Zucc. (Koeberliniaceae), *Pachycormus* Coville (Anacardiaceae), *Agave* L. (Agavaceae), *Yucca* L. (Agavaceae), *Olneya* A. Gray (Fabaceae), and several genera of large columnar cacti (such as *Carnegiea* Britton & Rose; Cactaceae).

Axelrod (1950, 1979a,b) was the first researcher to fully recognize and attempt to account for the difference in age between the North American deserts (including the Mojave, Sonoran, and Chihuahuan Deserts) and their characteristic plant groups. He has

developed a general evolutionary scenario, based on the analysis of fossil floras in conjunction with paleoclimatic reconstruction, that posits the autochthonous evolution of a drought-adapted flora drawn from local semi-arid or arid sites that existed among more mesic habitats (Axelrod 1967, 1972, 1979a). Specifically, he suggests that the earliest adaptations to seasonally dry climates began as early as the late Cretaceous (99-65 Ma; Upper Cretaceous of Gradstein et al. 1995), possibly in areas of edaphic aridity, within the widespread tropical forests and seasonally dry woodlands that then characterized the continent (Axelrod 1950, 1958, 1972, 1979a). By the late Eocene (approx. 37-33.5 Ma), plants typical of drier vegetation types such as sclerophyll woodland and arid tropic scrub had arisen and occupied at least some portion of the interior of southwestern North America (Axelrod 1958, 1979a, Leopold et al. 1992). A series of global climatic changes near the Eocene/Oligocene boundary (~33.5 Ma), however, resulted in a dramatic cooling and drying of the Earth, greatly increasing the extent of woodland and savanna habitats (although probably not grass-dominated savanna as is common at present) in western North America (Axelrod 1979a, Wolfe 1997, Graham 1999). While climate conditions fluctuated somewhat between the early Oligocene and the mid-Miocene (approx. 33.5-15 Ma; Wolfe 1997, Graham 1999), they remained wetter than at present; there is no evidence of widespread arid or semi-arid habitats during this time period, although they probably existed very locally under certain topographic or edaphic conditions (Axelrod 1979a). The mid-Miocene (~15 Ma) ushered in another period of global cooling and drying, and it is in this era that evidence of grass-dominated habitats first appears in North America (Leopold et al. 1992, Retallack 1997, Jacobs et al. 1999). Drier

environments such as these spread gradually throughout the late Miocene, until another sharp drop in temperature and concomitant increase in aridity near the end of the Miocene (~5 Ma) caused the spread of semi-desert conditions into at least some of the areas of the modern North American deserts (Axelrod 1950, 1979a; Graham 1999). However, even during the relatively dry periods of the Pliocene (approx. 5-1.75 Ma), true desert vegetation may have been no more than a local phenomenon (Axelrod 1979a). It was not until the glacial climate upheavals of the Pleistocene that deserts became widespread, and then only during the drier interglacials; there is abundant evidence that desert vegetation was greatly restricted during the pluvial periods (Wells 1966, Axelrod 1979a, Betancourt et al. 1990, Lowenstein et al. 1999, Thompson and Anderson 2000). The modern extent of desert vegetation in North America may then be the widest it has ever been (Axelrod 1979a, Betancourt et al. 1990).

From this summary, it is clear that climate in western North America has cooled and aridified in a stepwise fashion since the Eocene. Thus, each of the major increases in aridity during the Cenozoic would not only have stimulated the evolution of increasingly arid-adapted plants; each increase would have also favored the spread of those taxa already adapted to the new level of aridity. Despite the absence of the desert ecosystem in North America before the Pleistocene, it seems clear from fossil evidence and from the sheer mophological divergence of the plant groups involved that a significant drought-adapted flora had already accumulated in locally arid and semi-arid regions by the late Miocene/early Pliocene. This flora comprised all of the endemic plant groups that are considered characteristic of North American deserts (including all of the genera listed

above; Axelrod 1950, 1979a) and was able to spread quickly across the newly arid landscapes of Pleistocene interglacial periods.

Recent years have witnessed a remarkable increase in the use of fossil information to calibrate the ages of plant groups in molecular-based phylogenies (Sanderson et al. 2004), and while many of these studies have focused on correlating past cladogenetic events with significant climate changes or plate movements, very few have attempted to examine the evolution of aridland plant groups. For instance, a number of studies have attempted to test important biogeographic hypotheses, such as Gondwanan vicariance (examples include Davis et al. 2004 and Weeks et al. 2005) and the ages of the vicariant plant groups of the northern temperate forests (Wen 1999, Donoghue and Smith 2004), using molecular-based phylogenies in groups with good fossil records. Such studies have been completed for a few plant taxa in various arid regions of the world, such as in South Africa (Caujapé-Castells et al. 2001, Linder 2003, Linder and Hardy 2004, Klak et al. 2004) and Australia (Crisp et al. 2004), and for certain special problems such as the evolution of C_4 photosynthesis (Kadereit et al. 2003, Sage 2004). However, while there have been several recent molecular phylogenetic studies that have included endemic North American desert lineages, including Cactaceae (Butterworth et al. 2002, Nyffeler 2002), *Fouquieria* (Schultheis and Baldwin 1999), *Simmondsia* (Cuénoud et al. 2002), *Koeberlinia* (Hall et al. 2004), *Olneya* (Lavin et al. 2003), and Agavaceae (Bogler and Simpson 1995, 1996), none have attempted to date the evolutionary events inferred in their respective phylogenies.

How could we recognize that a given North American arid-adapted plant lineage evolved under the kind of stepwise aridification envisioned above by Axelrod, in the absence of fossil information? In other words, what kinds of patterns might we expect in a molecular-based phylogeny of such a desert plant group? Certain features of the phylogeny could be predicted under such a climatic scenario. First, the stem lineage of such a group should be relatively old and isolated phylogenetically with respect to its outgroups, due to its restriction to local arid pockets throughout much of its history. Second, it is likely that any of the deeper cladogenetic events within the ingroup would be correlated with one or more of the episodes of aridification that occurred during the Tertiary. Third, the species- or population-level diversity within these groups should be much younger, representing recent radiations in response to the spread of semi-arid and arid climates since the late Miocene. In order to test such hypotheses for a particular desert plant group, reliable dates would have to be inferred on a well-supported phylogeny of the group.

Tiquilia Pers. (Boraginaceae) provides a good opportunity to test these assumptions regarding evolution of North American aridland plants. The genus comprises approximately 30 species of prostrate herbs and subshrubs that inhabit the deserts of both North and South America (Richardson 1977). There are a number of reasons to suspect that *Tiquilia* may conform to the predictions outlined above. Although *Tiquilia* is clearly a member of subfamily Ehretioideae of Boraginaceae, its adaptation to aridity and its other unique morphological characters have left its position within the subfamily unclear (Al-Shehbaz 1991, Miller 2003). The genus is divided into two

subgenera that differ in a number of features, including morphology, chromosome number, substrate preference, and geographic distribution: subg. *Tiquilia* (~20 spp.) is found on sandy soils in the Sonoran, Mojave, and Great Basin Deserts of North America, as well as in the coastal deserts of Peru and the deserts of western Argentina, while subg. *Eddya* A.T. Richardson (~10 spp.) is entirely North American and is restricted to calcareous and gypseous substrates, with most species in the Chihuahuan Desert Region (Richardson 1977). Furthermore, the North American species of *Tiquilia* fall into several groups that are quite distinct morphologically; all of these species groups are quite common throughout the deserts and in many cases occur sympatrically without any evidence of hybridization. At first glance, then, *Tiquilia* would appear to be a good candidate group in which to test the hypotheses implied in Axelrod's work: it is almost certainly of North American origin; its isolated position within the Boraginaceae combined with its tremendous morphological diversity suggest a possible ancient age for the genus; and its fidelity to arid habitats together with its widespread and common nature in the deserts today suggest that it has benefited from the recent spread of the desert biome.

The purpose of the present paper, then, is to use several markers from the chloroplast and nuclear genomes to construct a well-supported phylogeny of *Tiquilia* and its potential outgroups, and to apply reasonable dates to the cladogenetic events in the resulting tree. From this information we can then test the predictions outlined above: (1) is *Tiquilia* a relatively old arid-adapted lineage that evolved in North America?; (2) when did the majority of diversification within the genus occur, and is any of this

diversification correlated with episodes of aridification during the Cenozoic?; and (3) is there any evidence of increased species- or population-level diversity with the onset of semi-arid and arid conditions since the late Miocene? Answers to such questions require both a well-supported phylogeny as well as reasonable calibration points to help date the evolutionary events in *Tiquilia*. In many molecular-based studies that seek to apply dates to specific events on phylogenetic trees, well-dated ingroup fossils with reasonably certain affinities to modern taxa are used as constraints on the ages of nodes. However, such an approach for *Tiquilia* is currently impossible, because of the complete absence of a fossil record for the genus. Such a situation is common in plant taxa of arid regions because the conditions required for fossil formation are usually not encountered in such areas (Axelrod 1979a,b). Nevertheless, it is still possible, although less desirable, to use other types of information, such as fossil-based ages from outgroups as well as known ages of past climatic or geologic events, to constrain the taxon ages in the phylogeny of a desert group. This approach has been used profitably in at least one other study involving the radiation of a drought-adapted plant group [in *Moraea* Mill. (Iridaceae); Goldblatt et al. 2002], and we have used a similar approach in the current study.

Materials and Methods

Taxon and marker selection—*Tiquilia*:

Preliminary phylogenetic analyses of 202 accessions representing 27 species of *Tiquilia* (see Appendix A) using sequences from the chloroplast *rps16* intron (Oxelman et al. 1997, Kelchner 2002), the nuclear ribosomal internal transcribed spacer (ITS; Baldwin et al. 1995, Álvarez and Wendel 2003), and part of the nuclear granule-bound starch

synthase gene (GBSSI, or *waxy*; Mason-Gamer et al. 1998) allowed the selection of a subset of 25 accessions representing all of the major lineages of *Tiquilia* for sequencing of two additional chloroplast markers, *ndhF* (Olmstead and Sweere 1994) and *matK* (including the 3' portion of the *trnK* intron; Steele and Vilgalys 1994, Johnson and Soltis 1994, Kelchner 2002). Using sequence data for all five markers for these 25 accessions, the incongruence length difference (ILD) test (Farris et al. 1994) was used to assess data combinability using each marker as a separate partition. Initial results of the ILD test suggested incongruence among three partitions ($p = 0.01$): chloroplast DNA (cpDNA), ITS, and *waxy*. The taxa causing the incongruence were identified and removed so that a reduced data set including just 14 species of *Tiquilia* (see Table 2.1), but still representing every major lineage in the genus, exhibited congruence among all data partitions. These 14 species were then included in the SH test of alternate topologies (see below).

The *ndhF* data set was selected for analyses involving dating of divergence events because the maximum likelihood topologies of the combined data tree and the *ndhF* tree were congruent with nearly identical relative branch lengths, and because a previous study of the Hydrophyllaceae had sampled the Boraginales widely for *ndhF* (Ferguson 1999). The *ndhF* analyses included all 14 species included in the combined analysis as well as an additional five species that were added to increase coverage within the more species-rich main lineages of *Tiquilia*, without adding zero length terminal branches (see below for molecular dating; see Table 2.1 for a list of all taxa included in these analyses). The addition of these five taxa did not significantly alter the topology or branch lengths with respect to the major lineages of *Tiquilia*.

Taxon selection—outgroups:

Four outgroups were initially chosen from among the Boraginales based on their potential affinities to *Tiquilia*: *Coldenia procumbens* was chosen because of earlier morphological hypotheses regarding it as the closest relative of *Tiquilia* (Richardson 1977); *Ehretia anacua* and *Bourreria succulenta* were chosen as representative taxa of the subfamily Ehretioideae, to which *Tiquilia* belongs (Richardson 1977, Miller 2003); and the holoparasite *Pholisma arenarium* Nutt. (Lennoaceae) was chosen because of some recent molecular work indicating a possible close relationship between it and *Tiquilia* (Smith 1998, Smith et al. 2000, Olmstead and Ferguson 2001). However, sequences derived from *P. arenarium* were always the most distant from *Tiquilia* in preliminary analyses (see Figure A1 for an example), and the amount of divergence present in these sequences was suggestive of an accelerated rate of evolution, a common feature of holoparasitic genomes (Nickrent et al. 1998, Barkman et al. 2004). Because of the problems such divergent sequences can pose in phylogenetic analysis (Nickrent et al. 1998, Stefanovic and Olmstead 2004), *Pholisma* was excluded from all analyses (although the sequences derived from *Pholisma* will be deposited in GenBank). The remaining three outgroups were included in the *ndhF* analyses, while only *Coldenia* and *Bourreria* were included in the SH tests to simplify the analyses.

In addition to these three outgroups, *ndhF* sequences of 19 additional outgroups were selected from the earlier work of Ferguson (1999) on Hydrophyllaceae (see Table 2.1). These taxa are representative of all the major groups of Boraginales, except Lennoaceae [although we had material of *Pholisma*, it lacks *ndhF* (Bremer et al. 2002)].

In addition, *ndhF* sequences representing several families of euasterids closely related to Boraginales were selected to provide a means for calibrating the stem lineage of Boraginales in the r8s analysis using the dates of Bremer et al. (2004; see Table 2.1).

DNA extraction, amplification, sequencing, and alignment:

All accessions of *Tiquilia* were field collected in silica gel, and vouchers deposited in TEX. Of the outgroups, *Ehretia anacua* was collected and extracted from fresh material at the University of Texas, while whole-genomic DNA of *Coldenia*, *Pholisma*, and *Bourreria* was provided by other labs (see Table 2.1). All accessions of *Tiquilia* and *Ehretia* were extracted using the 0.1 g protocol of the Nucleon Phytopure DNA extraction kit (Amersham Biosciences), with 10 μL of 2-mercaptoethanol added to each extraction.

Amplification of the five molecular markers utilized the primers listed in Table 2.2. The total *ndhF* region amplified includes the entire gene sequence with the exception of the first 23 bases at the 5' end, and the last 146 bases at the 3' end (relative to the sequence for *Nicotiana tabacum*). The *matK* primers amplified a region including the entire *matK* gene as well as the 3' end of the *trnK* intron. Both *ndhF* and *matK* were amplified in two separate, overlapping fragments, and were sequenced using 8 and 7 primers, respectively. All other markers were sequenced using the same primers used for amplification. *Waxy* is present in a single copy in most angiosperms (Mason-Gamer et al. 1998), although it is present in multiple copies in some groups (Evans et al. 2000, Winkworth and Donoghue 2004). It has been found to be useful for reconstructing phylogenies at the species level in several plant families, notably in Poaceae (Mason-

Gamer et al. 1998, Mathews et al. 2002, Ingram and Doyle 2004) and Solanaceae (Peralta and Spooner 2001, Walsh and Hoot 2001). This study represents the first use of this marker in the Boraginales. The *waxy* primers used in this study amplified a region approximately 600 bp long that is homologous to the region from exon 9 to exon 11 in *Ipomoea* (Convolvulaceae; Miller et al. 1999). However, the primers designed specifically for this study were located just inside the 5' and 3' primers of Miller et al. (1999).

A few taxa exhibited clearly evident sequence polymorphism in directly sequenced PCR product for ITS and/or *waxy*. These accessions were cloned using the TOPO TA kit (Invitrogen) with vector pCR 2.1-TOPO using one-third the recommended reaction volumes. For each cloning reaction, from 5 to 15 positively transformed colonies were reamplified and the products sequenced. In all but one case, cloned sequences from a given accession formed a monophyletic group and in these cases one cloned sequence per accession was included in the combined phylogenetic analyses. The cloned *waxy* sequences of the allotetraploid *Tiquilia elongata* fell into two monophyletic groups, the first nearly identical in sequence to *T. palmeri*, and the second clearly more distant and basal to the preceding clade. Because the other four markers agree with the phylogenetic position implied for *T. elongata* by the second group of *waxy* sequences, one of these sequences was selected for the combined analyses in this paper. Cloned accessions are listed in Table 2.1.

PCR volumes of 25 μL included 10-100 ng of template DNA, 12.5 μL of FailSafe PCR 2X Premix J (Epicentre), 0.4 mM of each primer, and 0.5 U *Taq* polymerase. ITS

and the *rps16* intron were amplified using an initial denaturation of 96°C (3 min), followed by 36 cycles of 94°C (1 min), 54°C (1 min), and 72°C (45 sec + 3 sec/cycle), followed by a final extension at 72°C (7 min). Amplification of *waxy* used the same program parameters but with a 50°C annealing temperature. Amplification of *ndhF* and *matK* proceeded as follows: an initial denaturation of 96°C (2 min 30 sec), followed by 35 cycles of 94°C (1 min), 50°C (1 min), and 72°C (2 min), followed by a final extension at 72°C (15 min). PCR parameters were identical for cloned template except for the substitution of an initial 10 min hot start denaturation at 95°C, followed by the addition of *Taq* polymerase.

PCR products were cleaned using QiaQuick columns (Qiagen) and the amount of product was quantified using agarose gel electrophoresis with a low mass DNA ladder. Cycle sequencing reactions included 20-40 ng of template DNA, 2 μL Big Dye terminator (Perkin Elmer), and 0.5 mM of each primer per 20 μL reaction volume. The cycle sequencing program included an initial denaturation of 96°C (2 min) followed by 26 cycles of 96°C (10 sec), 50°C (5 sec) and 60°C (4 min). Samples were then cleaned using Centri-Sep columns (Princeton Separations) packed with G-50 Sephadex (Amersham Biosciences), and then sequenced using an MJ Research BaseStation automated sequencer.

Raw sequences were trimmed and edited using Sequencher v. 3.0 and v. 4.0 (GeneCodes). Sequences from each marker were aligned initially using ClustalX (Thompson et al. 1997), and the resulting alignments were adjusted manually using SeqApp (Gilbert 1992), with the exception of the *ndhF* data set, which was aligned

manually. Several short regions of ITS (amounting to 13.7% of the total ITS alignment; see Appendix E) that were difficult to align among the major lineages of *Tiquilia* and/or the outgroups were eliminated from the combined analyses. All sequences will be deposited in GenBank, and the final alignments will be made available at TreeBASE (http://www.treebase.org).

<u>Phylogenetic analyses</u>:

Parsimony and maximum likelihood (ML) analyses were conducted on both the 5-marker combined data set and the *ndhF* 41-taxon data set in PAUP* v. 4.0b10 (Swofford 2002) using heuristic searches with TBR branch swapping, multrees in effect, and with gaps treated as missing data. Gaps were included in parsimony analyses, but were coded separately using the simple gap coding method of Simmons and Ochoterena (2000). Parsimony heuristic searches included 100 random sequence addition replicates, and clade support was assessed using nonparametric bootstrap analyses with 100 replicates. ML analyses incorporated the best fitting model of sequence evolution as selected by Modeltest v. 3.6 (Posada and Crandall 1998) using the Akaike Information Criterion (Posada and Buckley 2004), with the separate gap characters excluded and 10 random sequence addition replicates.

Bayesian analyses were also conducted on both data sets, using the Metropolis Coupled Markov Chain Monte Carlo simulation program, MrBayes v. 3.0b4 (Huelsenbeck and Ronquist 2001). Modeltest selected GTR + I + Γ (Rodríguez et al. 1990) as the appropriate model for the combined data, and TVM + I + Γ (Posada and Crandall 1998) for the *ndhF* data; because the latter model is not implemented in

MrBayes, we set the model to GTR + I + Γ for the *ndhF* data. Three replicate analyses were run for 3 million generations each to ensure that the runs were converging on the appropriate posterior probability distribution. Analyses were conducted with four chains, with the heating set to 0.15 and proposal parameters adjusted to ensure acceptance rates between 10% and 70%. Trees were sampled every 100 generations, and the point of stationarity was determined by examining plots of the values of the estimated parameters against generation time in Microsoft Excel. All trees prior to reaching stationarity were discarded and the remaining trees were used to compute majority rule consensus trees.

To assess whether *Tiquilia* is of North or South American origin, the current distributions of the various taxa within *Tiquilia* were included as separate characters in the 5-marker combined data set. The distribution of the outgroup *Bourreria succulenta* was defined as the composite distribution of the monophyletic group composed of the sister genera *Bourreria* P. Browne (North, South, and Central America, and the Caribbean) and *Hilsenbergia* Tausch ex Meisn. (Africa; Gottschling and Hilger 2001, Miller 2003). The resulting distribution character was mapped using MacClade v. 4.06 (Maddison and Maddison 2003) onto the tree found from the ML search of the combined data set.

<u>Test of alternate topologies</u>:

Preliminary analyses of each data partition indicated potential incongruence in the branching order of the well-supported major lineages within each subgenus of *Tiquilia*. Even though the 5-marker, 16-taxon reduced data set indicated strong support for all nodes involving the major lineages of *Tiquilia*, a Shimodaira-Hasegawa (SH) test

(Shimodaira and Hasegawa 1999, Goldman et al. 2000) was therefore conducted to determine if any other topologies within *Tiquilia* were reasonable alternatives to the highest likelihood topology. The SH test is an appropriate test of topology if one of the topologies is selected on an *a posteriori* basis (Goldman et al. 2000). In order to ensure that the topology with the highest likelihood is always available when running the SH test, it is desirable to include every possible reasonable topology when running the analysis (Goldman et al. 2000). To this end we exhaustively tested two sets of topologies: in one set of topologies, every possible branching order among the four well-supported major lineages of subg. *Eddya* (*T. canescens* + *T.greggii*, the *T. gossypina* clade, the *T. hispidissima* clade, and *T. purpusii*), while holding the topology of subg. *Tiquilia* constant, was included (15 possible topologies), along with a tree in which all four lineages formed a basal polytomy; while in the other set of topologies, we tested the two alternative branch orders of the three well-supported lineages of subg. *Tiquilia* (the *T. plicata* clade, the *T. palmeri* clade, and *T. nuttallii*) while holding subg. *Eddya* constant, again with a tree in which the three lineages formed a polytomy. This resulted in a total of 19 topologies that were included in the SH test, which employed the same ML settings used for estimating the ML topology of the 5-marker data set, with RELL optimization (1000 replicates; Shimodaira and Hasegawa 1999, Goldman et al. 2000) as implemented in PAUP* 4.0b10.

<u>Molecular-based dating</u>:

The tree topology with the highest likelihood score for the 41-taxon *ndhF* data set was used to infer divergence times. Because a different outgroup topology of

((*Ehretia*,*Bourreria*),*Tiquilia*) was recovered instead of (*Ehretia*,(*Bourreria*,*Tiquilia*)) in some of the individual marker preliminary analyses (see Chapter 4), this alternate topology was also included in the dating analyses. All topologies that could not be rejected by the SH test of the combined data set were also included in the dating analyses, to see what effect these alternative topologies would have on divergence times. Branch lengths for all alternate topologies were estimated using the model selected by Modeltest for the *ndhF* data set.

We utilized the age inferred for the divergence of *Vahlia* and *Borago* (~104 Ma) in the chronogram of asterid ages in Bremer et al. (2004) as a fixed node age in the *ndhF* tree. The ages of Bremer et al. are derived from an analysis based on six well-dated and well-placed fossil-based constraints, and represent the best current effort at dating divergence times in asterid angiosperms. This divergence date was selected because it is the closest node to the Boraginales in the tree, and because the same node is present in our *ndhF* tree. During test runs of our data set in r8s, it was observed that using this age resulted in a divergence date several million years older for the clade uniting *Nicotiana* with *Vahlia* + Boraginales than the 106 Ma date inferred in Bremer et al. for the same node. Thus we also applied this 106 Ma divergence time to this node as an alternate fixed age. One possible alternate calibration point is available for Boraginaceae subfamily Ehretioideae: a well-dated fossil fruit with clear affinities to *Ehretia* P. Browne that has been dated to the early Eocene (~50 Ma; Gottschling et al. 2002). This fossil has been utilized by Gottschling et al. (2004) to date divergence times in the woody Boraginales. However, we could not appropriately apply this date to *Ehretia* in the *ndhF*

tree because the only node in the *ndhF* phylogeny for this genus occurs in a more derived position compared to the node to which the 50 Ma date was applied in Gottschling et al. (2004). Instead, we constrained subfamily Ehretioideae to have a minimum age of 50 Ma, as this is a logically appropriate, although less desirable, placement for such a constraint (Magallón 2004).

Divergence times were estimated using the program r8s v. 1.70 (Sanderson 2003). A likelihood ratio test of molecular clock-constrained and unconstrained ML trees indicated a significant level of rate heterogeneity ($p << 0.001$) in the *ndhF* data set, and for this reason we applied a rate smoothing approach using penalized likelihood (Sanderson 2002) with the TN algorithm, as implemented in r8s v. 1.70, to estimate divergence times. The two furthest outgroups, *Luculia* (Rubiaceae) and *Logania* (Loganiaceae), were pruned from the analysis to eliminate the presence of arbitrary zero-length branches at the root of the tree, and all zero-length internal branches were collapsed. The cross-validation procedure outlined in Sanderson (2002) was performed to determine the appropriate smoothing rate for penalized likelihood, and gradient checks of the correctness of each optimization solution were also performed. A total of five starts per optimization was conducted to check for multiple optima.

Standard errors of divergence dates were estimated using a parametric bootstrapping approach (Huelsenbeck et al. 1996, Swofford et al. 1996, Goldman et al. 2000) similar to that used by Davis et al (2002). Using the *ndhF* ML tree topology, branch lengths, and parameter estimates, 100 data sets were simulated with the program Mesquite v. 1.05 (Maddison and Maddison 2004). Branch lengths were then reestimated

for each simulated data set in PAUP* using the original *ndhF* ML topology and parameter estimates. The resulting trees with branch lengths were then imported into r8s, and standard deviations were then estimated using the 'profile' command using both the 104 Ma and 106 Ma calibration points.

Results

Preliminary phylogenetic analyses confirmed that *Tiquilia* is part of Boraginaceae subfamily Ehretioideae, and is sister to either *Bourreria* alone or to *Bourreria* + *Ehretia* (see Chapters 3 and 4 and Appendix A). These same analyses suggested the monophyly of both *Tiquilia* and its subgenera, and suggested the existence of seven well-supported major lineages. Subgenus *Eddya* contains four of these lineages: (1) *Tiquilia canescens* + *T.greggii*; (2) the *T. gossypina* clade; (3) the *T. hispidissima* clade; and (4) *T. purpusii*. Subgenus *Tiquilia* contains the remaining three major lineages: (5) the *T. plicata* clade; (6) the *T. palmeri* clade; and (7) *T. nuttallii*. This preliminary work was used to select both the 16 taxa to include in the 5-marker combined analyses and the 19 species of *Tiquilia* to include in the *ndhF* dating analyses.

The aligned lengths of the 5-marker (partitioned by marker) and the *ndhF* data sets are indicated in Table 2.3, along with general data set characteristics and the number of gap characters used in parsimony analyses. The 5.8S, *ndhF*, and *matK* genes had the least amount of variation, while the coding regions of *waxy* and the non-coding *trnK* and *rps16* introns had an intermediate amount of variation. The two introns of *waxy* and ITS were highly variable, with regions of ITS unalignable, even in some cases between major lineages of *Tiquilia*.

Waxy sequences of *Tiquilia* and its outgroups showed high similarity to the regions from exon 9 through exon 11 of *waxy* sequences of Solanaceae and Convolvulaceae in GenBank. Exon and intron boundaries were determined by sequence comparison to other GenBank *waxy* sequences and to 5' and 3' consensus sequences for intron splice sites in plants (Csank et al. 1990). *Waxy* sequences of *Tiquilia* and the three outgroups contained portions of exon 9 (ranging from 26-43 bp of the 3' end) and exon 11 (ranging from 126-147 bp of the 5' end), and all of intron 9 (ranging from 91-103 bp), exon 10 (all sequences contained 177 bp), and intron 10 (ranging from 105-120 bp). No indels were detected in any exon sequences.

Parsimony, ML, and Bayesian searches produced identical topologies within each data set and are therefore treated together in each data set. The replicate Bayesian analyses that were performed for both data sets were consistent with each other and indicated that the posterior distribution was properly sampled. Modeltest selected the GTR + I + Γ model as the model that best described the 5-marker combined data, while it selected the TVM (one transition and four transversion types; Posada and Crandall 1998) + I + Γ model for the *ndhF* data. The maximum likelihood tree for the 5-marker data set is depicted in Fig. 2.1, with Bayesian posterior probabilities indicated (bootstrap values are omitted from the tree because they were identical to Bayesian posterior probabilities). The parsimony analysis of the 5-marker data set returned 10 most parsimonious trees which differed only in the branching order among the taxa of the *T. gossypina* clade. Biogeographic analysis supports a North American origin for *Tiquilia*, including both subgenera and all seven major lineages, with dispersals to South America and the

Galápagos Islands (see Chapter 3). The biogeographic reconstruction is shown on the phylogeny in Fig. 2.1.

The maximum likelihood tree for the *ndhF* data set is depicted in Fig. 2.2, with parsimony bootstrap percentages and Bayesian posterior probabilities also indicated. The parsimony analysis of the *ndhF* data set returned a single most parsimonious tree.

Seven of the 19 alternate combined data set topologies could not be rejected at the $p = 0.05$ level by the SH test. These seven topologies are depicted in a simplified form in Fig. 2.3.

Maximum likelihood analysis of the *ndhF* data set returned a tree with an unconstrained likelihood score of -11529.351 and a clock-constrained likelihood score of -11682.074. A χ^2 test of twice the difference in these likelihoods strongly rejected clock-like evolution in this data set (2 × difference = 305.446; df = 39; $p << 0.001$).

The node ages derived from the r8s analysis of the *ndhF* data set using the 104 Ma calibration point are depicted in the chronogram in Fig. 2.4. These values are also given in Table 2.4, along with the node ages derived from the 106 Ma calibration point. The mean ages and standard deviations obtained for each node in the parametric bootstrap analysis are also indicated in Table 2.4. A smoothing rate of 100 was suggested as the most appropriate by cross-validation analyses, and gradient checks of optimization solutions passed in all cases. Dates obtained from the 104 and 106 Ma calibration points did not conflict with the 50 Ma minimum age constraint for Boraginaceae subfam. Ehretioideae. The analysis of the alternate topology of ((*Bourreria*,*Ehretia*),*Tiquilia*) produced divergence times that varied only slightly from

the ML toplogy, and hence these times are not listed. Node ages derived from the analyses of the SH trees were essentially identical in most cases, with significant differences restricted to nodes that were not present in the original ML topology. These different node ages are depicted on the SH trees themselves in Fig. 2.3; other dates are omitted.

Discussion

Phylogenetic relationships of *Tiquilia*:

Tiquilia, its two subgenera, and all seven of the major lineages suggested by the preliminary data are strongly supported as monophyletic in both the *ndhF* and the 5-marker combined analyses (these seven lineages are indicated in the combined data phylogeny in Fig. 2.1). The relationships within some of these seven lineages are sometimes difficult to recover using a larger sample of taxa, particularly within the *T. gossypina*, *T.hispidissima*, and *T. plicata* clades. Incongruence among one or more data partitions exists in all three of these lineages (see Chapters 3 and 4). Such incongruence is frequently encountered when reconstructing relationships among closely related species (Wendel and Doyle 1998, McKinnon 2005), and since it is of no consequence to the deeper-level relationships in the genus, it is not discussed further here.

Reduction of the combined data matrix to include 14 exemplar taxa within *Tiquilia* results in a non-significant ILD test result (p = 0.15) among all five data partitions, suggesting congruence among all markers in the relationships among the seven major monophyletic lineages. This congruence is corroborated by the strong bootstrap and Bayesian support for the branching order in the combined data phylogeny (see Fig.

2.1). However, the SH test of the branching order among these seven lineages fails to reject seven of 19 alternate topologies (see Fig. 2.3), including all three alternate topologies in subg. *Tiquilia*. The remaining four alternate topologies are restricted to subg. *Eddya* and do not appear to follow an easibly discernible pattern. Therefore the SH test would seem to indicate a lack of confidence in the deeper-level branch order within *Tiquilia*. However, the topology in the combined data phylogeny makes sense from a morphological standpoint (e.g., the *T. gossypina*, *T. hispidissima*, and *T. purpusii* clades share a similar fruit morphology, and it would make sense for them to form a monophyletic group; the same could be said about the *T. palmeri* clade and *T. nuttallii* based on foliar morphology), and it is possible, if not likely, that with even more sequence data the SH test would converge on the combined data topology. Because the combined data and *ndhF* topologies are identical, with very similar proportional branch lengths, and because a much larger sample of outgroups is available for *ndhF*, it seems reasonable to utilize the *ndhF* phylogeny for analyses of divergence times.

The *ndhF* data allow for a reasonably clear picture of the position of *Tiquilia* within the Boraginales (see Fig. 2.2). The *ndhF* phylogeny is largely congruent with the ITS phylogeny of Boraginales in Gottschling et al. (2001) with respect to the major lineages of Boraginales, and supports the position of *Tiquilia* within Ehretioideae, with *Tiquilia* sister to the American/African tropical sister genera *Bourreria* and *Hilsenbergia* (represented by *Bourreria* only in the *ndhF* phylogeny; Gottschling and Hilger 2001, Miller 2003). However, there is only moderate support for this relationship (72% bootstrap proportion; Bayesian posterior probability = 0.72). In some of the preliminary

individual marker analyses that include both *Ehretia* and *Bourreria*, the two instead form a sister group to *Tiquilia*. The *ndhF* phylogeny also strongly supports the monophyly of Ehretioideae (excluding *Coldenia*, which in our analysis is sister to *Cordia*) and of Boraginales, and also supports the sister relationship of *Vahlia* and Boraginales suggested by Bremer et al (2002). The other subfamilial taxa of Boraginaceae (including subfamilies Boraginoideae, Cordioideae, and Heliotropioideae) and Hydrophyllaceae (including clades I and II of Ferguson 1999) are recovered as monophyletic lineages, with reasonably strong support. The branching order among these lineages is also well-supported by *ndhF*, although the branch lengths are relatively short.

Molecular dating in the Boraginales:

The divergence dates obtained using the 104 Ma calibration point (*Vahlia* + Boraginales) are approximately 5-6% older than those obtained from the 106 Ma calibration point (*Nicotiana* + *Vahlia*/Boraginales) across the *ndhF* phylogeny (Table 2.4). This difference results in only a minor discrepancy in node ages for most of the Boraginales, particularly in the younger nodes (Table 2.4); henceforth they are discussed together. In neither case do the calibration points conflict with the 50 Ma minimum age constraint imposed on the Ehretioideae (node ages of 61.0 Ma and 57.5 Ma, respectively).

The penalized likelihood analysis applied to the *ndhF* phylogeny suggests that extant Boraginales began diversifying in the late Cretaceous (79.7-75.6 Ma; see Fig. 2.4 and Table 2.4). This was followed by a relatively rapid diversification of the major lineages of Boraginales, including *Codon*, the Boraginoideae, Hydrophyllaceae clades I

and II (*sensu* Ferguson 1999), and the Heliotropioideae, Cordioideae (including *Coldenia*), and Ehretioideae, from the late Cretaceous to the early Tertiary. Within the Ehretioideae, an initial rapid radiation into the stem lineages of *Ehretia*, *Bourreria*, and *Tiquilia* is implied in the Paleocene (61.0-55.6 Ma). After a relatively long isolation, the *ndhF* analysis implies that the split between the two subfamilies of *Tiquilia* dates to the early Oligocene (33.1-31.4 Ma). This was followed by diversification of the major lineages of the genus during the early to mid-Miocene (23.3-13.6 Ma). The *ndhF* analysis suggests that the within-lineage diversification of *Tiquilia* did not begin until the late Miocene (~7 Ma), and likely continued at least through the Pliocene (after 5.3 Ma).

It should be noted that our dates for the initial diversification and radiation of the major clades of Boraginales are considerably younger than the dates implied in Gottschling et al. (2004), who based their analysis largely on the 50 Ma *Ehretia* fossil fruit described above in the Materials and Methods. They interpret their analysis as suggesting a mid-Cretaceous origin for the primarily woody Boraginales (a group including Boraginaceae subfamilies Cordioideae, Heliotropioideae, and Ehretioideae), whereas the same node in our analysis suggests an origin for these subfamilies near the Cretaceous/Tertiary boundary (67.1-63.3 Ma; Table 2.4). Constraining this node to have a mid-Cretaceous date (90 Ma) in the *ndhF* tree results in a highly improbable mid-Jurassic age for the divergence of Vahliaceae and Boraginales. While we admit that our analysis does not rely on ingroup fossil data (and likely never will), we feel that the use of inferred divergence dates from the work of Bremer et al. (2004), which utilizes

multiple well-placed and well-dated fossils from throughout the asterids, represents a more conservative approach.

The role of aridity in the evolutionary history of *Tiquilia*:

We can now examine whether *Tiquilia* conforms to the general outline of desert plant evolution implied by Axelrod (1950, 1979a,b) by revisiting the three questions asked in the introduction:

(1) *Is* Tiquilia *a relatively old arid-adapted lineage that evolved in North America?*

Even in the absence of molecular data, there are a number of reasons to suspect that *Tiquilia* is a relatively old North American group. Most of the extant diversity in the genus is found in North America (in spite of the great diversity of coastal South American *Tiquilia*; Richardson 1977; see Chapter 3), including all of subg. *Eddya*, which is suggestive of an origin on that continent. That this origin may be relatively ancient is suggested by the morphological divergence of *Tiquilia* from its relatives, and by the sheer diversity of extant *Tiquilia*. *Tiquilia* possesses a number of unique traits compared to the rest of Ehretioideae, including a low habit, distinctive leaf venation, nutlets instead of drupes, dichotomous branching, and a number of other characters indicative of adaptation to aridity (including a number of leaf characters; Richardson 1977). In fact, *Tiquilia* is the only genus in Ehretioideae that is restricted to arid and semi-arid habitats (Miller 2003). A potentially long history for the genus is also suggested by its great diversity, which is best illustrated by contrasting the two subgenera, *Tiquilia* and *Eddya*. Morphologically, these two monophyletic groups differ substantially in their leaf

venation, branching pattern, and fruit morphology (Richardson 1977). They also differ in base chromosome number (subg. *Eddya*, $x = 9$; subg. *Tiquilia*, $x = 8$) and habitat preferences (subg. *Eddya*, calcareous substrates; subg. *Tiquilia*, loose sand; Richardson 1977). The major lineages within each subgenus are themselves quite distinct from each other morphologically, and frequently occur sympatrically (in the case of subg. *Eddya*, with up to five species growing together) with no signs of hybridization. This tremendous diversity within *Tiquilia* is therefore likely to be the result of relatively ancient speciation events, and not the product of a recent, rapid radiation.

The *ndhF* and combined analyses strongly suggest that *Tiquilia* is in fact an old North American lineage. Biogeographic reconstruction confirms a North American origin for *Tiquilia* and all of its major lineages (Fig. 2.1), and molecular-based dating of the *ndhF* phylogeny indicates that the stem lineage of *Tiquilia* may have diverged from its nearest extant relatives as early as the Paleocene (Fig. 2.4). Furthermore, the long branch separating *Tiquilia* from its nearest relatives suggests a potentially long period of isolation (on the order of 20-25 million years) for the stem lineage of *Tiquilia*. Such isolation would help to explain the great morphological divergence between *Tiquilia* and the rest of the Ehretioideae. This long branch is probably a real phenomenon, and not an artifact of undersampling of potential sister groups in the Ehretioideae. Various molecular-based phylogenies of the taxa within Ehretioideae (Gottschling et al. 2001, Gottschling 2003) suggest that no other genera within the subfamily are closer to *Tiquilia* than the sister genera *Bourreria* and *Hilsenbergia*. According to the penalized likelihood analysis, the divergence of the two subgenera occurred in the early Oligocene, with all of

the major lineages diverging by the mid-Miocene, roughly 14 Ma (Table 2.4, Fig. 2.4). Such relatively ancient divergence times would account for the considerable morphological diversity of *Tiquilia* and for the reproductive isolation suggested by the morphological integrity of the sympatric species within each subgenus.

The early divergence dates suggested by the *ndhF* analyses for the stem and crown lineages of *Tiquilia* have important implications for the origin of adaptation to aridity in the genus. Because all extant species of *Tiquilia* occupy arid habitats, it is very likely that the most recent common ancestor of *Tiquilia* was also arid-adapted. The *ndhF* r8s analysis indicates that this ancestor lived no later than the late Eocene/early Oligocene, judging from the age of the divergence implied for the two subgenera (Fig. 2.4). However, *Tiquilia* in the morphological sense (that is, an arid-adapted subshrub) may have existed many millions of years prior to this in the middle Eocene, in local xeric sites among the subtropical dry forest and savanna that existed in western North America at the time (Axelrod 1950, 1979a; Graham 1999). This drier vegetation type probably first spread significantly in the middle to late Eocene as a response to the gradual cooling and drying that had been occurring since the Eocene peak warmth of around 55-50 Ma (Graham 1999). Based on the current knowledge of the phylogeny, distribution, and ecology of the taxa of Ehretioideae, it is possible that the earliest Ehretioids were members of a seasonally dry tropical forest or scrub community (many extant species of Ehretioideae grow in such environments at present; Standley 1924, Leon and Alain 1957, Miller 1989, Miller 2002a), and it is likely that at least part of this ancient range included North America. These putative North American ancestral Ehretioids could have

inhabited the middle Eocene subtropical/tropical dry scrub or sclerophyll woodland communities of western North America, eventually giving rise to *Tiquilia* in areas of local aridity or semi-aridity in the middle to late Eocene. Such a scenario, whereby a tropical plant group becomes adapted to aridity during the Eocene in time to benefit from the increasingly dry climates that followed, is exactly in line with the hypotheses of Axelrod (1950, 1979a) concerning the evolution of the current American desert flora. However, it should be emphasized that the lack of fossil information for early *Tiquilia* and Ehretioids in general precludes the confirmation of such a hypothesis, in spite of the suggestiveness of the *ndhF* phylogeny and r8s analyses.

Tiquilia may also conform to two of Axelrod's corollary arguments concerning the origin of arid-adapted plant lineages. First, Axelrod argued that local islands of edaphic aridity would promote the evolution of an autochthonous, arid-adapted flora (Axelrod 1967, 1972, 1979a). Both subgenera of *Tiquilia* prefer their own, often arid, substrates: the species of subg. *Tiquilia* are restricted to loose sand, while the species of subg. *Eddya* are restricted to calcareous substrates (Richardson 1977). Perhaps more importantly, however, species in both subgenera nearly always establish in substrate lacking plant cover, and consequently are frequently encountered in recently disturbed substrate. Such bare sites are characterized by high insolation of the soil, resulting in relatively high soil temperatures and low soil moisture near the surface, particularly in arid or semi-arid regions (MacMahon 1999, Brady and Weil 2002). It is likely that ancestral *Tiquilia* evolved as a specialist of these dry, open habitats, which were probably more restricted in extent prior to the advent of regional deserts. Evolution of new species

or ecotypes can proceed rapidly when populations occupy extreme environments such as these [the catastrophic selection of Lewis (1962); see also Stebbins 1952, Axelrod 1967, Levin 2005], and it is possible that such was the case for ancestral *Tiquilia*.

Second, Axelrod (1979a) pointed out that essentially all of the modern plants thought of as desert-adapted either occur outside of the deserts or have close relatives that do, and thus the ancestors of these plants could have evolved and/or survived in local patches of semi-arid woodland or tropic scrub, even in the complete absence of truly arid sites. *Tiquilia* follows just such a pattern, particularly with respect to the members of subg. *Eddya*. Although they are more common in arid sites, all of the species of subg. *Eddya* also occur outside of the deserts in semi-arid regions of desert grassland or oak-juniper-piñon savanna. In these semi-arid environments individuals grow either in recently disturbed substrate or on arid, calcareous slopes with low plant cover. Although we cannot determine the exact place and substrate of origin for ancestral *Tiquilia*, such dry microhabitats would have existed in the Eocene prior to the advent of regional deserts, just as they exist today. It is certainly possible that ancestral *Tiquilia* evolved in such arid microhabitats during the Eocene, and that the genus was able to survive in these habitats, albeit with a restricted range, throughout much of its evolutionary history.

(2) *When did the majority of diversification within the genus occur, and is any of this diversification correlated with episodes of aridification during the Cenozoic?*

There appear to have been two episodes of deeper-level diversification within *Tiquilia*. The divergence of the modern subgenera of *Tiquilia*, which the molecular dating analyses imply occurred soon after the end of the Eocene ~33.5 Ma (Fig. 2.4),

constitutes the first episode. If properly dated, this divergence would coincide with the end of one of the greatest episodes of Cenozoic aridification. This significant cooling of the Earth began near the end of the Eocene, and resulted in a terrestrial average temperature drop of 6-8°C in the middle latitudes of North America (Wolfe 1992, 1997). The mean annual range of temperature also increased signficantly near the Eocene/Oligocene boundary, perhaps as much as 8-10°C (Wolfe 1992). These temperature shifts were accompanied by decreases in rainfall over the interior of North America (Graham 1999), causing the tropical forests that dominated large swaths of southern North America during the middle Eocene to retreat southward (Wolfe 1992, 1997). Also by the early Oligocene, oak-pine savanna occupied at least portions of the western United States (Graham 1999), and drier chaparral vegetation may have also spread over portions of this same area (Graham 1999). The drier, cooler climates of the Oligocene also favored the beginning of the rise of mostly herbaceous angiosperm plant families (such as Asteraceae; Wolfe 1997). The sudden increase in aridity at the Eocene/Oligocen boundary may have provided new opportunities for the regional and ecological expansion of *Tiquilia*, which hitherto may have been restricted to very local dry areas.

The second episode of diversification seems to have been confined to the early to mid-Miocene, and involved the radiation of the subgenera into all of the major lineages of extant *Tiquilia* (Fig. 2.4). It is more difficult to correlate any of these earlier divergence events, which the r8s analysis suggests occurred over ~9 million years, with a particular episode of aridification. The Oligocene and early Miocene were climatically

similar times, but were followed by a mid-Miocene warming trend that ended around 15 Ma (Wolfe 1997, Graham 1999). After 15 Ma, cooling and drying resumed, resulting in the first appearance of true grasslands, the expansion of piñon-juniper woodland, and the possible appearance of some semi-desert vegetation in the modern Sonoran Desert region (Axelrod 1979a, Graham 1999). According to the molecular dating analyses, the divergence of *Tiquilia canescens* and *T. greggii* dates to this time (~14 Ma; Fig. 2.4), as does the divergence of *T. purpusii* and *T. hispidissima*/*T.latior*, and it is possible that these diversifications are somehow linked to the aridifying conditions of the time. Other drought-adapted plant lineages are also thought to have diversified after the mid-Miocene cooling event, including C_4 grasses (Sage 2004) and North American tarweeds (Asteraceae; Baldwin and Sanderson 1998). Finally, it should be noted that alternative reconstructions of the branching order within the major lineages of *Tiquilia*, as suggested by the SH test, result in age differences for several nodes (see Fig. 2.3). However, in all cases these differing ages are confined to the early to mid-Miocene, further corroborating the importance of this time to the evolution of *Tiquilia*.

Although widespread desert habitat was probably nonexistent prior to the late Miocene (Wolfe 1985, Graham 1999), the fact that the modern subgenera of *Tiquilia* likely began to diversify ~20 Ma indicates that suitable habitat for *Tiquilia* must have existed. Axelrod (1979a) predicts that local semi-arid or arid sites may have been present at this time across southwestern North America, and it is probable that *Tiquilia* was restricted to these limited areas. Such a fragmented distribution may have promoted allopatric speciation in *Tiquilia*. If isolated long enough, localized species such as these

would evolve postzygotic reproductive barriers, thereby prohibiting subsequent hybridization (as opposed to closely related species, which are often isolated by geography or by prezygotic barriers; Levin 2004). This type of scenario would account for the current sympatric ranges of many of the major lineages within each subgenus, which is otherwise difficult to explain due to the seemingly similar habitat requirements and pollination syndromes of the lineages involved.

(3) *Is there any evidence of increased species- or population-level diversity in* Tiquilia *with the onset of semi-arid and arid conditions since the late Miocene?*

The *ndhF* analyses suggest that diversification within each of the major lineages of *Tiquilia* dates to the late Miocene (~7 Ma) or later (see Table 2.4 and Fig. 2.4), and is likely correlated with the expansion of semi-arid and later arid habitats during the Pliocene. From the late Miocene, mountain ranges and plateaus in Asia and western North American were uplifted (Graham 1999). The changes in global air circulation that followed increased the extent of polar ice and eventually initiated the glacial/interglacial cycles of late Pliocene and Pleistocene times. Until the late Pliocene, it is likely that deserts were at best a local phenomenon (Axelrod 1979a); nevertheless, semi-arid habitats expanded, allowing for the continued spread of arid-adapted plants such as *Tiquilia* (Graham 1999). The *ndhF* analysis indicates that *Tiquilia* underwent a new round of diversification within several of the major lineages during the early Pliocene (e.g., in the *T. gossypina* and the *T. hispidissima* clades), resulting in the origin of a number of new species. However, even though *Tiquilia* likely spread significantly during the late Miocene and early Pliocene, it probably still had a patchier distribution than

today. It was only during the late Pliocene that climate conditions aridified enough to favor the spread of regional deserts. Because *Tiquilia* was preadapted to the newly arid conditions, it would have expanded significantly and potentially rapidly into its new desert surroundings. This expansion may have brought into contact a number of hitherto isolated species or populations of *Tiquilia*, and it is possible that the widespread sympatry of the major lineages of the genus may initially date from this time. The glacial/interglacial cycles of the Pleistocene created a concomitant cycle of contraction (during glacial periods) and expansion (during interglacial periods) of the desert biome in southwestern North America (Axelrod 1979a, Graham 1999, Lowenstein et al. 1999), which could be expected to cause simultaneous range contractions and expansions in populations of *Tiquilia*. The *T. gossypina* clade exhibits a somewhat confusing pattern of molecular variation across its range that may relate to the climate upheavals of the Pleistocene. The closely related species of this lineage show evidence of probable hybridization among formerly isolated populations (data are presented in Chapter 4) that could have resulted from repeated isolation and contact among these populations during the Pleistocene. Finally, the spread of *Tiquilia* in North America also increased the likelihood of amphitropical dispersal to the newly arid regions of the Pacific coast of South America and the Galápagos Islands (see Chapter 3). The *ndhF* phylogeny suggests that the *T. palmeri* and *T. plicata* clades did not disperse to these regions prior to the late Miocene approximately 6-7 Ma (Table 2.4 and Fig. 2.4), which correlates well with the initial advent of aridity in these regions at about the same time (Hartley and Chong 2002).

Conclusions

All of the evidence derived from the molecular dating and phylogenetic analyses of *Tiquilia* supports Axelrod's hypotheses of the evolution of the modern North American desert flora. Although its exact place of origin within North America cannot be determined, it is possible that the Ehretioid ancestors of *Tiquilia* inhabited dry tropic scrub in southwestern North America during the Paleocene. *Tiquilia* likely first appeared in the middle to late Eocene as a specialist of open, edaphically dry sites, and probably remained restricted to locally semi-arid sites through much of its early history. The divergence of the two modern subgenera of *Tiquilia* probably occurred shortly after the great aridification event of the late Eocene/early Oligocene, ~33 Ma. The current major lineages of *Tiquilia* all likely arose in the early to mid-Miocene, perhaps evolving allopatrically in locally arid or semi-arid pockets. The drying and cooling of the late Miocene and early Pliocene prompted further expansion and diversification within the major lineages of *Tiquilia* (from ~7-4 Ma), followed by a potentially rapid range expansion with the onset of widespread aridity over southwestern North America in the late Pliocene and Pleistocene interglacial periods.

Although this sequence of events corresponds well with Axelrod's ideas of North American desert flora assembly, we must emphasize that this evolutionary scenario is based on a molecular dating analysis that relies on a single calibration point, far removed from the ingroup. The ages inferred from this type of analysis should therefore be treated cautiously. Nevertheless, there are a number of independent reasons to accept the ancient divergence dates implied within *Tiquilia* by the molecular analyses: the calibration point

used in this study is based on a thorough and well-calibrated analysis; the phylogeny of *Tiquilia* and its outgroups is well-supported, allowing a good degree of confidence in our interpretation of the sequence of evolutionary events; and the morphological and ecological divergence of *Tiquilia* from its nearest relatives, combined with the great morphological diversity within *Tiquilia*, independently suggests an ancient age for the genus and its main lineages. In arid-adapted plant groups such as *Tiquilia* that lack a fossil record, there is no choice but to seek alternative methods for calibrating divergence times. This study demonstrates that the utilization of a reasonably dated calibration point located far from the ingroup, while less desirable, can still yield useful information about evolutionary diversification in fossil-poor groups.

Table 2.1. List of accessions included in *ndhF* and 5-marker combined phylogenetic analyses. All vouchers are deposited at TEX, unless otherwise noted (standard herbarium acronyms are used).

Species	GenBank/Collection Info.
Boraginales	
Boraginaceae subfam. Boraginoideae	
Borago officinalis L.	L36393 (Olmstead and Reeves 1995)
Cryptantha flavoculata Payson	AF047803 (Ferguson 1999)
Boraginaceae subfam. Cordioideae	
Cordia nodosa Lam.	AF047808 (Ferguson 1999)
Boraginaceae subfam. Ehretioideae	
Bourreria costaricensis (Standl.) A.H. Gentry	AF047797 (Ferguson 1999)
Bourreria succulenta Jacq.	Cuba: Pinar del Rio, *R. G. Olmstead 96-114* (WTU)
Coldenia procumbens L.	Ghana: Bolgatanga, *Jongkind & Nieuwenhuis 1973* (MO)(cloned for *waxy*)
Ehretia acuminata R.Br. (in GenBank as *E. ovalifolia* Hassk.)	AF047800 (Ferguson 1999)
Ehretia anacua I.M. Johnst.	USA, Texas: Travis County, *M. J. Moore s.n.*
Tiquilia canescens (DC.) A.T. Richardson	USA, Nevada: Clark County, *M. J. Moore 239*
Tiquilia conspicua (I.M. Johnst.) A.T. Richardson	Perú: Dpto. Arequipa, *M. J. Moore 294* (cloned for *waxy*)
Tiquilia cuspidata (I.M. Johnst.) A.T. Richardson	México: Baja California Sur, *M. J. Moore 223*
Tiquilia darwinii (Hook.f.) A.T. Richardson	Galápagos Islands: Isla Santiago, *A. Tye 573*
Tiquilia “Durango” (undescribed species)	México: Durango, *M. J. Moore 260*
Tiquilia gossypina (Wooton and Standl.) A.T.Richardson	Accession 134—USA, Texas: Brewster County, *M. J. Moore 134*; Accession 263—México: Coahuila, *M. J. Moore 263*
Tiquilia greggii (Torr. & A. Gray) A.T. Richardson	USA, Texas: Brewster County, *M. J. Moore 133*

Tiquilia hispidissima (Torr. & A. Gray) A.T. Richardson	Accession 131—USA, Texas: Brewster County, *M. J. Moore 131*; Accession 154—USA, Texas: Culberson County, *M. J. Moore 154*
Tiquilia latior (I.M. Johnst.) A.T. Richardson	Accession 211—USA, Arizona: Navajo County, *M. J. Moore 211*; Accession 216—USA, Utah: Wayne County, *M. J. Moore 216*
Tiquilia mexicana (S. Watson) A.T. Richardson	USA, Texas: Terrell County, *M. J. Moore 117*
Tiquilia nuttallii (Benth.) A.T. Richardson	USA, Washington: Grant County, *M. J. Moore 218*
Tiquilia palmeri (A. Gray) A.T. Richardson	Accession 197—USA, California: Riverside County, *M. J. Moore 197*; Accession 202—USA, Arizona: Yuma County, *M. J. Moore 202*
Tiquilia paronychioides (Phil.) A.T. Richardson	Perú: Dpto. Arequipa, *M. J. Moore 300* (cloned for ITS, *waxy*)
Tiquilia plicata (Torr.) A.T. Richardson	USA, California: Riverside County, *M. J. Moore 196*
Tiquilia purpusii (Brandegee) A.T. Richardson	México: San Luís Potosí, *M. J. Moore 109*
Tiquilia turneri A.T. Richardson	México: Coahuila, *M. J. Moore 89*
Boraginaceae subfam. Heliotropioideae	
Tournefortia acutiflora M. Martens & Galeotti	AF047813 (Ferguson 1999)
Heliotropium arborescens L.	AF014000 (Ferguson 1999)
Hydrophyllaceae	
Codon schenckii Schinz	AF047776 (Ferguson 1999)
Eriodictyon californicum Greene	AF047820 (Ferguson 1999)
Hydrophyllum virginianum L.	AF019646 (Ferguson 1999)
Nama sericeum Willd.	AF047798 (Ferguson 1999)
Phacelia congesta Hook.	AF047780 (Ferguson 1999)
Romanzoffia californica Greene	AF047804 (Ferguson 1999)
Tricardia watsonii Torr. ex S. Watson	AF047775 (Ferguson 1999)
Wigandia urens Urb.	AF047763 (Ferguson 1999)

Loganiaceae	
Logania vaginalis (Labill.) F. Muell.	AJ235837 (Backlund et al. 2000)
Rubiaceae	
Luculia gratissima Sweet	AJ011987 (Oxelman et al. 1999)
Solanaceae	
Nicotiana tabacum L.	NC_001879 (Shinozaki et al. 1986)
Vahliaceae	
Vahlia capensis Thunb.	AJ429112 (Bremer et al. 2002)

Table 2.2. List of primers utilized in this study. Primers denoted with an asterisk (*) were used only for sequencing. Numbers in primer names for *ndhF* and *matK* primers refer to approximate base pair position downstream of the transcription start position. F = forward primer; R = reverse primer.

Marker/Primer Name	Primer Sequence (5´→3´)	Reference
ndhF		
ndhF 1F	ATG GAA CAK ACA TAT SAA TAT GC	Olmstead and Sweere 1994
ndhF 536F*	TTG TAA CTA ATC GTG TAG GGG A	Olmstead and Sweere 1994
ndhF 972F	GTC TCA ATT GGG TTA TAT GAT G	Olmstead and Sweere 1994
ndhF T972F	GTC TCA GTT RGG TTA TAT GAT G	designed for this study
ndhF 1318F*	GGA TTA ACY GCA TTT TAT ATG TTT CG	Olmstead and Sweere 1994
ndhF T1318F*	GGA TTA ACT GCA TTT TAT ATG TTT CG	designed for this study
ndhF T730R*	CAT ACA TGA AGT GGA AAT TGT GCA	designed for this study
ndhF 972R*	CAT CAT ATA ACC CAA TTG AGA C	Olmstead and Sweere 1994
ndhF T972R*	CAT CAT ATA ACC YAA CTG AGA C	designed for this study
ndhF 1318R	CGA AAC ATA TAA AAT GCR GTT AAT CC	Olmstead and Sweere 1994
ndhF T1318R	CGA AAC ATA TAA AAT GCA GTT AAT CC	designed for this study
ndhF 1603R*	GCA TAG TAT TGT CCG ATT CAT RAG G	Olmstead and Sweere 1994
ndhF T1603R*	ACA TAG TAT TAT CCG ATT CC	designed for this study
ndhF 2110R	CCC CCT AYA TAT TGA TAC CTT CTC C	Olmstead and Sweere 1994
matK		
matK 1F	ACT GTA TCG CAC TAT GTA TCA	Sang et al. 1997
matK 230F*	GTT CAC TAA TTG TGA AAC GT	Sang et al. 1997 (= *matK* 2F)
matK T230F*	CAG TTT ACT AAT TGT GAA ACG T	designed for this study
matK 590F	AAG ATG CCT CTT CTT TGC AT	Sang et al. 1997 (= *matK* 3F)
matK T590F	AAG ACC CCT CTT CTT TGC AT	designed for this study
matK 1320F*	TCT CAT TAT CAC AGC GGA TC	Sang et al. 1997
matK T1320F*	TCT CAT TAT TAT AGC GGA TC	designed for this study
matK 1320R	GAT CCG CTG TGA TAA TGA GA	Sang et al. 1997 (= *matK* 3R)
matK T1320R	GAT CCG CTA TAA TAA TGA GA	designed for this study
matK 1580R*	TTC ATG ATT GGC CAG ATC A	Sang et al. 1997 (= *matK* 2R)

matK T1580R*	TTG ATG ATT GGC CAG ATC A	designed for this study
matK 1820R	GAA CTA GTC GGA TGG AGT AG	Sang et al. 1997 (= *matK* 1R)
***rps16* intron**		
rps16 F	GTG GTA GAA AGC AAC GTG CGA CTT	Oxelman et al. 1997
rps16 R2	TCG GGA TCG AAC ATC AAT TGC AAC	Oxelman et al. 1997
ITS		
ITS 1A (modified ITS 5)	GGA AGG AGA AGT CGT AAC AAG G	Downie and Katz-Downie 1996
ITS 4	TCC TCC GCT TAT TGA TAT GC	White et al. 1990
waxy		
waxy 9F	GAT ACC CAA GAG TGG AAC CC	Miller et al. 1999
waxy T9F-alt	GCA ACT GAT AAA TAC ATT GAT GTT C	designed for this study
waxy 11R	GTT CCA TAC GCA TAG CAT G	Miller et al. 1999
waxy T11R-alt	CAA TTG AAT GAG ACC ACA AGG CTC	designed for this study

Table 2.3. Summary of data set characteristics. MPTs = most parsimonious trees; RCI = rescaled consistency index; RI = retention index; I = proportion of invariant sites; Γ = rate heterogeneity. References: GTR (Rodríguez et al. 1990), TVM (Posada and Crandall 1998).

Data set	Aligned length (bp)	G + C content	# pars. inform. characters	# gap characters (# inform.)	# MPTs	RCI	RI	Model selected
5-marker combined (all markers)	6040	37.5%	644	153 (59)	1	0.613	0.767	GTR + I + Γ
5-marker combined (*ndhF*)	2062	33.0%	152	2 (1)	n/a	n/a	n/a	n/a
5-marker combined (*matK*)	1775	33.3%	137	13 (6)	n/a	n/a	n/a	n/a
5-marker combined (*rps16*)	900	33.0%	69	41 (15)	n/a	n/a	n/a	n/a
5-marker combined (ITS)	677	67.6%	143	60 (27)	n/a	n/a	n/a	n/a
5-marker combined (*waxy*)	626	39.5%	143	37 (10)	n/a	n/a	n/a	n/a
ndhF	2087	32.8%	418	n/a (gaps not included)	10	0.481	0.754	TVM + I + Γ

Table 2.4. Crown ages of nodes derived from the r8s molecular dating analyses, along with standard deviations derived from parametric bootstrap analyses. Clade names refer only to the taxa in Figs. 2.2 and 2.4. The calibration point ages are depicted in bold.

Clade name	Crown age of clade—104 Ma cal. point (Ma)	Crown age of clade—106 Ma cal. point (Ma)	Mean crown age ± S.D., from parametric bootstrap (Ma)	Mean crown age ± S.D., from parametric bootstrap (Ma)
root	115.8	**106.0**	116.5 ± 5.6	**106.0**
Vahlia + Boraginales	**104.0**	99.1	**104.0**	99.1 ± 3.0
Boraginales	79.7	75.6	80.4 ± 4.0	76.3 ± 4.8
Codon + Boraginoideae	69.1	65.6	70.0 ± 4.4	66.5 ± 4.7
Boraginoideae	31.9	30.4	32.6 ± 3.1	31.0 ± 2.9
Hydr. Clade II + Clade I/woody Borag.	73.0	69.0	74.2 ± 5.2	70.2 ± 5.9
Hydrophyllace Clade II	53.6	50.6	54.9 ± 5.0	51.8 ± 5.2
Tricardia + *Romanzoffia/Phacelia*	52.6	49.7	53.5 ± 4.8	50.6 ± 4.8
Romanzoffia + *Phacelia*	46.1	43.6	47.3 ± 5.2	44.7 ± 5.2
Hydroph. Clade I + woody Boraginales	70.3	66.2	71.1 ± 5.7	67.0 ± 6.2
Hydrophyllaceae Clade I	59.5	55.7	61.6 ± 7.9	57.7 ± 8.4
Wigandia + *Eriodictyon*	34.8	31.6	38.5 ± 10.5	35.1 ± 10.2
woody Boraginales	67.1	63.3	68.0 ± 5.7	64.1 ± 6.2
Heliotropioideae	23.4	21.9	25.1 ± 4.8	23.4 ± 4.6
Coldenia + *Cordia*	60.4	57.0	61.2 ± 5.8	57.8 ± 6.1
Ehretioideae	61.1	57.5	62.3 ± 6.2	58.8 ± 6.6
Ehretia	19.8	17.9	22.4 ± 9.0	20.1 ± 8.2
Bourreria + *Tiquilia*	59.0	55.6	59.6 ± 5.9	56.0 ± 6.2
Bourreria	7.3	6.9	7.4 ± 3.1	6.9 ± 3.0
Tiquilia	33.1	31.4	34.4 ± 3.7	32.7 ± 3.9
T. subg. *Tiquilia*	19.8	18.8	20.3 ± 2.4	19.3 ± 2.4

T. nuttallii + *T. palmeri* clade	18.3	17.4	18.2 ± 2.3	17.3 ± 2.3
T. palmeri clade	6.8	6.5	7.1 ± 2.0	6.8 ± 1.9
T. plicata clade	17.4	16.6	17.8 ± 2.4	16.9 ± 2.5
T. darwinii + *T. paron./T. cusp.*	6.0	5.8	6.5 ± 1.5	6.2 ± 1.5
T. paronychioides + *T. cuspidata*	4.7	4.5	4.4 ± 1.3	4.2 ± 1.2
T. subg. *Eddya*	23.3	22.2	23.5 ± 2.9	22.3 ± 2.9
T. goss. clade + *T. hisp.* clade/*T. purp.*	20.5	19.5	20.6 ± 2.5	19.5 ± 2.5
T. hispidissima clade + *T. purpusii*	14.3	13.6	14.4 ± 2.6	13.7 ± 2.5
T. hispidissima clade	4.9	4.6	5.2 ± 1.5	4.9 ± 1.4
T. latior	2.0	1.9	2.2 ± 1.1	2.1 ± 1.0
T. hispidissima	3.4	3.2	3.2 ± 1.1	3.0 ± 1.1
T. gossypina clade	4.4	4.2	4.7 ± 1.0	4.5 ± 1.0
T. gossypina + *T. turneri*	3.8	3.6	3.9 ± 1.1	3.7 ± 1.0
T. gossypina.263 + *T.turneri*	2.7	2.6	2.8 ± 0.9	2.7 ± 0.8
T. canescens + *T. greggii*	14.6	13.9	14.4 ± 2.4	13.6 ± 2.3

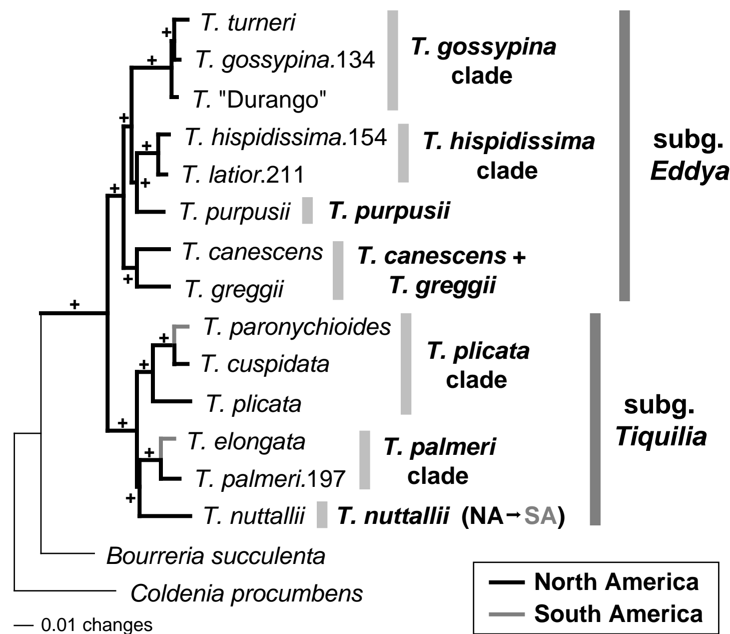

Fig. 2.1. Phylogram of the tree resulting from the maximum likelihood search of the 5-marker combined data set (lnL = -19287.127), with a reconstruction of dispersal history mapped onto the tree. Plus signs (+) above or below the branches indicate Bayesian posterior probabilities of 1.0. The distributions of outgroup taxa are not shown because they have either Old World or pantropical distributions and because they do not influence the reconstruction of dispersal history in *Tiquilia*. The single dispersal event from North America to South America in *T. nuttallii* is not shown on the tree. The Galápagos taxa of subg. *Tiquilia* were not included in this analysis, but their absence does not affect the North American origin of all major lineages of *Tiquilia*.