**Title**

The genus Artemisia: Biogeography and chemosystematics in a phylogenetic context (phylochemistry)



**Author**

Hobbs, Christopher Rollin

**Publication Date**

2014

Peer reviewed|Thesis/dissertation

# The genus *Artemisia*: Biogeography and chemosystematics in a phylogenetic context (phylochemistry)

by

Christopher Rollin Hobbs

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

In the GRADUATE DIVISION

of the

University of California, Berkeley

Committee in charge:

Professor Bruce Baldwin, Chair
Professor Brent Mishler
Professor Allen Goldstein

Fall 2014

Abstract

The genus *Artemisia*:
Biogeography and chemosystematics in a phylogenetic context (phylochemistry)

By Christopher Rollin Hobbs

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Bruce G. Baldwin, Chair

*Artemisia* is the most diverse genus in tribe Anthemideae, consisting of 350 to 500 species (Hillebrand, 1888; Vallès & McArthur, 2001; Vallès *et al.*, 2003). These wind-pollinated herbaceous and shrubby taxa, including mugworts, sagebrushes, and wormwoods, occur widely in the Northern Hemisphere and sparingly in southern South America and sub-Saharan Africa. I studied the biogeography of the genus with an emphasis on the three Hawaiian taxa in the context of world-wide distribution.

I sought to determine whether Hawaiian *Artemisia* (Compositae–Anthemideae), with lowland and subalpine species, represents such an example by investigating the origin and relationships of the Hawaiian taxa.

Molecular phylogenetic analyses of Hawaiian *Artemisia* were conducted using nuclear ribosomal transcribed spacers, chloroplast DNA intergenic spacers, and morphology. Timing of divergence events associated with inferred dispersals was estimated with calibration from fossil pollen records. Historical biogeographic analyses based on molecular trees and ecological modeling of distributions of extant taxa were used to aid interpretation of geographic and habitat shifts associated with diversification and long-distance dispersal.

The findings indicate that the Hawaiian-endemic species (*A. australis*, *A. kauaiensis*, and *A. mauiensis*) constitute a clade that is sister to southeast Asian *A. chinensis.*, which, like the Hawaiian endemics, has ribbed fruit walls and, unlike other members of *Artemisia* except *A. kauaiensis*, has a distinct pappus, often associated with dispersal ability in Compositae. The clade encompassing *A. chinensis* and Hawaiian *Artemisia* was resolved to be most likely of Asian origin.

An ecological shift in Hawaiian *Artemisia* from tropical coastal habitats to drier and colder

subalpine slopes is consistent with evidence from recent studies by Tkach and others, and data reported here, for repeated colonisation of the arctic by diverse lineages of *Artemisia*. *Artemisia* appears to be prone to such anticlimatic ecological shifts, which may explain this exceptional example of an ancestrally lowland tropical lineage in the Hawaiian high-montane flora.

Since the genus *Artemisia* has an abundance of secondary metabolites (SM), from a variety of distinct biosynthetic pathways, I sought to determine if these compounds had biogeographical signal, and were more influenced by environment, or mostly under genetic control and conserved. The odor compounds, or volatile organic constituents (VOCs), with up to 200 possible in one species was characterized by gas chromatograph/mass spectroscopy, and are a strong characteristic of the genus. Based on their odors, they seem to be distinct and varied within individual species.

First I sought to determine if compounds from all major constituent classes (alkaloids, terpenes, fatty acids, phenolic compounds) were under environmental or genetic control. The use of SM for solving taxonomic problems is called chemotaxonomy or chemosystematics. Use of SM in systematics, by tradition called chemosystematics, has been extensive for more than a half century and predates macromolecular approaches.

A comprehensive search of the chemosystematics literature was undertaken, and a previously published dataset of 24 leaf fatty acids for 123 taxa from all major gymnosperm clades (cycads, ginkgos, gnetophytes, and conifers) was compared with trees from *rbc*L chloroplast sequence data for the same taxa using comparative analyses in R to determine phylogenetic signal and to evaluate the mode of evolution of the chemical traits. Strong evolutionary constraint of the fatty acid data adds weight to the idea that fatty acids are more suitable for use as characters in phylogenetic analyses than non-structural secondary compounds. The relative benefits and drawbacks of chemical data are briefly discussed, along with potential problems in their collection, extraction, analysis, and interpretation.

Volatile organic constituents are prominent in the genus *Artemisia*, and many of these volatile compounds such as thujone and 1,8-cineol, are associated with defensive behavior such as anti-herbivory, anti-encroachment, and antimicrobial effects in plants. I sought to test the hypothesis that these toxic compounds might be lost in island settings over evolutionary time because of a demonstrable lack of herbivores and other encroaching species within the time frame of early colonization. I show show a dramatic example of island tameness in plants in *Artemisia*. In the 74 species sampled across the genus, 10 abundant and widely-studied, toxic terpene compounds that are known to inhibit seed germination, root development, and herbivore feeding[9] demonstrate repeated loss, in two separate island lineages in the Hawaiian and Canary Islands. This is the first phylogeny-based study demonstrating this in plants, whereas island tameness has been widely discussed in animals, and well-known examples such as the dodo of Mauritius, Galápagos island iguanas are common knowledge. The tameness of the Falkland Island wolf, was noted by Darwin.

I dedicate this work to my dad, Ken Hobbs, Ph.D., the second scientist in the family. He taught me to identify plants when I was very young, and inspired my love for the natural world that I have carried with me throughout my life. He was inspired by his mother's brother, Wyatt Jones, the first botanist and scientist in the family. It brings me great joy to carry on the lineage. With hopes that my son Ken will continue the passion.

Table of Contents

**Acknowledgements**

I acknowledge support from the Lawrence R. Heckard Endowment Fund of the Jepson Herbarium (to CH). Special thanks to Mary Guo, Gracie Martin, and Monica Neff for critical lab assistance, and Ben Carter, and especially Nick Matzke for statistical support. I also thank John L. Strother for helpful reviews; Napua Harbottle, Will McClatchey, Abby Moore, Cliff Morden, Mike Park, John Twibell, Bridget Wessa, and Ken Wood for helpful field, herbarium, or laboratory assistance; curators of BISH, JEPS, MO, PTBG, RM, SD, UC, and UPS for access to herbarium specimens; Cliff Morden and the Hawaiian Plant DNA Library collection (Cliff Morden) for DNA samples; and the Plant Heritage/ NCCPG *Artemisia* collection (John Twibell), University of California Botanical Garden, and Regional Parks Botanical Garden (Tilden Regional Park) for plant material.

It was a great joy and honor to work with my advisor, Bruce G. Baldwin, whose love of plants and great knowledge of ecology, evolutionary biology, systematics, and phylogenetics inspired and guided me through the long process of my science training at Cal. Also great thanks to Brent Mishler for his mentorship, council, and teachings. Thanks to the third member of my dissertation committee, Allen Goldstein for his VOC work and support.

My labmates Mike Park and Abby Moore were instrumental in getting me on the right track, and answering a million questions. They were excellent friends as well and my time in the lab was much the richer for their being there. Thanks to Bridget Wessa for her excellent lab training and abundant experience.

To the professors, Staff, GSIs, and students of the Integrative Biology Department—what a long and crazy party!

# Chapter 1

## Asian origin and upslope migration of Hawaiian *Artemisia* (Compositae—Anthemideae)

## INTRODUCTION

Origins of Hawaiian flora are becoming increasingly well known from molecular phylogenetic studies (see Baldwin & Wagner, 2010; Price & Wagner, 2011) and a predictive framework for examining long-distance dispersal to the archipelago has been proposed based in part on the growing body of phylogenetic data for Hawaiian organisms in general (Gillespie *et al*., 2012). Recent evidence for a much larger contribution of temperate and boreal American lineages to Hawaiian plant and animal diversity than earlier believed has led to a focus on high-elevation habitat in the Hawaiian Islands as an important source of ecological opportunity for colonists from higher latitudes (Ballard & Sytsma, 2000; Baldwin & Wagner, 2010; Gillespie *et al*., 2012). Although ecological diversification has been resolved in various Hawaiian plant clades (e.g., Baldwin & Robichaux, 1995; Givnish *et al*., 2009), evidence for ecological shifts from lowland tropical settings to the cooler and drier conditions associated with alpine or subalpine vegetation of the high volcanoes is lacking for Hawaiian plants, as noted earlier by Skottsberg (1931), notwithstanding the wide elevational distribution of some endemic Hawaiian species (Wagner *et al*., 1999). Examples of such shifts are of interest to understand the relative importance of niche conservatism versus ecological lability in the evolution and assembly of Hawaiian vegetation.

One of the few Hawaiian plant genera with high-montane endemism for which phylogenetic data have been lacking is *Artemisia* L. (Compositae). *Artemisia* is the most diverse genus in tribe Anthemideae, consisting of 350 to 500 species (Hillebrand, 1888; Vallès & McArthur, 2001; Vallès *et al.*, 2003). These wind-pollinated herbaceous and shrubby taxa, including mugworts, sagebrushes, and wormwoods, occur widely in the Northern Hemisphere and sparingly in southern South America and sub-Saharan Africa. As many as seven endemic taxa of *Artemisia* have been described from the Hawaiian archipelago (see Hillebrand, 1888; Degener & Degener, 1946; Skottsberg, 1927; Shultz, 1999), three of which have been recognised recently (Shultz, 1999; Wagner *et al*., 2005): *A. australis* Less., *A. kauaiensis* (Skottsb.) Skottsb., and *A. mauiensis* (A. Gray) Skottsb.

The geographic origin(s) and closest relatives of Hawaiian *Artemisia* have long been obscure. Guppy (1903) suggested that the Hawaiian taxa descended from an ancient colonist from North America. Skottsberg (1931) hypothesised that the Hawaiian taxa descended from a boreal ancestor referable to sect. *Abrotanum*, now treated within sect. *Absinthium*. Fosberg (1948) also suggested a boreal affinity for the Hawaiian endemics, one of which (*A. mauiensis*) is restricted to subalpine settings on Maui. Ballard & Sytsma (2000), who demonstrated a boreal (probably North American) ancestry for Hawaiian *Viola* (Violaceae), listed *Artemisia* among 15

genera in the Hawaiian Islands that warranted study for possible boreal ancestry based on having arctic congeners. The high proportion of bird taxa from boreal regions among those that migrate from continental areas to the Hawaiian Islands indicates that the arctic or subarctic is likely a relatively rich source of avian-dispersed propagules reaching the archipelago (Ballard & Sytsma, 2000; Baldwin & Wagner, 2010; Gillespie *et al*., 2012).

Recent molecular studies have allowed for refinements of Bremer & Humphries's (1993) morphology-based circumscriptions of *Artemisia* and its subgenera *Dracunculus* (including tarragon and relatives), *Seriphidium* (shrubby Asian taxa), and *Tridentatae* (shrubby North American taxa) (Watson *et al*., 2002; Sanz *et al*., 2008; Garcia *et al*., 2010). More work is needed to resolve subg. *Artemisia* (see Sanz *et al*., 2008; Tkach *et al*., 2008a), which now includes subg. *Abrotanum*, and has been suggested recently to include the three Hawaiian endemics (Shultz, 1999). To date, the Hawaiian taxa have not been included in any published molecular analyses.

Here, I investigate Hawaiian *Artemisia* from molecular, morphological, and ecological perspectives to: (1) discern whether the Hawaiian taxa constitute a clade or arose from independent introductions; (2) determine the sister clade(s) of the Hawaiian species; (3) consider the most likely source area(s), migration route(s), and dispersal mechanisms for the Hawaiian colonising ancestor(s); and (4) examine adaptive evolutionary and biogeographic trends in the Hawaiian *Artemisia* clade in relation to such trends in *Artemisia* as a whole. I was especially interested in determining whether the occurrence of Hawaiian *Artemisia* from sea level to subalpine habitats is an example of niche conservatism from lineages of different ancestral ecology or, alternatively, an exceptional instance of extreme ecological change across the topographic and climatic gradient in the Hawaiian Islands, and, if so, why.

## MATERIALS AND METHODS

### Taxon sampling

Taxon samples in this molecular phylogenetic analysis represent all recently recognised subgenera and sections of *Artemisia* (i.e., subgenera *Artemisia*, *Absinthium*, *Dracunculus*, *Seriphidium*, and *Tridentatae*) and all geographic regions where the genus is native.

DNA sequences of the 18S-26S nuclear ribosomal internal transcribed spacer region (ITS-1, 5.8S subunit, and ITS-2; hereafter, ITS), the external transcribed spacer upstream of the 18S subunit (ETS), and the chloroplast *trn*L-*trn*F and *psb*A-*trn*H intergenic spacer regions (including some flanking exon sequence) were obtained from fresh or silica-dried plant collections and herbarium specimens. Voucher information and GenBank accession numbers of sequences are presented in Appendix S1).

## DNA extraction, amplification and sequencing

DNA was extracted from leaf material using the Qiagen DNEasy Plant Mini-Kit, and for two accessions was obtained from the Hawaiian Plant DNA Library (Randell & Morden, 1999). Primers used for PCR amplification were Ast-1 (Markos & Baldwin, 2001) and 18S-ETS (Baldwin & Markos, 1998) for ETS, and ITS-I and ITS4 (Urbatsch *et al.*, 2000; White *et al.*, 1990) for the ITS region. cpDNA primers e and f (Taberlet *et al.*, 1991) were used to amplify the *trn*L-*trn*F intergenic spacer region, and psbA (Sang *et al.*, 1997) and trnH-GUG (Tate & Simpson, 2003) were used to amplify the *psb*A-*trn*H intergenic spacer region.

PCR was performed on an MJ Research PTC-200 thermal cycler. For ITS amplification, the initial temperature was 95° C for 2 minutes, then 40 cycles of 96° C for 10 sec, 48° C for 30 sec, and 72° C for 20 sec with a 4 sec extension per cycle. Final extension was at 72° C for 7 minutes. For ETS, the initial hold at 96° was for 1 minute, followed by the same cycling conditions used for amplifying the ITS region except for an annealing temperature of 60°. The chloroplast gene regions were amplified with an initial temperature of 94° C for 3 minutes, then 29 cycles of 94° C for 1 minute, 55° C for 1 min, 72° C for 1 min, followed by 72° C for 5 min. Exo-SAP (Affymetrix) was used to clean the products for sequencing, and Sanger sequencing was performed at the University of California, Berkeley DNA Sequencing Facility, Berkeley, CA. The unedited sequences were processed with Chromas Pro (Technelysium Pty Ltd) and Geneious ver. 4.8.5 (Drummond *et al.*, 2009) to check base identities on each strand. Sequences were aligned manually using Simmon's (2004) similarity criterion in Seaview ver. 4.32.0.0 (Galtier, 1996), in comparison with an automated alignment using MAFFT (Katoh, 2009; Miller *et al.*, 2009, 2010).

## Morphological character analysis

Morphological characters and character states were evaluated from fresh plant material and herbarium specimens. Type specimens of *Artemisia australis*, *A. kauaiensis*, and *A. mauiensis*, including a photograph of the holotype of *A. kauaiensis* (destroyed during World War II, Berlin, Germany), were examined at the Bishop Museum (BISH) and at the National Tropical Botanical Garden (PTBG) herbarium. Skottsberg's observations made from type specimens for these taxa were also consulted (Degener & Degener, 1946). Many of the taxa included in this study were also examined from collections in the University Herbarium (UC), University of California, Berkeley, including two specimens of *A. chinensis*. Floristic treatments consulted to augment character evaluation of taxa included those in the *Flora of North America, North of Mexico* (Shultz, 2006), *The Flora of China*, *Asteraceae, Tribe Anthemideae* (Ling *et al.*, 2008), *Manual of the Flowering Plants of Hawai'i* (Wagner *et al.*, 1999), the *Flora of Siberia* (Poljakov, 1995), *Flora of the USSR* (Poljakov, 1961), and *Flora Europaea* (Tutin *et al.*, 1976). See Appendix S2 for a summary of the morphological characters and character states examined and used in this

study. For the Hawaiian taxa and the southeast Asian *A. chinensis*, the fruits (cypselae) were photographed with a Microptics XLT Digital Imaging System.

**Phylogenetic analyses**

For the molecular data, nDNA (ITS, ETS) and cpDNA data sets were analysed separately and together to investigate phylogenetic congruence between the nuclear and chloroplast regions. Phylogenetic analyses using both maximum likelihood (ML) and Bayesian inference (BI) were performed within CIPRES (Cyberinfrastructure for Phylogenetic Research) Portal versions 2.2 (Miller *et al.*, 2009, 2010) for all data sets. Maximum likelihood (ML) tree searches with bootstrapping were performed on the separate and combined cpDNA and nDNA datasets. The ITS + ETS dataset included 114 taxa, and, based on the more limited number of high-quality cpDNA sequences that could be obtained from herbarium specimens, the combined cpDNA and nDNA dataset included 84 taxa, with good representation of all subgenera and sections.

The ML tree searches were performed with PHYLIP-formatted input using RAxML-VI-HPC, ver. 7.2.6 (Stamatakis, 2006; Stamatakis *et al*., 2008) via the RAxML GUI 0.9 beta 2 user interface (Silvestro & Michaliak, 2010), within CIPRES (Cyberinfrastructure for Phylogenetic Research) Portal, version 2.2 (Miller *et* al., 2009, 2010). The best-scoring ML tree with clade support values was obtained from 10 independent runs, with 500 LSR (lazy subtree rearrangement) bootstrap replicates each. RAxML chose model parameters, including alpha, invar, and rates, with an accuracy of 0.001 Log Likelihood units for each of the 10 independent runs, and then the best-fitting parameters were chosen for final optimization at the end of the run.

The BI analyses were performed with the NEXUS-formatted input using MrBayes, ver. 3.1.2 (Huelsenbeck & Ronquist, 2001) within CIPRES Portal 2.2, with 20 million Markov chain Monte Carlo (MCMC) generations and four independent runs (nruns = 4), including three heated chains and one "cold" chain per run (nchains = 4). The sample frequency was set to 1000, and the heating parameter, 0.20. The best-fitting DNA substitution model chosen by MrModeltest 2.3 (Nylander, 2004), GTR+I+G for the combined dataset as well as for the individual cpDNA and nDNA datasets, was implemented. Clade credibility values (posterior probabilities) for the combined dataset were obtained from a 50% majority-rule consensus of the retained trees (burnin = 0.25, 40,004 trees produced, 30,004 trees sampled) from the 4 independent runs. Graphing the standard deviation of split frequencies was done to determine when stationarity was reached (at 40,000 generations), and analyses were run until the average standard deviation of split frequencies dropped below 0.01 (by 13.8 million generations, reaching 0.005 at the end). The separate nDNA (ITS, ETS) 114-taxa and cpDNA analyses were performed with the same parameters; for the nDNA analysis the average standard deviation of split frequencies dropped below 0.01 after 4.78 million generations, and reached 0.004 after 20 million generations. For

the cpDNA BI analysis, the average standard deviation of split frequencies was 0.009997 at 10.2 million generations, reaching 0.006 at 20 million generations. Tracer, ver. 1.5 (Rambaut & Drummond, 2007) and AWTY Online (Wilgenbusch *et al.*, 2004) were both used to confirm stationarity. Tracer plotted the ln likelihood, and AWTY was used to check the split frequencies of the 4 runs.

Phylogenetic analysis of the morphological data matrix was performed using PAUP* 4.0 (Swofford, 2003) within CIPRES ver. 3.1 (Miller *et al.,* 2009, 2010). The starting tree was produced by stepwise addition. The heuristic search used TBR branch swapping with 200 iterations of parsimony ratchet (Nixon, 1999). Consense, ver. 3.66 (Felsenstein, 2008) was then used within CIPRES to calculate a strict consensus tree (of 6,015 trees).

**Divergence time analysis**

To estimate divergence times pertinent to origin of the Hawaiian clade, a BI analysis of the 114-taxa nDNA (ETS and ITS) data set was performed using BEAST (Drummond & Rambaut, 2007; Drummond *et al.*, 2002), with an uncorrelated lognormal relaxed molecular clock model and time calibration based on palynological fossil records. Age estimation for the most recent common ancestor of the Hawaiian–*A. chinensis* clade was carried out through Bayesian stochastic modeling of the rates of molecular evolution independently on each tree branch (relaxed molecular clock), with Bayesian statistical analyses that allow incorporation of prior knowledge, in this case of *Artemisia* fossil pollen data reported for Eurasia and North America (see below). A relaxed clock was chosen because the coefficient of variation as determined with Tracer (Rambaut & Drummond, 2007) indicated significant rate heterogeneity across branches, i.e., a molecular clock was rejected (Drummond & Ho, 2007).

The age constraint calibration was applied as a Bayesian prior with an exponential parametric statistical distribution, as suggested for modeling data with fossil calibration (Ho, 2007). The distribution mean of 31 Ma was chosen so that 95% of the probability in the distribution was constrained by a rigid lower bound (32 Ma in the case of the monophyletic group *Artemisia*), with a flexible maximum. The exponential distribution describes the rapid decrease in probability for the maximum age as the distance between the estimated nodal age and the fossil calibration point increases.

A Yule process of speciation, and best-fit model of sequence evolution for the 114-taxa nDNA (ITS and ETS) dataset (HKY+G from Bayesian Information Criterion in JModelTest 0.1.1, Posada, 2008; Guindon & Gascuel, 2003), with 4 gamma categories, was chosen for the BEAST analysis. BEAUti 1.6.2 was used to generate the XML file for input to BEAST. Each of 4 independent MCMC analyses was terminated at 30 million generations (saving 1 tree per 1000 generations), with a burn-in of 10%. Tracer was used to check for stationarity, which was reached after 500,000 generations. The 4 independent runs were combined in LogCombiner

v1.6.2 (in BEAST package). TreeAnnotator was used to read in 120,004 trees and ignore the first 500 (identifying 13,172 unique clades), and to find and annotate the maximum clade credibility tree (i.e., the sum of all posterior probabilities of all clades of each tree and an estimate of the total probability of the tree topologies), which was displayed by FigTree (Rambaut, 2009).

**Time calibration of molecular trees**

Use of fossil pollen to calibrate the most recent common ancestor of modern taxa of *Artemisia* at 31 Ma was based on extensive palynological evidence of the genus from Cenozoic records. Fossil pollen of *Artemisia* has been widely reported from sites in Asia, Western Europe, and North America, and is easy to identify because ornamentation with short spinules is a recognisable taxonomic marker for the genus and its close relatives (Martin *et al*., 2001, 2003). *Artemisia* macrofossils have not been reported in the literature, likely because the leaves are delicate and frequently wither before being shed (Axelrod, 1987).

Three credible lines of fossil evidence were the basis for determination of the first occurrence of *Artemisia* pollen, from central Asia. These include (1) a stratigraphic study of samples taken from the Pavlodar section of the Irtysh Valley, in the northern Aral Region of the Turgai Lowland in present-day west central Kazakhstan (Zaklinskaya & Leopold, 1957), (2) an integrated paleomagnetic and stratigraphic study of fossils found in the Hoh Xil basin in the northern Qinghai-Tibet Plateau (Liu *et al.*, 2003), and (3) a biostratigraphic and paleomagnetic study in the Qaidam Basin (Zhu *et al.*, 1985; Song *et al.*, 2004; Wang, 2004). In these studies, *Artemisia* pollen was identified in deposits from the early Oligocene, about 30–32 Ma. *Artemisia* pollen was not found in earlier layers at the same sites. In another high-resolution magnetostratigraphic study from the Xining Basin, Tibetan Plateau, pollen of Compositae, but not *Artemisia,* was identified from the Eocene-Oligocene boundary at about 34 Ma (Dupont-Nivet et al, 2007, 2008), with evidence of xerophytic shrubland similar to what would later include *Artemisia* in abundance.

Fossil pollen records for North America were compared to tree-based age estimates for the North American clade, thereby allowing validation of the basal calibration chosen for *Artemisia*. In North America, the earliest reported Compositae pollen is from lower Miocene (ca. 23 Ma) sediments in the Pacific Northwest (Gray, 1958), coinciding with world-wide appearance of Compositae pollen in other Northern Hemisphere sites. Pollen records from three credible fossil floras (Split Rock, the Troublesome Formation, and Succor Creek) suggest that *Artemisia* likely made its first appearance in North America in the mid-Miocene, about 16.5–18 Ma, based on stratigraphic dating, supplemented with radiometric dating (Taggart & Cross, 1980; Cross & Taggart, 1982; Leopold & Denton, 1987).

***Biogeographic analyses***

The biogeographic range of each taxon was determined from the same floras that were consulted for the morphological analysis. These ranges were coded as a multistate character in Mesquite as [1 = Asia, 2 = North America, 3 = South Africa, 4 = South America, 5 = Circumboreal, 6 = Hawaiian Islands, 7 = Middle East, 8 = North Africa, and 9 = Europe]. The characters were traced over the "best tree" found by RAxML for the data set that included both nDNA (ETS, ITS) and cpDNA sequences, with ancestral states estimated using the parsimony criterion in Mesquite (Fig. 1). The same molecular phylogeny and biogeographical range matrix was analysed using Lagrange (Ree *et al.*, 2005; Ree & Smith, 2008), version 20090327, to estimate rates of dispersal and extinction and the likelihoods of different areas of ancestral origin at each node.

**Climate modeling**

For *Artemisia chinensis* (= *Crossostephium chinense*), georeferenced occurrence data from WorldClim for 22 locality records (GBIF, 2011) were used to predict the likely distribution of the taxon in the same area as its known distribution, based on habitat similarity, using Maxent (Phillips *et al.*, 2011). Nineteen environmental variables such as mean, maximum, and minimum temperature and rainfall for each month were considered, at two different levels of resolution, 1 km or 2.5 arc-minute (Hijmans *et al.*, 2005). Maxent finds the probability distribution of maximum entropy, and has been reported to perform well compared with the accuracy of other similar methods (Ortega-Huerta & Peterson, 2008), even when collection and observation data is limited (Benito *et al*., 2009). A recent model comparison (Elith *et al.*, 2006) found Maxent to be the best-performing model algorithm among sixteen.

Maxent was also used to statistically model habitat suitability based on presence-only data, with associated environmental variables, of geographical areas outside the known range of *A. chinensis*, to determine whether the Hawaiian Islands would provide similar environmental conditions to specific locations around Taiwan and surrounding Pacific islands where *A. chinensis* is known to occur. Geographic maps were generated using Maxent to show estimates of probability of occurrence, where the lowest value, 0, is blue, and higher probabilities, with a high of 1, have increasingly warmer colours. Maxent was set to a convergence threshold of $10^{-5}$, 5,000 iterations (to ensure convergence was reached), and 10,000 background points. Predictive accuracy of the model was not determined by the area under the ROC (receiver operating characteristic curve), the AUC, which measures discrimination ability (Gibson *et al.*, 2007), because it may not be suitable with presence data alone (Boyce *et al*., 2002) due to bias and other problems (Lobo *et al.*, 2008). To assess the predictive accuracy of models using small sample sizes (for my study, 11 training samples and 4 test samples), I used the jackknife validation approach of Pearson *et al*. (2007), who found that Maxent had significant predictive ability with only 5 locality points and that 11–17 records could accurately predict 90 to 100% of presences

when removed sequentially in random order.

Principal components analysis (PCA) was performed in R (R Development Core Team, 2008) to determine the most influential derived climate variables from WorldClim consistent with the clustering of specimen data in climate space. A MANOVA analysis also was conducted in R to determine the significance of separation by species of these climate data extracted at each specimen location.

## RESULTS

Fig. 1 shows the ML tree from analysis of combined nDNA and cpDNA sequences. Comparison of trees and clade support from separately analysed nDNA and cpDNA datasets showed that phylogenetic incongruence between nuclear and chloroplast data was not robust or was not relevant to relationships among major clades, including the Hawaiian clade (Appendices S3, S4). The ETS + ITS analyses (without the cpDNA sequences) resulted in the BI tree in Appendix S3. The clade that includes only the 3 Hawaiian taxa and *A. chinensis* received 100% BI PP and ML bootstrap support in both the nDNA and combined data trees (Fig. 1, Appendix S3). BI and ML analyses of the cpDNA (*psb*A-*trn*H + *trn*L-F spacer) sequences also resolved a sister-group relationship between *A. chinensis* and the 3 Hawaiian taxa, with weak support (see TreeBASE (accession number to be forthcoming)). Clades corresponding to the major subgenera of *Artemisia* were resolved with high support by both ML and BI analyses of the combined dataset (Fig. 1).

A heuristic search of tree space based on the morphological data under the parsimony criterion resolved four major subgenera of *Artemisia* as clades (see TreeBASE (accession number to be forthcoming)), as well as the clade including the Hawaiian taxa and *A. chinensis*. The characters listed in Appendix S2 include those that have been used to diagnose the subgenera in major floras (Poljakov, 1995; Ling, 2006; Tutin *et al*., 1976; Shultz, 2006). The morphological analysis revealed that the Hawaiian taxa and the Asian *A. chinensis*, unlike other members of *Artemisia*, have distinctly ribbed cypselae, and *A. chinensis* and *A. kauaiensis* are further distinguished within the genus by having pappus crowns that include irregular, sharp, curved teeth (Figs. 2, 3).

Parsimony mapping of geographic ranges in Mesquite produced the results in Fig. 1, with the Hawaiian taxa resolved as having a common Hawaiian origin. This analysis indicated an Asian origin for the Hawaiian taxa. The probability that the common ancestor of the 3 Hawaiian taxa and *A. chinensis* was Asian is 0.9 based on the likelihood analysis using Lagrange. The probability that the 3 extant Hawaiian taxa all derived from a common Hawaiian ancestor is 1.0 based on the same analysis.

The probability of suitable habitat for *A. chinensis* occurring close to the geographic

distribution of Hawaiian taxa of *Artemisia*, based on the Maximum Entropy model of prediction (implemented by the software Maxent), is ~0.9 for the westerly Leeward Islands (at the western end of the Hawaiian chain, see Fig. 4), and 0.6 for Johnston Atoll and the most easterly part of the of the Leeward Islands (Lisianski and Laysan). Areas with a high probability (> 0.9) of having suitable habitat for *A. chinensis* in southeast Asia include the coasts of Guangdong and Fujian Provinces, China (directly west of Taiwan), the Ryukyu Islands, Okinawa, southern Japan, and all of the Bonin Islands. Most of the Pacific Islands show a 0.5–0.7 probability of climatic suitability. The most important contributors to these predictions were the mean diurnal temperature range [mean of monthly (maximum temp - minimum temp)], 55.4%, mean temperature of warmest quarter, 14.4%, the annual mean temperature, 11.1%, minimum temperature of coldest month, 8.9%, and the precipitation seasonality (coefficient of variation), 2%. Maxent predicted suitable habitat for *A. chinensis* with low omission rates, which was well fit to the predicted omission rate. Pearson's pvaluetest software computed the correlation between the prediction success rate and the probability of success under randomness as 90% with a p-value of $< 0.0001$.

Projection of potential habitat for *A. chinensis* in the low-elevation Leeward Islands, but not the high islands of the Hawaiian chain, by the Maxent analysis is consistent with significant separation of *A. chinensis* and the Hawaiian taxa in the PCA and MANOVA analyses along vectors associated with degree of climatic tropicality (isothermality, mean diurnal range of temperature, and temperature seasonality). Significant separation of mean climatic data for Hawaiian *A. australis* and *A. mauiensis* in the same analyses was primarily along vectors associated with the elevational gradient in temperature and humidity (mean temperature of driest quarter, mean temperature of coldest quarter, and minimum temperature of coldest month), with drier and colder temperatures in subalpine habitats of *A. mauiensis* than on lower slopes, where *A. australis* is found (results not shown).

Using BEAST, the mean root height of the clade that includes the Hawaiian taxa and *A. chinensis* was estimated at 3.93 Ma with a standard error of 0.041 and an effective sample size (ESS) of 2,339. The 95% highest posterior density (HPD; the set containing 95% of samples or draws from the posterior distribution) was found to be 1.03–7.96 Ma. Age of the clade with only the 3 Hawaiian taxa was estimated as 1.45 Ma. The BEAST age estimate for the most recent common ancestor of the North American clade corresponding to subg. *Tridentatae* was 14.3 Ma with a standard error of 0.119 and ESS of 465 and a 95% HPD of 9.4 to 19.5 Ma. The mean mutation rate estimated by BEAST was $1.828 \times 10^{-9}$ substitutions per site per year, with a standard error of $1.30 \times 10^{-11}$.

The dates provided by the BEAST analysis must be viewed cautiously because the calibration date, input as prior knowledge, has some uncertainty. The oldest known (31 Ma) *Artemisia* pollen was found in central Asia (see Materials and Methods, "Time Calibration of Molecular Trees"). The BEAST date of 14.3 Ma calculated for the North American clade

corresponding to subg. *Tridentatae* is consistent with the first reported appearance for *Artemisia* pollen in the North American fossil record. Although I have calibrated the root of the *Artemisia* clade with a date of 31 Ma, it is possible that pollen samples from this time represent ancient taxa from a branch within the genus and not the root.

## DISCUSSION

### Relationships of Hawaiian *Artemisia*

Both nDNA and cpDNA evidence supports the novel hypothesis that *Artemisia chinensis* of southeast Asia is the closest living relative of the Hawaiian *Artemisia* taxa, which in turn constitute a clade. Gray (1884) originally hypothesised a close relationship between *A. chinensis* and the Hawaiian *A. australis* (prior to the description of additional Hawaiian species), notwithstanding his confusion about the relationship of *A. californica* (subg. *Tridentatae*) to *A. chinensis* and *A. australis*: "*Here (with* A. californica*) belongs* A. australis *Less. of the Hawaiian Islands, as well as the anomalous* A. chinensis *L.* (p. 370)." Gray's conclusion was based on his observation of ribbed fruits in all three taxa; however, fruits of *A. californica* are often only faintly or not ribbed.

The nested position of *A. chinensis* in a well-supported clade with the Hawaiian taxa and with members of subgenera *Artemisia* (Asian), *Tridentatae* (North and South American), and *Seriphidium* (Asian) confirms that the anomalous *A. chinensis* belongs in the genus *Artemisia*, where it was originally described (Linnaeus, 1735), and not in the monotypic genus *Crossostephium* Makino, where it is often placed. The southeast Asian–Hawaiian clade that consists of the 4 closely-related taxa *A. chinensis*, *A. australis*, *A. kauaiensis*, and *A. mauiensis* is evidently sister to multiple subgenera of *Artemisia* and is best treated as a new subgenus, ***Artemisia* subg. *Hawaiiana* C.R. Hobbs** (type species: *Artemisia australis* Less.). Morphologically, subg. *Hawaiiana* is diagnosed by all four having a small shrubby habit and trailing branches, with leaves clustered near the tips, and glandular, conspicuously 5-ribbed fruits (cypselae), sometimes with irregular pappus teeth (in *A. chinensis* and *A. kauaiensis*).

### Biogeographic origin of the Hawaiian clade

The Lagrange likelihood results indicate that the common ancestor of the *A. chinensis*–Hawaiian (subg. *Hawaiiana*) clade was likely Asian, as resolved by parsimony mapping, as well. The same analyses provide support for a common Hawaiian origin of the Hawaiian endemics, *A. mauiensis*, *A. kauaiensis*, and *A. australis*, by either a single dispersal event or multiple dispersals closely spaced in time (prior to insular diversification), by an ancestor of Asian origin. *Artemisia* is unusual among Hawaiian angiosperms with arctic congeners and high-elevation island endemics in not having a boreal or temperate North American origin (see Baldwin &

Wagner, 2010). Skottsberg (1931), who believed that barriers to dispersal were too great to allow for North American origins of Hawaiian angiosperms (e.g., Skottsberg, 1925; see Baldwin & Wagner, 2010), appears in this instance to have been correct in surmising an Old World origin of Hawaiian *Artemisia*, in contrast to the American origin of *Artemisia* (and other Compositae) suggested by Bentham (1873), Wallace (1881), Guppy (1903), and Brown (1922).

Estimated divergence times for subg. *Hawaiiana* and the Hawaiian clade indicate that *Artemisia* reached the Hawaiian Islands and diversified there within the timeframe of the modern high islands. As discussed by Baldwin & Wagner (2010), early biogeographers, such as Wallace (1881), Guppy (1903), and Brown (1922), viewed Hawaiian Compositae as especially ancient (e.g., Mesozoic or early Cenozoic) elements in the Hawaiian flora. Since the application of molecular data to questions of Hawaiian origins, most endemic angiosperm clades, including those in Compositae, appear to be contemporary with the high islands (Price & Clague, 2002). The BEAST analysis finds the posterior distribution of the age of the most recent common ancestor of the *A. chinensis*–Hawaiian *Artemisia* (subg. *Hawaiiana*) clade to have a mean of 3.93 Ma, and the Hawaiian lineage to have a mean of 1.45 Ma. These dates place the introduction of the Hawaiian *Artemisia* ancestor after the emergence of Kaua`i, dated radiometrically by McDougall (1979) at 5.1 Ma, and, at earliest, about the same time as the emergence of O`ahu (Waianae volcano) at 3.7 Ma (Doel & Dalrymple, 1973). The results therefore indicate that initial colonisation of the archipelago probably did not predate the oldest modern high island, Kaua`i, where *A. australis* and *A. kauiensis* occur, or the lower island of Ni`ihau (4.9 Ma; Clague & Dalrymple, 1987), where *A. australis* is documented.

As noted below, presence of pappus is often associated with enhanced dispersal ability in Compositae and may well be ancestral in the Hawaiian clade, where selection has been implicated in reduced dispersal ability by loss of such structures in various Pacific lineages of the family (Carlquist, 1966). According to this interpretation, the absence of pappus in *A. australis* and *A. mauiensis* is a derived state, and island colonisation history is in keeping with Funk & Wagner's (1995) progression rule of older-to-younger island dispersal and diversification, which has been followed repeatedly by Hawaiian plant and animal groups.

**Dispersal routes and mechanisms**

The early suggestion by Engler (1879) and Hillebrand (1888) that the highly endemic Hawaiian flora must descend from ancestors that arrived via long-distance dispersal has been supported by geological evidence of long-term isolation of the volcanic chain from source areas for terrestrial organisms (e.g., Clague, 1996). In addition, molecular phylogenetic evidence indicates that endemic clades are generally contemporary with the modern high islands (Price & Clague, 2002) and are often nested within source lineages with characteristics or geographic occurrences consistent with a dispersal hypothesis (e.g., Baldwin & Wagner, 2010; Gillespie *et*

*al*., 2012). Hawaiian *Artemisia* not only fits the general pattern in terms of time of origin but also in having characteristics that are conducive to dispersal by biotic and, possibly, abiotic means.

*Bird dispersal*. Birds have been widely implicated as the most important dispersal vectors for angiosperm lineages that have diversified in the Hawaiian Islands, including lineages of Compositae (Guppy, 1903; Carlquist, 1967, 1996; Sakai *et al*., 1995; Price & Wagner, 2004; Baldwin & Wagner, 2010). Carlquist (1967) concluded that internal transport of seeds and fruits by birds was responsible for 38% to 50% of arrivals to the higher islands of Polynesia, and that attachment of seeds or fruits to bird feathers with hooks, barbs, or a viscid coating is also a significant mechanism of dispersal to Pacific islands. Sakai *et al*. (1995) judged that all founders of native Hawaiian Compositae lineages arrived by some means of dispersal by bird. Although regular bird traffic to the Hawaiian Islands is primarily along the Central Pacific flyway, most naturally visiting bird taxa that have been recorded in the islands are accidentally displaced migrants, including taxa that normally migrate between Eurasia and Australasia (see Gillespie *et al*., 2012). Notable among birds that have potential to disperse plants into the Pacific from eastern source areas along their natural migration route is the sooty shearwater, which has been mapped through electronic tracking tags and shown to travel in a figure-8 pattern around the Pacific Ocean, from Japan, Alaska, and California to Chile and New Zealand (Shaffer *et al.*, 2006). Sooty shearwaters are frequently seen close to shore in the Hawaiian Islands, feeding in groups, and, although pelagic, rarely have been sighted onshore (Birding Hawaii, 2002).

Guppy (1903) judged that *Artemisia* fruits were devoid of any adhesiveness or pappus, and so dispersal was largely attributable to the small size and light weight of the fruits, through adherence to bird feet or feathers, or through ingestion and passage through the gut (note: the indehiscent 1-seeded fruits of Compositae, not the seeds alone, are minimal dispersal units in the family). The fruits of Hawaiian *Artemisia*, however, are covered with sticky resinous glands, and may provide a potential dispersal advantage for the lineage (Degener & Degener, 1946; Wagner *et al.*, 1999). Sakai *et al*. (1995) concluded that *Artemisia* was likely dispersed to the Hawaiian Islands based on the viscid quality of its capitula and fruits. As mentioned previously, fruits of both *A. kauaiensis* and *A. chinensis* also have a distinct pappus crown of stiff, uneven teeth, and in some cases a long recurved tooth (see Figs. 2, 3), unlike all other *Artemisia* species. These pappus teeth may facilitate lodging of fruits in bird feathers. Pappus has been regarded as important in promoting fruit dispersal by birds (Carlquist, 1967), which have been observed to harbor fruits of Compositae in their feathers (Ridley, 1930).

The ability of *Artemisia* fruits to be dispersed over vast distances is supported by other findings of this study: The inferred dispersal of a member of subg. *Hawaiiana* from southeast Asia to the Hawaiian archipelago does not span as much distance as another long-distance dispersal from western North American to South America that also was resolved in this study. The South American *A. copa*, *A. sodiroi*, and *A. mendozana* evidently share a common ancestor with extant taxa of shrubby sagebrushes in subg. *Tridentatae* from North America. The molecular data show two of the South American plants, *A. mendozana* and *A. copa*, to be sister

taxa, which in turn are sister to North American taxa of subg. *Tridentatae*, sect. *Tridentatae*. The third South American taxon, *A. sodiroi*, is sister to North American *A. californica* and *A. nesotica*, and therefore belongs to subg. *Tridentatae*, sect. *Nebulosae* (Shultz, 2009) (Fig. 2). The South America taxa are native to Ecuador, Chile, Bolivia, and Argentina. The distance between the Great Basin of western North America, where members of subg. *Tridentatae* are abundant, to southern South America is ~9,500 km or nearly twice that between Taiwan, where *A. chinensis* is native, and the high islands of the Hawaii chain (~5,000 km). These relationships between North and South American *Artemisia* taxa represent the widely noted pattern of New World amphitropical disjunction, which has been primarily attributed to long-distance dispersal by birds (see Wen and Ickert-Bond, 2009).

*Wind dispersal*. Long-distance wind dispersal of *Artemisia* from a coastal habitat in southeast Asia to the Hawaiian Islands cannot be completely dismissed, although appears relatively unlikely. The fruits of Hawaiian *Artemisia* are 0.75–1 mm long and 0.5 mm wide (Degener & Degener, 1946) and are as small or smaller than seeds of *Metrosideros polymorpha* Gaud. (Myrtaceae), which has been extensively studied in relation to wind dispersal in the Hawaiian Islands (Drake, 1992). Carlquist (1967, 1974) discussed *Metrosideros* as one of the few angiosperm species that may have been able to colonise the Hawaiian Islands by wind dispersal, although he noted the potential for all of those taxa to be dispersed by birds, as well. The subtropical jet stream has been implicated in carrying small propagules (e.g., fern spores) from Asia to the Hawaiian Islands, although this mechanism requires that storm updrafts could propel the propagules into the upper atmosphere, where the jet stream prevails (see Carlquist, 1980; Gillespie *et al*., 2012). The prevailing trade winds are relatively weak, and hurricanes and cyclones in the northeast Pacific generally track from east to west, so are not favorable for wind dispersal from Asia to the Hawaiian Islands (see Gillespie *et al*., 2012).

*Oceanic dispersal by rafting*. Potential for *Artemisia* dispersal to the Hawaiian Islands by oceanic rafting needs to be considered in light of the occurrence of Asian *A. chinensis* and Hawaiian *A. australis* in littoral or near-shoreline environments. Within subg. *Hawaiiana*, the Asian *A. chinensis* has been described as highly salt-tolerant by Huang *et al.* (2007), who found that plants of that species grew well even under highly saline conditions. Guan *et al.* (2010) noted that among 32 taxa tested for salt tolerance, *A. chinensis* was even more salt tolerant than *A. vulgaris*, which was among the most salt-tolerant of those taxa tested in Hanslin & Eggen's (2005) experiments on seashore halophytes. In Hanslin & Eggen's study, *A. vulgaris* germinated at the highest salinity, 400 millimoles after 21 days of exposure, and germinated at a high rate after rinsing with distilled water. Potential for water dispersal in at least some members of *Artemisia* was noted by Cappers (1993), who reported that seeds of *A. maritima* were a common component of the litter washed ashore on north coast beaches in the Netherlands, and that they could still germinate after collection.

Carlquist (1967) discussed frequent oceanic drift and infrequent rafting as effective means of long-distance dispersal in Polynesia for salt-resistant seeds that are able to germinate after

prolonged saltwater contact, citing source and recipient ecology as an important factor: "*Hawaiian flora contains more species traceable to rafting or similar infrequent arrival via oceanic drift than do other Pacific floras*" (p. 158). A scenario might be envisioned of island hopping between small volcanic islands from southeast Asia to the Bonin Islands (1500 km eastward), where *A. chinensis* is native, to the Midway Islands (where Maxent modeling suggests presence of a suitable climate for *A. chinensis*), and down the Hawaiian chain to the modern high islands without travelling more than 800 km per dispersal interval, provided favorable currents. Extensive island-hopping by drift or rafting over such large distances appears unlikely for *Artemisia*, however, based on lack of evident adaptations for flotation or drift, absence of remnant populations in intervening areas, and endemic status of the Hawaiian taxa. *Artemisia* appears to be a better candidate for infrequent rafting, which would be more consistent with observed patterns, including divergence of the Hawaiian taxa, without recurrent gene flow from outside the archipelago (see Gillespie *et al*., 2012).

**Ecological history of Hawaiian *Artemisia***

The ecological linkage between *A. chinensis* and *A. australis* is notable. Both grow in warm littoral and near-littoral habitats on raised coral and along coastal cliffs. The Maxent model for *A. chinensis*, reinforced by the PCA and MANOVA results, indicates that some of the leeward Hawaiian Islands, at least, may have suitable climatic settings in their exclusively low elevation, littoral environments. Insofar as the ecology of *A. chinensis* can be assumed to be similar to that of the most recent common ancestor of the clade corresponding to subg. *Hawaiiana*, the Leeward Islands conceivably could have served as a staging ground for *Artemisia* during initial colonisation of the archipelago, with subsequent extinction on the small islets.

Understanding the ecological history of Hawaiian *Artemisia* requires brief consideration of the habitats and distribution of the members of subg. *Hawaiiana*. The Asian *A. chinensis* has been documented as naturally occurring from Taiwan, nearby Orchid Island, the southernmost Ryukyu Islands (just north of Taiwan), Okinawa, and the Bonin Islands, nearly 2,000 km to the east. Native status of *A. chinensis* along the coast of Guangdong, Fujian, and Zhejiang is questioned in the *Flora of China* (Ling *et al*., 2008), but climatic suitability there is predicted by the Maxent analysis. *Artemisia chinensis* is widely cultivated for medicinal and ornamental purposes in China, Japan, and the Philippines, making determination of native status difficult. The Maxent analysis predicts that climatic suitability in the Philippines is unlikely.

*Artemisia australis* is the most widespread of the Hawaiian taxa, found on all of the main islands and occupying sites along coastal cliffs from close to sea level (C. Hobbs, personal observation) to about 800 m elevation (Wagner *et al.*, 1999). The Maxent model appeared to accurately predict its occurrence on all islands according to *Manual of Flowering Plants of*

*Hawai`i*, based on 17 records of occurrence data. *A. chinensis* is limited to mostly tropical environments due at least in part to observed lack of cold-tolerance, with *A. australis* exhibiting less cold-sensitivity than *A. chinensis* and more cold-sensitivity than *A. mauiensis* (C. Hobbs, unpublished results).

Observations of living plants and locality data from herbarium specimens suggest that the Kaua`i endemic *A. kauaiensis* occurs mainly at about 1,000 m elevation, perhaps only on cliffs above Waimea Canyon, where abundant feral goats may limit its distribution. The possibility of a past distribution of *A. kauaiensis* that extended further downslope cannot be ruled out. Burney *et al*. (2001) documented pre-Polynesian plant (and animal) subfossils in lava tubes near sea level of taxa currently known only from mid to high elevations on Kaua`i, where low elevation habitats have been highly modified by human activities.

*Artemisia mauiensis* is elevationally discontinuous from the other two Hawaiian taxa, growing upslope at 1,900 to 2,300 m elevation on exposed cliffs of volcanic rock in subalpine shrubland in the crater and on south-facing slopes of Haleakala volcano on East Maui (Shultz, 1999). The highly dissected leaves of *A. mauiensis,* which are divided into thread-like segments, and its general silvery appearance are putative adaptations to intense exposure to ultraviolet light, porous depauperate soils, and extreme cold, dry conditions high on the volcano. *Artemisia mauiensis* exclusively occurs on some of the youngest features of the Hawaiian Islands outside Hawai`i Island; the Haleakala volcano has been K-Ar dated to 0.75 Ma (Naughton *et al.*, 1980).

The nested phylogenetic position of the subalpine *A. mauiensis* within subg. *Hawaiiana*, which otherwise consists of primarily low-elevation taxa, provides evidence for an unusual example of colonisation of relatively cool and dry subalpine habitats from lower habitats of tropical climate in the Hawaiian Islands. The finding here of upslope migration accounting for the habitat occurrence of *A. mauiensis* (and, to a relatively minor extent, *A. kauaiensis*) does not support the hypothesis of boreal ancestry for Hawaiian *Artemisia* suggested earlier by Engler (1879), Skottsberg (1931), and Fosberg (1948).

*Artemisia* appears to be an exception to an increasingly evident pattern of boreal ancestry of high-elevation Hawaiian endemic angiosperms, as predicted by Ballard & Sytsma (2000). They based that prediction on north-south bird migratory pathways intersecting the Hawaiian Islands and climatic matching between boreal American and high-elevation Hawaiian sites, and it has been subsequently supported by phylogenetic data from additional clades (see Baldwin & Wagner, 2010; Gillespie *et al*., 2012). In other words, although high-elevation endemism in angiosperms of the Hawaiian Islands commonly appears to be a result of colonisation of the archipelago from temperate or boreal areas of similar climate (synclimatic transitions sensu Ackerly, 2009), Hawaiian *Artemisia* appears instead to be an example of an ecological shift from tropical lowlands to temperate or boreal highlands within the archipelago (anticlimatic transitions sensu Ackerly, 2009). Such extreme anticlimatic transitions are evidently rare in the Hawaiian Islands, as predicted by Skottsberg (1931), who noted absence or near-absence of

Polynesian and Indo-Malayan (including southeast Asian) elements in the flora of the high Hawaiian volcanoes and who concluded "*(w)e cannot expect to find this thermophilous element in the highest regions*" (p. 56).

**Parallel trends in ecological evolution**

The inferred ecological expansion in Hawaiian *Artemisia* from tropical coastal habitats to subalpine settings, where *A. mauiensis* is endemic, is consistent with evidence from recent studies by Tkach *et al*. (2008a, b), who showed molecular evidence for repeated colonisation of the Arctic by 13–18 diverse lineages of *Artemisia* and concluded that members of the genus showed "*multiple and uncomplicated colonisations of the Arctic*" (Tkach *et al*., 2008a, p. 195). Modern members of *Artemisia* often occur in boreal and high-elevation sites in Asia, Europe, and North America, and high-plains habitats may represent the ancestral ecology of the genus, based in part on fossil pollen evidence. For example, the Hoh Xil Basin, one of the oldest sites in central Asia where *Artemisia* fossil pollen from the early Oligocene was reported, has an average elevation of 5,000 m and was likely of high elevation when the pollen was shed (Garzione *et al*., 2000; Liu *et al*., 2001). Some mainland taxa of the genus have sparingly colonised tropical areas, either directly to alpine areas or with subsequent migration upslope. Examples include *A. afra* Jacq. ex Willd. in South Africa (to 2400 m), and *A. copa*, *A. mendozana*, and *A. sodiroi* in Chile, Argentina, and Bolivia (3000 to 3600 m elevation). In conclusion, a worldwide perspective on the evolutionary and biogeographic history of *Artemisia* indicates that the exceptional anticlimatic ecological shift from tropical lowland to subalpine habitats in the Hawaiian setting may be an example of a clade-specific tendency for transitions to high-elevation or high-latitude environments.

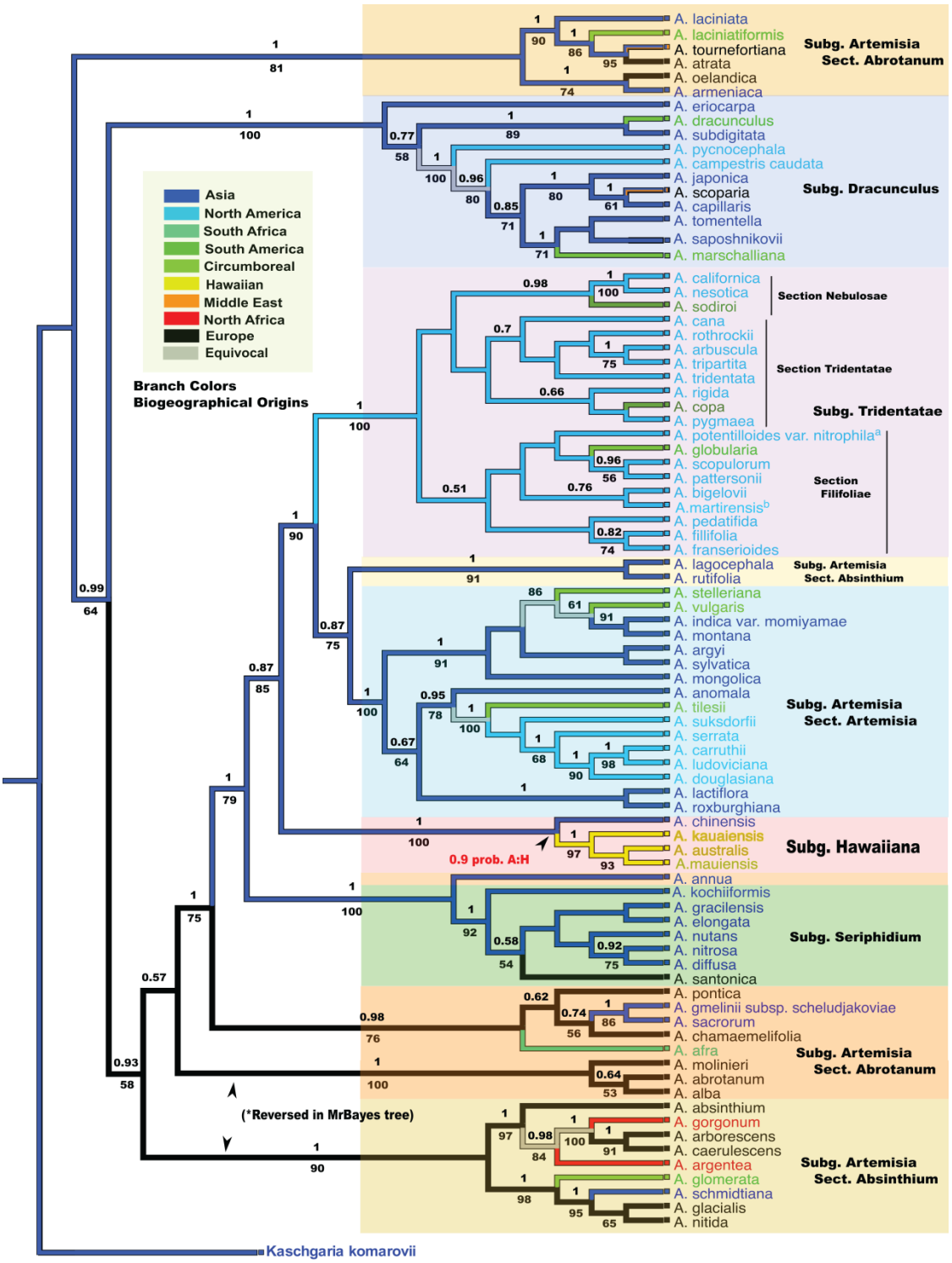



**Figure 1** Nuclear ribosomal (ITS + ETS) + chloroplast (*trn*L-F + *psb*A-*trn*H) maximum-likelihood tree for *Artemisia* (Compositae–Anthemideae), with biogeographical ranges mapped using parsimony. Branches are coloured according to the biogeographic key (bicoloured branches indicate endemism to both Asia and North America). Support values: Bayesian-inference (BI) posterior probabilities above branches, maximum-likelihood (ML) bootstrap values (>50%) below branches. Differences in topology between ML and BI trees are marked with an asterisk (*). Subgenera and sections of *Artemisia* are denoted by shaded boxes. Probability of an Asian origin of subg. *Hawaiiana* (based on LaGrange analysis) is indicated on the root node. Abbreviations: A. = *Artemisia*; sect. = section; subg. = subgenus. Name changes for two taxa were necessitated by recent reclassification of the genus *Sphaeromeria* (Garcia *et al*., 2011): **Artemisia martirensis** (Wiggins) C.R. Hobbs,

comb. nov. [basionym: *Tanacetum martirensis* Wiggins; Proc. Calif. Acad. Sci. ser. 4, 30: 254 (1965)]; **Artemisia potentilloides** A. Gray var. **nitrophila** (Cronquist) C.R. Hobbs, comb. nov. [basionym: *Tanacetum potentilloides* (A. Gray) A. Gray var. *nitrophilum* Cronquist; Leafl. W. Bot. 6: 49 (1950)].

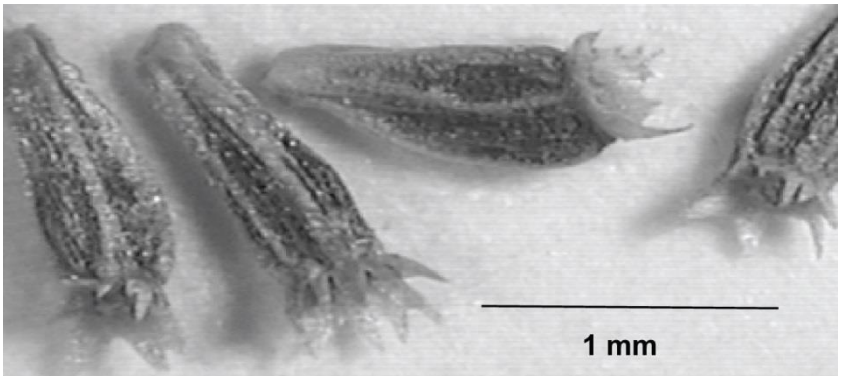



**Figure 2** *Artemisia chinensis* fruits (cypselae), showing prominently ribbed fruit walls and coroniform pappus.


**Figure 3** *Artemisia kauaiensis* fruits (cypselae), showing prominently ribbed fruit walls and coroniform pappus, with dried corolla/style (in lower image)

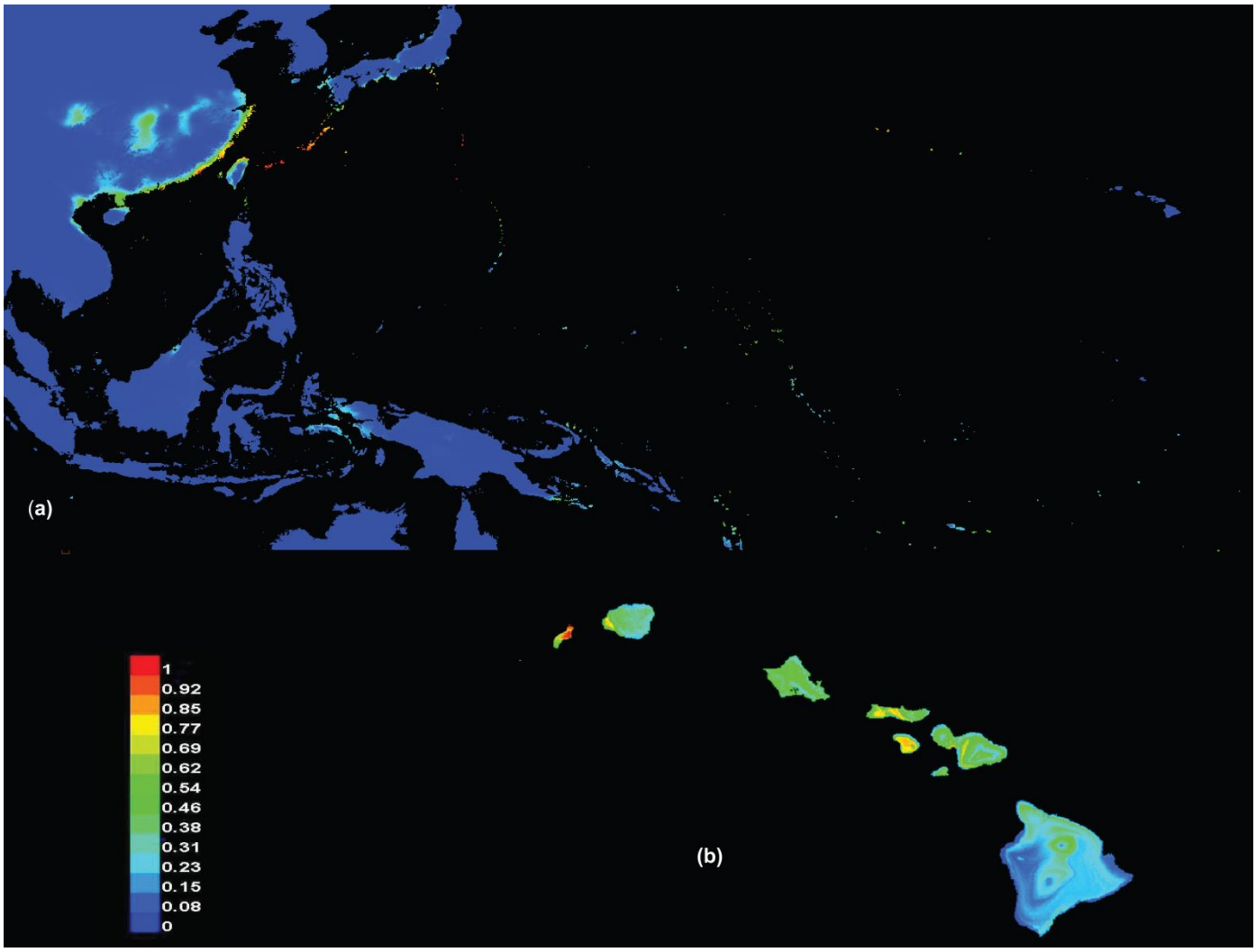



**Figure 4** Maximum entropy models of climate suitability for taxa of *Artemisia* subg. *Hawaiiana*. Map (a): Hawaiian prediction for *A. australis*. Map (b): Asian and Pacific prediction for *A. chinensis*. Map (c): Prediction for *A. chinensis* in vicinity of Hawaiian chain (enlarged from Map (b)).

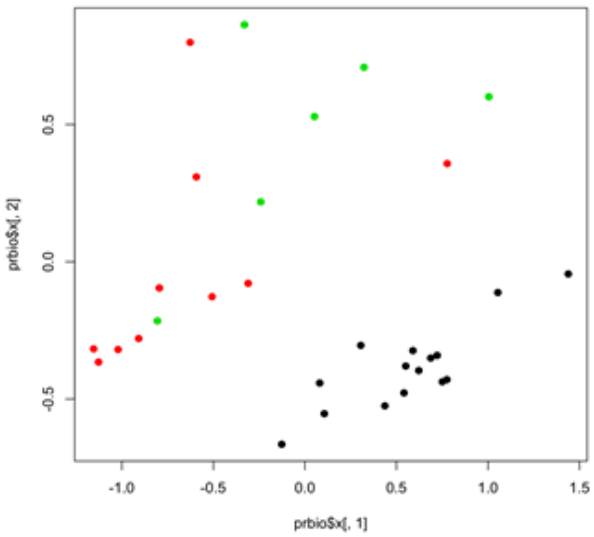
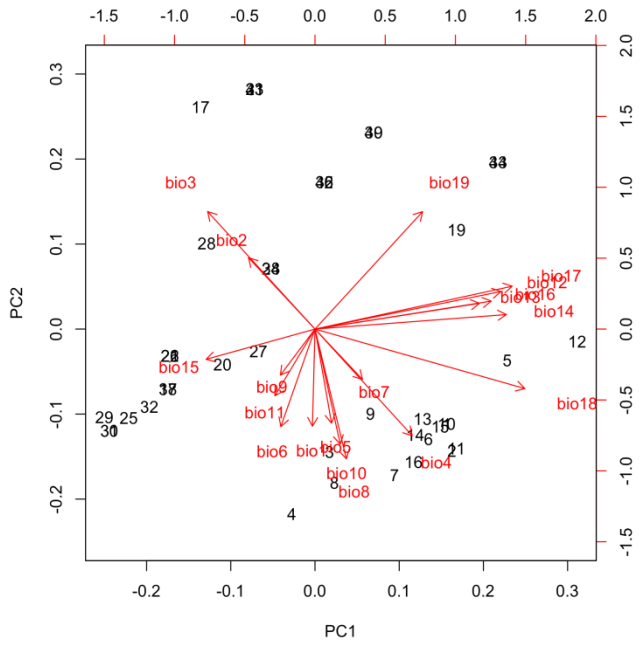



**Figure 5a** PCA analysis, plot of 19 derived climate variables sampled at specific reported locations of *A. chinensis* (black), *A. australis* (red), and *A. mauiensis* (green) showing species divergence in climate space.




BIO1 = Annual Mean Temperature
**BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)**
**BIO3 = Isothermality (BIO2/BIO7) (* 100)**
**BIO4 = Temperature Seasonality (standard deviation *100)**
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
**BIO7 = Temperature Annual Range (BIO5-BIO6)**
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month
BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
**BIO19 = Precipitation of Coldest Quarter**

**Figure 5b** PCA analysis from 5a with arrows showing the amount of variance accounted for A. chinensis-A. australis divergence in climate space; blue-highlighted text explains A. australis-A. australis divergence.

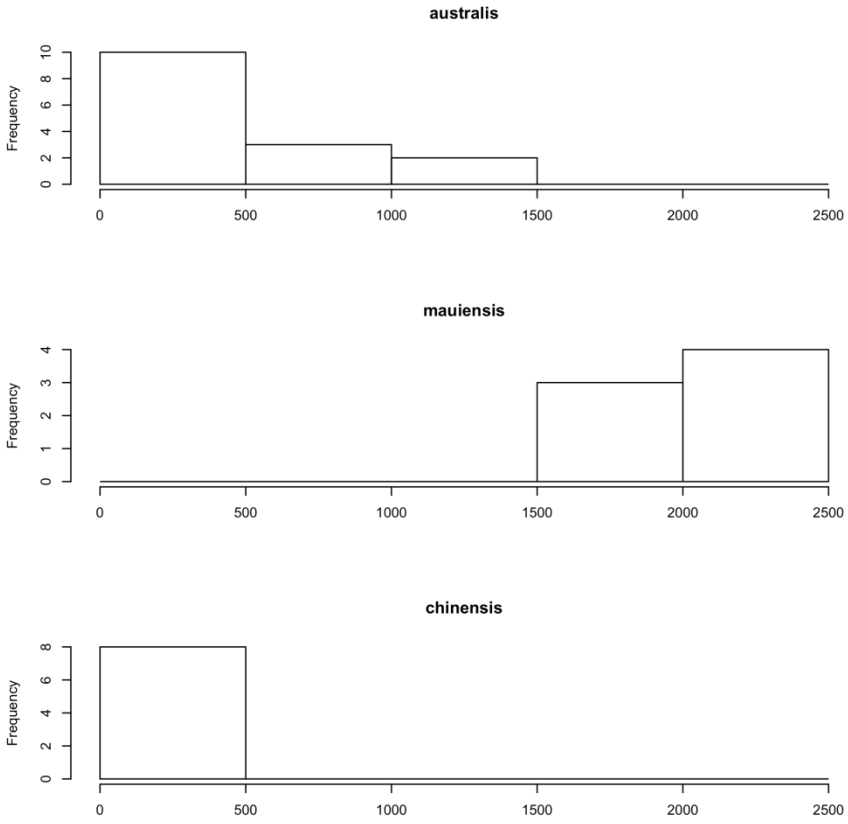



**Figure 6** Graph of Elevation of specimen datapoints

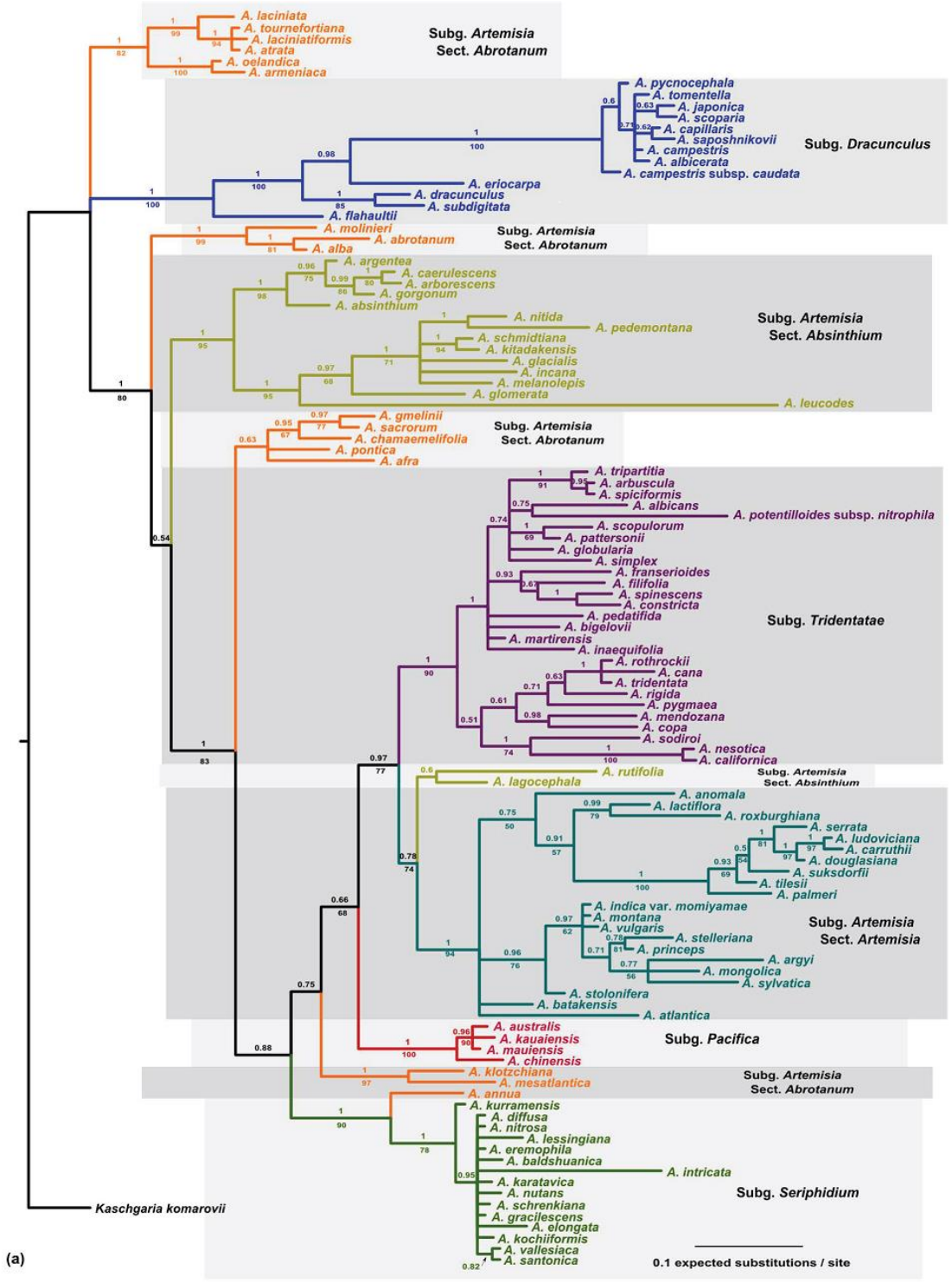



**Figure 7** *Artemisia nuclear ribosomal DNA tree*

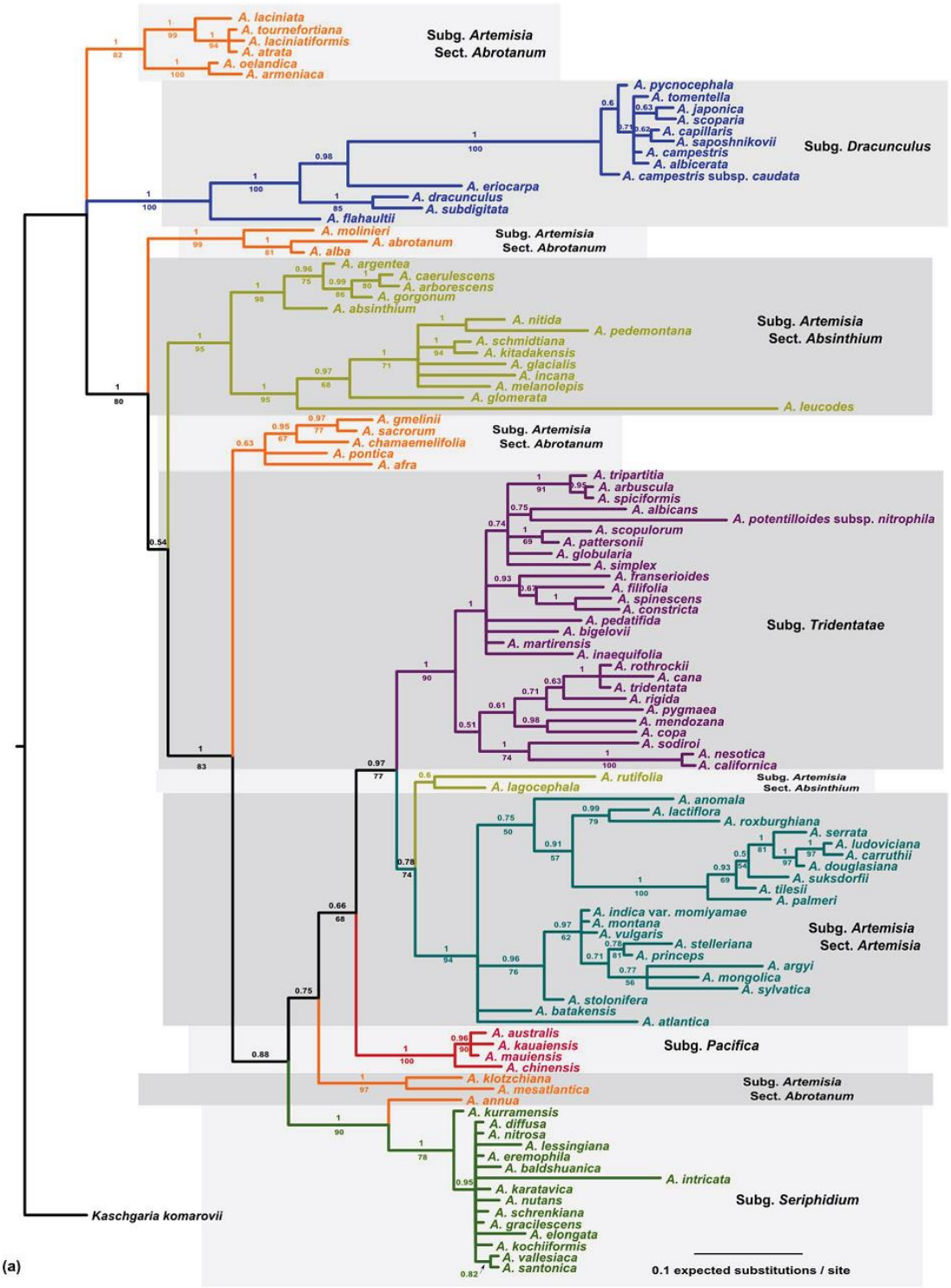



**Figure 8** Chloroplast DNA tree

Table S1 Accession numbers of samples and sequences.

| **Taxon** | Collector | Collect. # (date) | Source | **Voucher** | **Collection Locality** | **GB-ITS** | **GB-ETS** | **GB-*trn*L-F** | **GB-*psb*A** |
|---|---|---|---|---|---|---|---|---|---|
| *A. abrotanum* L. | *leg. Ign.* | *s.n.* | UCBG (49.0525) | UC | Europe | JX051694 | JX069394 | JX073778 | JX073862 |
| *A. absinthium* L. | Hobbs | H113 | Hobbs | UC | Berkeley, CA | JX051763 | JX069443 | JX073827 | JX073911 |
| *A. afra* Jacq. ex Willd. | Twibell, J. | 1992.08 | NPCA | UC | South Africa | JX051743 | JX069431 | JX073815 | JX073899 |
| *A. alba* Turra | *leg. Ign.* | *s.n.* | UCBG (98.0686) | UC | Europe | JX051695 | JX069395 | JX073779 | JX073863 |
| *A. albicans* S. Garcia, Garnatje, McArthur, Pellicer, S.C. Sand. & Vallès-Xirau | Trehm, A. | 15345 | UC | UC | Nye Co., NV | JX051705 | JX069459 | | |
| *A. albicerata* Krasch. | Michelson, A. | *s.n.* (8/25/1910) | UC | UC | Dzarkent, Kazakhstan | JX051744 | JX069467 | | |
| *A. annua* L. | Hilbig, W.; Torrell & Valles; *leg. Ign.*; Liu and Ji | 23.07.1983; BCN_12486; MPS 001275; *s.n.* | Genbank | HAL; HAL; KMH; _ | Mongolia:Bulgan Aymag; Spain; _ ; Yunnan, China | AM398847 | DQ028879 | FJ692341 | FJ418749 |
| *A. anomala* S. Moore | Twibell, J. | 1999.33 | NPCA | UC | China | JX051674 | JX069377 | JX073761 | JX073845 |
| *A. arborescens* | Aita, L., *et al* | *s.n.* (7/6/1993) | UCBG 94.1041) | UC | Varigotti, S.V. Italy | JX051741 | JX069429 | JX073813 | JX073897 |
| *A. arbuscula* Nutt. | Raiche, R. | 70182 | UCBG (87.1053) | UC | Mendocino Co., CA | JX051666 | JX069369 | JX073753 | JX073837 |
| *A. argentea* L'Hér. | Twibell, J. | 1994.9 | NPCA | UC | Madeira, African Atlantic coast | JX051696 | JX069396 | JX073780 | JX073864 |
| *A. argyi* Lév. & Van. | *leg. Ign.* | *s.n.* (11/16/2009) | UCBG (2009.0558) | UC | Shandong Prov., China | JX051681 | JX069383 | JX073767 | JX073851 |
| *A. armeniaca* Lam. | Leonova, T. | *s.n.* (8/3/1968) | UC | UC | Kazakhstan, Tselinograd region | JX051693 | JX069393 | JX073777 | JX073861 |
| *A. atlantica* Coss. & Durieu | Maire, R. | *s.n.* (12/8/1924) | UC | UC | Taouarit, Morocco | JX051745 | JX069468 | | |
| *A. atrata* Lam. | Twibell, J | 1989.12 | NPCA | UC | Italy | JX051761 | JX069442 | JX073826 | JX073910 |
| *A. australis* Less. | Hobbs | H164 | Hobbs | UC | O'ahu, HI | JX051753 | JX069435 | JX073819 | JX073903 |
| *A. baldshuanica* Krasch. & Zapr. | Filatova, N. | *s.n.* (9/30/1982) | UC | UC | Uzbekistan, Zeravshan Range | JX051658 | JX069449 | | |
| *A. batakensis* Spreng. | Tanaka, T. | *s.n.* (3/8/1931) | UC | UC | Taiwan | JX051684 | JX069453 | | |
| *A. bigelovii* A. Gray | Ackerman, T.L. | 83-783 | UC | UC | Lincoln, NV | JX051710 | JX069405 | JX073789 | JX073873 |
| *A. caerulescens* L. | *leg. Ign.* | *s.n.* (9/17/2002) | UCBG ( 2002.1033) | UC | Rovigo Province, Italy | JX051697 | JX069397 | JX073781 | JX073865 |
| *A. californica* Less. | Roderick, W. | *s.n.* | UCBG | UC | San Mateo Co., CA | JX051687 | JX069388 | JX073772 | JX073856 |
| *A. campestris* L. | Twibell, J | 1988.1 | NPCA | UC | Breckland, Norfolk, England | JX051736 | JX069426 | JX073810 | JX073894 |
| *A. campestris* subsp. *caudata* Michaux | Leonova, T. | 438 | UC | UC | Aktyubinsk region, Kazakhstan | JX051714 | JX069409 | JX073793 | JX073877 |
| *A. cana* Pursh | Edwards, S. | 89.234 | TRBG | UC | Modoc Co., CA | JX051668 | JX069371 | JX073755 | JX073839 |
| *A. capillaris* Thunb. | Hobbs | *s.n.* (9/12/2007) | UCBG (2007.0467) | UC | Ibaraki Prefecture, Honshu, Japan | JX051715 | JX069410 | JX073794 | JX073878 |
| *A. carruthii* Alph. Wood ex Carruth. | Hobbs | H119 | Hobbs | UC | Sedona, AZ | JX051722 | JX069416 | JX073800 | JX073884 |
| *A. chamaemelifolia* Vill. | Twibell, J | 1982.03 | NPCA | UC | Swiss Alps | JX051730 | JX069420 | JX073804 | JX073888 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *A. chinensis* L. | Chung, MH | 1076 | UC | UC | Fukien Prov., China | JX051756 | JX069438 | JX073822 | JX073906 |
| *A. constricta* S. Garcia, Garnatje, McArthur, Pellicer, S.C. Sand. & Vallès-Xirau | Leary *et al.* | 6195 | UNLV | UNLV | Clark Co., NV | JX051752 | JX069470 | | |
| *A. copa* Phil. | Arroyo, P.J. *et al.* | 94019 | MBG | MO | Antofagasta Prov., Chile | JX051759 | JX069440 | JX073824 | JX073908 |
| *A. diffusa* Krasch. ex Poljakov | Filatova, N. | *s.n.* (10/8/1983) | UC | UC | Uzbekistan, vicinity of Bukhara | JX051653 | JX069362 | JX073746 | JX073830 |
| *A. douglasiana* Besser | Hobbs | H261 | Hobbs | UC | San Bernardino Co., CA | JX051723 | JX069417 | JX073801 | JX073885 |
| *A. dracunculus* L. | Raiche, R. | 52.93 | UCBG | UC | San Luis Obispo Co., CA | JX051718 | JX069413 | JX073797 | JX073881 |
| *A. elongata* Filatova & Ladygina | Ladishna, J.M. | 1892 | UC | UC | Tian Shan, Kyrgyzstan | JX051746 | JX069432 | JX073816 | JX073900 |
| *A. eremophila* Krasch. ex Poljak. | Filatova, N. | *s.n.* (9/19/1982) | UC | UC | Tajikistan, Turkestan range | JX051657 | JX069448 | | |
| *A. eriocarpa* Bunge | Belianina, Proscuriakova | 7/1975 | UC | UC | Turkmenistan, Lake Mollakara | JX051742 | JX069430 | JX073814 | JX073898 |
| *A. filifolia* Torr. | Shultz, L. & McReynolds, J. | 20176 | UC | UC | San Juan Co., CA | JX051670 | JX069373 | JX073757 | JX073841 |
| *A. flahaultii* Emb. & Maire | Maire, R. | *s.n.* (6/25/1927) | UC | UC | Guelb, Morocco | JX051719 | JX069460 | | |
| *A. franserioides* Greene | Hobbs | A174 | Hobbs | UC | Telluride, CO | JX051669 | JX069372 | JX073756 | JX073840 |
| *A. glacialis* L. | Twibell, J | 1995.37 | UCBG | UC | Piedmont regions of north-west Italy | JX051700 | JX069400 | JX073784 | JX073868 |
| *A. globularia* Cham. ex Bess. | Haley, G. | *s.n.* (summer, 1929) | UC | UC | St. Paul Is., Alaska | JX051757 | JX069439 | JX073823 | JX073907 |
| *A. glomerata* Ledeb. | Wiggins, I.L. | 12668 | UC | UC | Barrow, AK | JX051738 | JX069427 | JX073811 | JX073895 |
| *A. gmelinii* subsp. *scheludjakoviae* Korobkov | Twibell, J. | 1989.18 | NPCA | UC | Russia | JX051662 | JX069365 | JX073749 | JX073833 |
| *A. gorgonum* Webb | Twibell, J. | 1995.09 | NPCA | UC | Cape Verde | JX051721 | JX069415 | JX073799 | JX073883 |
| *A. gracilescens* Krasc. & Iljin | Saphronova, I. | *s.n.* (1968) | UC | UC | Kazakhstan, near Aktobe | JX051731 | JX069421 | JX073805 | JX073889 |
| *A. inaequifolia* S. Garcia, Garnatje, McArthur, Pellicer, S.C. Sand. & Vallès-Xirau | Hobbs | H231 | UC | UC | Wasatch Co., UT | JX051676 | JX069472 | | |
| *A. incana* (L.) Druce | McNeal *et al.* | *s.n.* (9/16/1976) | UC | UC | Armenia | JX051762 | JX069455 | | |
| *A. indica* var. *momiyamae* (Kitam.) H. Hara | Twibell, J. | 2005.03 | NPCA | UC | Japan | JX051701 | JX069379 | JX073763 | JX073847 |
| *A. intricata* Franch. | Ikonnikov-Galitzky, N. and V. | *s.n.* (7/11/1912) | UC | UC | Central Mongolia | JX051659 | JX069450 | | |
| *A. japonica* Thunb. | *leg. Ign* | *s.n.* (1985) | UCBG (86.0408) | UC | Sichuan Province, China | JX051713 | JX069408 | JX073792 | JX073876 |
| *A. karatavica* Krasch. & Abolin ex Poljakov | Poljakov, P.P. | 277 | UC | UC | Kazakhstan | JX051660 | JX069451 | | |
| *A. kauaiensis* (Skottsb.) Skottsb. | Hobbs | A189 | Hobbs | UC | Kaua'i, HI | JX051754 | JX069436 | JX073820 | JX073904 |
| *A. kitadakensis* Hara & Kitam. | Furuse, M. | *s.n.* (8/19/1956) | UC | UC | Nakakoma-gun, Japan | JX051732 | JX069465 | | |
| *A. klotzschiana* Besser | Vasquez, M. | 2121 | UC | UC | Veracruz, Mexico | JX051704 | JX069458 | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *A. kochiiformis* Krasch. & Lincz. ex Poljakov | Chukavina, AP and NC Filatova | *s.n.* (10/15/1981) | UC | UC | Pripiandzhski Karatay, southern Tajikistan | JX051747 | JX069433 | JX073817 | JX073901 |
| *A. kurramensis* Qazilb. | Togasi, M | *s.n.* (11/2/1957) | UC | UC | Honshu: Kyoto, Japan | JX051751 | JX069469 | | |
| *A. laciniata* Willd. | Hiroe, M | 7453 | UCBG (2006.0529) | UC | Rishiro Island, Japan | JX051691 | JX069391 | JX073775 | JX073859 |
| *A. laciniatiformis* Kom. | Petrovsky, V and Piieva, T | *s.n.* (8/20/1985) | UC | UC | Bilibino district, Russia | JX051750 | JX069434 | JX073818 | JX073902 |
| *A. lactiflora* Wall. ex DC. | Simmons, J. | Guiz 137 | UCBG (2000.0284) | UC | Guizhou Prov., China | JX051675 | JX069378 | JX073762 | JX073846 |
| *A. lagocephala* (Besser) DC. | *leg. Ign.* | H210 | UCBG | UC | Primorsky Region, Russia | JX051673 | JX069376 | JX073760 | JX073844 |
| *A. lessingiana* Besser | Leonova, T. | 542 | UC | UC | Aktyubinsk region, Kazakhstan | JX051655 | JX069447 | | |
| *A. leucodes* Schrenk | O. Linchevski *et al.* | 3200 | UC | UC | Near Tal-Tal, Kazakhstan | JX051748 | JX069473 | | |
| *A. ludoviciana* Nutt. | Raiche, R. | 40332 | UCBG (84.0546) | UC | Russel Co., Kansas | JX051708 | JX069403 | JX073787 | JX073871 |
| *A. martirensis* (Wiggins) C.R. Hobbs syn. *Sphaeromeria martirensis* (Wiggins) A.H. Holmgren, L.M. Shultz & Lowrey | Thorne, R.F. | 57280 | UC | UC | Sierra San Pedro Martir, Baja California, Mexico | JX051734 | JX069423 | JX073807 | JX073891 |
| *A. mauiensis* (A. Gray) Skottsb. | Hobbs | A163 | Hobbs | UC | Maui, HI | JX051755 | JX069437 | JX073821 | JX073905 |
| *A. melanolepis* Boiss. | Rechinger, K.H. and F. | 6026 | UC | UC | Gorgon, Iran | JX051703 | JX069457 | | |
| *A. mendozana* DC. | Hunziker, J.H and Gamerro, J.C. | *s.n.* (3/25/1989) | MO (11626) | MO | San Juan, Argentina | JX051689 | JX069454 | | |
| *A. mesatlantica* Maire | Maire, R. | *s.n.* (8/10/1924) | UC | UC | Aghbalou Larbi, Morocco | JX051749 | JX069474 | | |
| *A. molinieri* Quezel, M. Barbero & R.J. Loisel | Twibell, J. | 1990.16 | NPCA | UC | southern France | JX051690 | JX069390 | JX073774 | JX073858 |
| *A. mongolica* (Fisch. ex Besser) Fisch. ex Nakai | McNeal *et al.* | 796 | UC | UC | Yakutsk region, U.S.S.R. | JX051683 | JX069385 | JX073769 | JX073853 |
| *A. montana* (Nakai) Pamp. | *leg. Ign.* | *s.n.* (11/02/2006) | UCBG (2006.0769) | UC | Hokkaido Prefecture, Hokkaido, Japan | JX051677 | JX069380 | JX073764 | JX073848 |
| *A. nesiotica* P.H. Raven | Smith, N. | 1.2 | UCBG (2001.0301) | UC | Ventura Co., CA | JX051686 | JX069387 | JX073771 | JX073855 |
| *A. nitida* Bertol. | Twibell, J. | 1996.77 | NPCA | UC | Italy | JX051698 | JX069398 | JX073782 | JX073866 |
| *A. nitrosa* Weber ex Stechm. | Leonova, T.A. | 520 | UC | UC | Aktyubinsk region, Kazakhstan | JX051654 | JX069363 | JX073747 | JX073831 |
| *A. nutans* Willd. | Biagi, F. & P. Castagnini | *n.s.* (10/13/2005) | UCBG | UC | Siena Prov., Italy | JX051661 | JX069445 | JX073829 | JX073913 |
| *A. oelandica* (Besser) Krasch. | Twibell, J. | 2000.73 | NPCA | UC | Denmark | JX051692 | JX069392 | JX073776 | JX073860 |
| *A. palmeri* A. Gray | *leg. Ign.* | *s.n.* (5/5/2003) | UCBG (2004.0162) | UC | San Diego Co., CA | JX051740 | JX069466 | | |
| *A. pattersonii* A. Gray | Teare, KA & Talow, DW | 1202 | UC | UC | Boulder, Colorado | JX051737 | JX069425 | JX073809 | JX073893 |
| *A. pedatifida* Nutt. | Hartman, RE & Nelson, TW | 40341 | UC | UC | Natrona Co., Wyoming | JX051671 | JX069374 | JX073758 | JX073842 |
| *A. pedemontana* Balb. | Costa, M *et al.* | 12587 | UC | UC | Teruel Prov., Spain | JX051702 | JX069456 | | |
| *A. pontica* L. | Twibell, J. | 1984.02 | NPCA | UC | Pontic Alps | JX051663 | JX069366 | JX073750 | JX073834 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *A. potentilloides* var. *nitrophila* (Cronquist) C.R. Hobbs syn. *Sphaeromeria potentilloides* var. *nitrophila* (Cronquist) A.H. Holmgren, L.M. Schultz & Lowrey | Hobbs | H232 | UC | UC | Bridgeport, CA | JX051760 | JX069441 | JX073825 | JX073909 |
| *A. princeps* Pamp. | *leg. Ign.* | 2010.0359 | UCBG (2010.0359) | UC | Tsukuba, Ibaraki Prefecture, Japan | JX051726 | JX069462 | | |
| *A. pycnocephala* (Less.) DC. | Hobbs | A228 | Hobbs | UC | Pescadero, CA | JX051711 | JX069406 | JX073790 | JX073874 |
| *A. pygmaea* A. Gray | Hobbs | H243 | Hobbs | UC | Austin, NV | JX051724 | JX069418 | JX073802 | JX073886 |
| *A. rigida* (Nutt.) A. Gray | Sondenaa, AC | 268 | UC | UC | Idaho Co., Idaho | JX051720 | JX069414 | JX073798 | JX073882 |
| *A. rothrockii* A. Gray | Hobbs | H253 | Hobbs | UC | Mammoth, CA | JX051665 | JX069368 | JX073752 | JX073836 |
| *A. roxburghiana* Wall. ex Besser | Twibell, J. | 2007.37 | NPCA | UC | Himalaya | JX051682 | JX069384 | JX073768 | JX073852 |
| *A. rutifolia* Steph. ex Spreng. | Filatova, N.C. | *s.n.* (9/12/1982) | UC | UC | Fergana, Kyrgyzstan | JX051672 | JX069375 | JX073759 | JX073843 |
| *A. sacrorum* Ledeb. | Lowdermilk, W.C. | *s.n.* (7/12/1924) | UC | UC | Shansi Prov., China | JX051733 | JX069422 | JX073806 | JX073890 |
| *A. santonica* L. | Leonova, T. | 6 | UC | UC | Saratov region, Russia | JX051656 | JX069364 | JX073748 | JX073832 |
| *A. saposhnikovii* Krasch. ex Poljakov | Lei SM, *et al*. | *s.n.* (7/29/1979) | UC | UC | Kyrgyz Republic | JX051716 | JX069411 | JX073795 | JX073879 |
| *A. schmidtiana* Maxim. | Twibell, J. | 1997.86 | NPCA | UC | Japan | JX051699 | JX069399 | JX073783 | JX073867 |
| *A. schrenkiana* Ledeb. | Leonova, T.G. | *s.n.* (7/8/1968) | UC | UC | Kazakhstan, Semipalatinsk region | JX051728 | JX069464 | | |
| *A. scoparia* Waldst. & Kitam. | Twibell, J. | 1998.113 | NPCA | UC | China | JX051717 | JX069412 | JX073796 | JX073880 |
| *A. scopulorum* A. Gray | Vanderhorst, JP | 1092 | UC | UC | Garfield Co., Colorado | JX051709 | JX069404 | JX073788 | JX073872 |
| *A. serrata* Nutt. | Shinners, LH & Catenhusen, J. | 2869 | UC | UC | Douglas Co., Wisonsin | JX051706 | JX069401 | JX073785 | JX073869 |
| *A. simplex* (A. Nelson) S. Garcia, Garnatje, McArthur, Pellicer, S.C. Sand. & Vallès-Xirau | Dorn, R. | 6976 | RM | RM | Natrona Co., Wyoming | JX051758 | JX069471 | | |
| *A. sodiroi Hieron.* | Balls, EK | B7078 | UC | UC | Schellin (near Cuenca), Equador | JX051667 | JX069370 | JX073754 | JX073838 |
| *A. spiciformis* Osterhout (*A. rothrockii* in gard.) | Roderick, W. | *s.n.* (9/13/1969) | TRBG | UC | Mono Co., CA | JX051725 | JX069461 | | |
| *A. spinescens* D.C. Eat. syn. *Picrothamnus desertorum* Nutt. | Pinzl, A. | 12811 | UC | UC | Nye Co., NV | JX051727 | JX069463 | | |
| *A. stelleriana* Besser | leg. Ign. | s.n. (2001-2002) | UCBG (2002.0626) | UC | Primorsky Region, Russia | JX051679 | JX069382 | JX073766 | JX073850 |
| *A. stolonifera* (Maxim.) Kom. | Twibell, J. | 1992.07 | NPCA | UC | China | JX051680 | JX069452 | | |
| *A. subdigitata* Mattf. | Sino-Amer. Exped. | 59 | UC | UC | Western Hubei Provence, China | JX051739 | JX069428 | JX073812 | JX073896 |
| *A. suksdorfii* Piper | Edwards, S. | 95.109 | TRBG | UC | Mendocino Co. | JX051707 | JX069402 | JX073786 | JX073870 |
| *A. sylvatica* Maxim. | Twibell, J. | 2007.18 | NPCA | UC | China | JX051685 | JX069386 | JX073770 | JX073854 |
| *A. tilesii* Ledeb. | Calder, J.A. | 6200 | UC | UC | Kenai Peninsula, Alaska | JX051735 | JX069424 | JX073808 | JX073892 |
| *A. tomentella* Trautv. | Leonova, T.G. | 467 | UC | UC | Kazakhstan, Karagandinskaya region | JX051712 | JX069407 | JX073791 | JX073875 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *A. tournefortiana* Reichb. | St. Lager, L. | *s.n.*, 9/29/1920 | UC | UC | Rhône, France (naturalized) | JX051729 | JX069419 | JX073803 | JX073887 |
| *A. tridentata* Nutt. | Hogan, S.B. | *s.n.*, 5/28/1993 | UCBG (UCBG 94.129) | UC | Lake Co., OR | JX051688 | JX069389 | JX073773 | JX073857 |
| *A. tripartita* Rydb. | Hobbs | H238 | UC | UC | Langell Valley, OR | JX051664 | JX069367 | JX073751 | JX073835 |
| *A. vallesiaca* All. | Twibell, J. | 1988.12 | UC | UC | Swiss Alps | JX051652 | JX069446 | | |
| *A. vulgaris L.* | *leg. ign* | *s.n.* | UCBG (UCBG 74.0211) | UC | Glantal, Austria | JX051678 | JX069381 | JX073765 | JX073849 |
| *Kaschgaria komarovii* (Krasch. & Rubtz.) P. Pol. | Bremer, K. | 3123 | UPSV | UPSV | Amanat., Kazakhstan | JX051764 | JX069444 | JX073828 | JX073912 |

## Chapter 2
## Phylochemistry
## The Utility of Secondary Metabolites in Phylogenetics

Despite the success and popularity of DNA for phylogenetic research, secondary metabolites (SM), also called secondary compounds, continue to be studied for their utility for solving taxonomic problems. Use of SM in systematics, by tradition called chemosystematics, has been extensive for more than a half century and predates macromolecular approaches. However, very few studies of SM characters using current phylogenetic methods have been published to date, so little is known about the relative phylogenetic utility of the individual groups of compounds from different biosynthetic pathways that occur throughout the tree of life.

The number of studies in the literature with the search terms "chemotaxonomy" or "chemosystematics" since 1950 is over 22,000 (Fig. 1). Since 1980, the number of studies that focus on chemotaxonomy or chemosystematics with phylogenetics as key words has gone from 68 in 1980 to over 500 in 2012, and has grown by 40% in the last 4 years (Fig. 2). Chemosystematic studies have been published on many groups of organisms, such as algae (Johns *et al.*, 1979), sponges (Erpenbeck & van Soest, 2006), fungi (Stadler *et al.*, 2010; Vynne *et al.*, 2011), lichens (Hawksworth, 1976; Printzen, 2010), plants (Hegnauer, 1962; Grayer *et al.*, 1999; Wink *et al.*, 2010), and especially bacteria (Schleifer, 2009). However, with a few exceptions, summarized below, a closer look at these studies shows that they mention the utility of chemical characters in phylogenetic analyses in an observational way, or in support of phylogenetic hypotheses based on *ad hoc*, non-statistical methods, or in some cases, distance measures.

At least 50,000 different SM have been characterized in nature: terpenes (30,000), fatty acids (300, with thousands estimated), and alkaloids and phenyl propanoids or phenolic compounds (3,000 characterized). Phenylpropanoid derivatives include the flavonoids, coumarins, anthocyanins, and alkaloids, among others (Southon and Buckingham, 1989; Connolly & Hill, 1991; Metcalf & Metcalf, 1992; Van de Loo *et al.*, 1993).

However, among that wide array of compounds, there are very few, if any, that have been successfully utilized in phylogenetic analysis, either for determining species trees by themselves, or when combined with gene data. I argue that one reason is the lack of analysis across the spectrum of SM groups available to determine the most appropriate compounds based on the particular organisms under study. Chemical groups generally have been chosen because of ease of obtaining the data and characterizing individual compounds (volatile terpenes), or based on previous attempts or ideas about the utility of the compounds in general (flavonoids, alkaloids).

In this study I question whether defense or signaling compounds such as alkaloids, flavonoids, and terpenes are the most appropriate types of compounds for chemosystematic studies, based on their presumed direct relationship with many environmental factors, such as predation, pollination, climate changes, radiation levels, available moisture, and other influences. On the other hand, compounds in organisms that are primarily involved with structure and nutrition, such as cell wall polysaccharides or fatty acids, may be less tied to

environmental influences and more conserved with less day-to-day fluctuation, making them more suitable for phylogenetic applications.

Compounds involved in ecological interactions, such as defense and signaling, e.g. pollinator attraction (Wink, 2003), appear to be mostly incongruent with phylogeny, and likely more tied to recent evolutionary and ecological process, based on a diversity of studies (e.g., Williams and Whitten 1999; Levin 2003; Agrawal *et al*. 2009; Armbruster *et al*. 2009; Harris 2009; Wink *et al*.2003, 2010), but very few studies involving phylogenies have utilized chemical data from compounds thought to be related to structural or nutritive functions in plants.

Wink *et al*. (2010), for example, published phylogenetic trees with plant orders (following APG-III, 2009) as terminal taxa, with 3 classes of alkaloids, cardiac glycosides, and glucosinolates mapped onto the branches (Wink *et al*., 2010). This work showed that these compounds, thought to be mostly utilized for defense, have arisen multiple times and for the most part appear to have little phylogenetic signal at this level, based on the lack of clade specificity of these compounds visually, not from formal phylogenetic signal tests. At the genus level in the Fabaceae, Wink *et al*. (2003, 2010) found only scattered clade specificity for protease inhibitors, cyanogenic glycosides, flavonoids, triterpenes, cardiac glycosides, iridoid glycosides, and various classes of alkaloids. Non-protein amino acids, which are thought to be used for nitrogen storage as well as defense, appear to be more tribe-specific in the Fabaceae, consistent with the idea that compounds related to nutrition or structure are more conserved.

As I will show, an individualized approach to selection of SM for phylogenetic or ecological studies in specific groups of organisms can provide insights into the relationship of chemical traits to phylogeny, maximizing their usefulness for phylogenetic inference. Methods such as coefficient of variance, comparative analysis for detection of phylogenetic signal, trait evolution estimation, phylogenetic inference with chemical datasets alone or combined with DNA datasets, and tree congruence tests are all performed to provide insights into the nature and utility of chemical characters. The importance of methodology for coding SM characters, and continuous characters in general, for phylogenetic analysis is also explored.

Attention to methodological considerations can maximize the utility of SM for use in phylogenetic analysis with specific groups of organisms. The identity and levels of chemical compounds can vary widely between different plant parts, so it is essential to perform analyses on the same part of the plant, collected in the same season, preferably close to the same time, and growing in representative conditions. Collection of samples from plants that are free from signs of fungal, viral, or bacterial infection or predation from insects and browsers can help reduce chemical variability (Hasegawa and Shirato, 1959). Extraction and analytical methods should also be consistent. Inducing methodological variability in the production of chemical datasets can add considerable noise and significantly reduce resolution.

In this study I utilized a dataset of 24 fatty acids (Mongrand *et al*., 2001) from 137 species of gymnosperms, representative of all families, in comparison with a reasonably well-resolved gene tree (*rbc*L) to test the idea that this structural and nutritive group of compounds are relatively conserved and less environmentally labile, i.e., less phenotypically plastic and

less homoplastic, and therefore more useful for phylogenetic analysis than other SM. Acyl lipids, including fatty acids, perform a number of important structural and metabolic functions in plants, animals, and microbes, notably for membrane fluidity, signaling (Bonaventure *et al.*, 2003), and energy storage in the form of triacylglycerol (TAG) in seeds, leaves, and other parts (Lin & Oliver, 2008), and are produced by a single group of coordinated enzymes in the plastid (Ohlrogge, 1997). Fatty acids can be quickly and accurately identified with fatty acid methyl ester analysis (FAME) using gas chromatography, and with high-resolution chromatographic methods such as Wall Coated Open Tubular (WCOT) columns (Eder, 1995; Ackman, 2009). With these methods, resolution is increased to the point where numerous minor compounds can be distinguished, adding to the number of characters for phylogenetic analysis. Rapid sample preparation and small polymer microextraction (SPME) can speed acquisition of data by up to 6 times (Lu & Harrington, 2010).

## MATERIALS AND METHODS

*Evaluation, Selection and Preliminary Analyses of Chemical Datasets*

The chemosystematics literature was searched to identify suitable chemical datasets from representative groups of secondary constituents. One hundred and seventy studies on plant groups reporting chemical data from 17 constituent groups were reviewed. Finally, 29 datasets from 11 groups (Table 1) were chosen for further analysis, based on the ease of extraction to a usable digital format, quality of the data, and number of taxa and chemical characters. Larger datasets, considering both number of taxa and characters, can be associated with higher accuracy and reliability of phylogenetic analyses and reduced noise, so dataset size was therefore a major selection criterion (Heath *et al.*, 2008). More specifically, datasets with more than 20 taxa have been shown to have a high occurrence of measurable signal, while datasets with ≥ 25 characters have reduced noise associated with poor analytical methodology and incomplete sampling (Freckleton *et al.*, 2002; Blomberg *et al.*, 2003). The mean taxon number and number of variable characters for the 15 "structure-nutrition" datasets was 35.57/16.0, and for the 14 "defense compounds" datasets was 25.86/66.21.

The coefficient of variance (CV) for each chemical character in relation to the mean of the samples (defined as the ratio of the standard deviation to the mean) for all taxa was measured using PC-ORD (McCune & Mefford, 2011) to test the lability of SM from different biosynthetic pathways (Table 1). The number of species (rows) represents the number of samples of each SM variable, and the CV is the % total variance.

As a framework in which to study the measurable phylogenetic signal in the SM datasets identified from the initial screening processes, DNA sequence alignments from GenBank were constructed when sufficient sequences were available, corresponding to a representative sampling of the taxa in each dataset. Thirteen SM datasets were found to have useable molecular datasets, and tree-building under parsimony and maximum likelihood criteria, as well as comparative analyses were performed for each (Stevens *et al.*, 1996; Maffei *et al.*, 1997; Löffler *et al.*, 1997; Rezanka, 1998; Umek *et al.*, 1999; Rønsted *et al.*,

2000; Mongrand *et al*., 2001; Wollenweber *et al*., 2003; Dunstan *et al.*, 2005; Fritsch & Keusgen, 2006; Dev *et al.*, 2010; Valant-Vetschera *et al.*, 2010; Hsieh & Harris, 2012).

Finally, one dataset of leaf fatty acids in the gymnosperms was chosen for intensive study (Mongrand *et al.*, 2001), based on the large size of the dataset (137 taxa and 24 SC traits) and availability of DNA sequences for most of the taxa from GenBank, which allowed for the production of well-resolved, albeit not completely resolved, trees for comparative analyses.

*Normalization and Discretization of the Chemical Data*

Chemical data are quantitative, measured and characterized through analytical methods such as gas chromatography/mass spectroscopy (GC/MS) and liquid chromatography/MS (LC/MS), and consist of continuous data. The Willi Hennig Society edition of TNT (Goloboff *et al.*, 2006, 2008) was used to analyze the original continuous data under parsimony, incorporating algorithms that assign a range to each terminal that varies from the mean plus or minus 1-2 standard error (SE). TNT allows states between 0-65 with 3 decimal places, and uses Farris' algorithms (Farris, 1970) for the down-pass optimization.

The original and log-transformed continuous data (relative %) were compared with discretized data using comparative methods to test for phylogenetic signal, and analyzed with TNT (maximum parsimony) and RAxML for phylogenetic inference. Data were discretized either by dividing the continuous range into 2 (presence-absence), 3 (0 = <2, 1 = 2.1-5, 2 = 5.1-30.0, 3 = > 30.0), 4, 9, or 10 states, or according to Thiele's gap-coding method. Discretization was performed in Mesquite version 2.75 (Maddison & Maddison, 2011), or gap-weighted by the method of Thiele (1993) using MorphoCode (Schols *et al.*, 2004). In some datasets where individual chemical compounds had large variance from the mean, log-transformation (x+1) was performed in PC-ORD before analysis.

*Phylogenetic Analysis of the Molecular and Chemical Data: Acquisition, Alignment, and Phylogenetic Tree Reconstruction*

DNA sequences of the taxa in the leaf fatty acid (FA) dataset were retrieved from GenBank. The highest number of sequences available for alignment by far in this group of taxa was for the large subunit of ribulose-1,5 bisphosphate carboxylase (*rbc*L). Sequences of this chloroplast encoded gene were available for 107 of the 137 taxa of the 14 families from the Mongrand *et al.* (2001) FA data set. GenBank accession numbers are given in Table S1 in the supplement. The alignment (1434 bp) was performed in Seaview ver. 4 (Gouy *et al.*, 2010) with MUSCLE (Edgar, 2004) and adjusted by hand (Simmons, 2004).

Sequences of *rbc*L are widely available in the literature for many taxa of seed plants because it was one of the first genes to receive wide attention as a useful region for plant phylogenetic inference in the 1980s (Palmer *et al*., 1988). Hollingsworth *et al.* (2009) determined that *rbc*L is highly universal and has good discriminating power. Although cpDNA markers by themselves have limitations, such as the inability to detect a history of hybridization, the lack of taxonomic coverage for gymnosperms using other molecular datasets outweighed those considerations.

Phylogenetic tree estimation using maximum likelihood (ML) and maximum

parsimony (MP) was performed separately for *rbc*L and the FA datasets with the original continuous log-transformed data, as well as with the data discretized as described above. Analyses were also done with *rbc*L and the discretized FA data combined.

The starting tree for MP searches was produced by stepwise addition. The heuristic search used TBR branch swapping with 1000 iterations of parsimony ratchet (Nixon, 1999). Tree clade-specificity (TCS) for all FA trees was measured by counting the distribution of terminal taxa from 10 genera (*Abies*, *Picea*, *Pinus*, *Juniperus*, *Larix*, *Cupressus*, *Ephedra*, *Taxus*/*Cephalotaxus*, *Araucaria*, and *Thuja*), as well as all taxa in the Zamiaceae for each analysis in clades with a separate most recent common ancestor (MRCA), compared with the *rbc*L tree as a reference. Chemical trees with a mean TCS of 1 had the same clade specificity of taxa as the *rbc*L tree, and so demonstrated a high level of congruence between the chemical and molecular trees. Higher numbers are consistent with less clade specificity.

The ML tree searches were performed using RAxML-VI-HPC, ver. 7.2.6 (Stamatakis, 2006; Stamatakis *et al.*, 2008) via the RAxML GUI 1.3 user interface (Silvestro & Michalak, 2012) on the desktop, or within CIPRES (Cyberinfrastructure for Phylogenetic Research) Portal, version 2.2 (Miller *et al*., 2009, 2010). The best-scoring ML trees with clade support values were obtained from 10 independent runs, with 500 LSR (lazy subtree rearrangement) bootstrap replicates each. RAxML chose model parameters for one or two partitions (when SC and molecular data were combined), including alpha, invar, and rates, with an accuracy of 0.001 log likelihood units for each of the 10 independent runs, and then the best-fitting parameters were chosen for final optimization at the end of the run. In all runs, the GTR+I+G model was used for the DNA partition and gamma-shaped rate variation or MK1 models were used for the fatty acid data (Lewis, 2001).

In addition to the gymnosperm FA and *rbc*L data, a corresponding morphological dataset with 123 categorical characters, coded 0-8 (Hart, 1987), was used to compare the phylogenetic signal and mode of evolution with the chemical and *rbc*L characters, and to perform additional separate and combined analyses incorporating the FA, morphological, and *rbc*L data for phylogenetic tree reconstruction. The combined morphological, FA, and nucleotide dataset had to be reduced to 39 taxa, representative of genera in the Pinaceae, Cupressaceae, Podocarpaceae, Araucariaceae, and Taxaceae, because of limited correspondence between the taxa included in the morphological dataset and the taxa in the FA-DNA dataset.

The advisability of combining morphological and genetic data, with possibly different evolutionary histories, in phylogenetic analyses is still debated (Bull *et al*., 1993; Scotland *et al.*, 2003; Wiens, 2004; Mishler, 2005; Wortley & Scotland, 2006; Garcia-Cruz & Sosa, 2006). Some feel they should not be combined, primarily because of potential incongruence between molecular and morphological data (Doyle, 1992; Graham *et al.*, 1998; Lecointre & Delaporte, 2004; Assis, 2009), and secondarily because of phenotypic variation in morphological data due to environmental and other influences (Caumul & Polly, 2005; Roncal *et al.*, 2012), as well as convergence (Wiens *et al.*, 2003). Others argue that any individual data set, whether morphological or genetic, might conflict with another, and that only "total evidence" analyses using all available data sets can resolve these conflicts (Kluge, 1989; Mishler 2000, 2005). In any case, studies that combine genetic, morphological, and SM data are rarely reported in the literature (Heethoff *et al.*, 2011).

Although some authors have recommended against combining different data when incongruence is detected (Lamboy, 1994; Hedges & Maxson, 1996; Givnish & Sytsma, 1997), the sole use of genetic data, to the exclusion of morphological and chemical data, is not a solution to incongruent data. Research has shown that phylogenetic trees from different genes are often highly discordant (Degnan & Rosenberg, 2009) and lack of a clear topology in multi-gene analyses, due to incomplete lineage sorting, is widespread (Syring *et al.*, 2007; Carstens & Knowles, 2007). Mishler (2000) proposed that combining several types of data could "cancel out" anomalous signals, including reticulation and incomplete lineage sorting, and bring to light new clades with high support. Consilience is the concept that independent evidence from unrelated sources adds strength to a hypothesis when the individual sources are less certain individually (Kluge, 1989). The results from the combined *rbc*L, morphological, and FA analysis here are consistent with this idea (Fig. 5).

*Measurement of phylogenetic signal*

Comparative analyses of the FA data were performed in R (R Core Team, 2012) to test for phylogenetic signal in all chemical characters. The following functions were used: phylo.d in Caper, version 0.5 (Fritz & Purvis, 2010) for binary chemical characters; fitDiscrete in Geiger version 1.99-3 (Pagel, 1999; Harmon *et al.*, 2008) or phylogenetic least squares (pgls) and a likelihood ratio test with the pgls function in caper (Pagel, 1994; Freckleton *et al.*, 2002) for multistate characters; and fitContinuous in Geiger (Harmon *et al.*, 2008), phylogenetic generalized least squares (pgls) in Caper (Freckleton *et al.*, 2002), Phylosig from the package phytools, version 0.2-30 (Revell, 2012)**,** and Blomberg's K statistic (Blomberg *et al.*, 2003) with Kcalc in Picante, version 1.6-0 (Kembel *et al.*, 2013) for continuous data. RStudio (2012), ver. 97.248 was the platform used for all analyses. The phylogenetic tree and matrix of FA trait values (original continuous data) were loaded for use in all analyses performed in the packages fitContinuous, pgls, and Phylosig. The discretized FA data (presence-absence) were also loaded for testing phylogenetic signal in the package phylo.d, and the discretized chemical and morphological data were coded as multistate characters for fitDiscrete in Geiger. The data and tree were prepared for these analyses by matching the tree tip taxon names to the taxon names in the data matrix in the same order, dropping any spaces or anomalous taxa that were not identical in both places. A comparative data object, needed for some of the analyses, was created with the comparative.data function in Caper.

*Model-testing with FitContinuous*

The best-fitting model of evolution for individual characters in the continuous FA data was found with parametric comparative methods by fitting 8 different likelihood models with the fitContinuous function of Geiger (Harmon *et al.*, 2008). The models, based on the Brownian motion (BM) model of evolution, include the lambda, delta, and kappa models, which change the branch lengths of the phylogeny in various ways, to test for the relatedness of traits to the phylogeny, for the rate of evolution, and whether the characters are evolving more under punctual equilibrium or gradualism (kappa model). The Ornstein-Uhlenbeck (OU) model is a random walk with a central tendency that approximates how stabilizing

selection and random genetic drift act on a specific phenotypic character (Lande, 1976; Felsenstein, 1988). White is a non-phylogenetic model that corresponds to no BM-like evolution (no signal). Both of these models measure the amount of evolutionary constraint. Smith *et al*. (2011) proposed that a high likelihood of this model for a character could mean very strongly constrained evolution is acting (Smith *et al*., 2011). Geiger also estimated alpha (selection-strength parameter) during model testing. Low values of α imply no selection and random drift (α =0), while higher values imply stronger selection pressure (Butler & King, 2004). The parameters that provide the best fit to the data were estimated by ML in Geiger, and the Akaike weights of all FA characters were calculated for each model using fitContinuous and Excel (Microsoft, 2010).

*Tree-Congruence Test*

Chemical FA trees were compared to corresponding *rbc*L trees from the same type of analysis (MP, ML, BI) using two different measures of congruence (the maximum agreement subtree--MAST index (de Vienne *et al*., 2007), and Kendall's W statistic of congruence with permutation using a Mantel test) to get a p value (Paradis, 2012; Paradis *et al*., 2004).

Phylogenetic trees developed with the ML and MP criteria, were input into Icong (de Vienne *et al*., 2007), which accepts web-based input of two Newick trees, to test the hypothesis that two trees are significantly more congruent than by chance. Icong performed three steps on the phylogenetic trees: 1) the topological congruence was tested with the maximum agreement subtree test (MAST) with COMPONENT v. 2.0 (Page, 1993), which measures the minimum number of tips that need to be removed in each phylogeny for the trees to be identical, 2) a large number of random trees were generated and the congruences between them measured, and 3) the values of the pair of trees being tested were compared to the random pairs. When less than 5% of the random pairs are more congruent than the two tested trees, the trees are considered more congruent than by chance ($p = <0.05$).

CADM.global in the R package ape (Paradis, 2012; Paradis *et al*., 2004; Campbell *et al*., 2011) was also used to test congruency by computing the distances between pairs of tips in two phylogenetic trees using branch lengths after creating a distance matrix. The amount of congruence between trees was determined using Kendall's W statistic (Kendall *et al*., 1939; Legendre, 2010), and then a generalized Mantel test of matrix correspondence (Mantel, 1967; Mantel & Valand, 1970) was performed to test the null hypothesis of complete incongruence between the trees being tested, generating a p-value.

## RESULTS

*Data Selection and Preliminary Analyses, Coefficient of Variance*

Considered as a group, the 15 datasets of secondary compounds of proposed nutritive compounds such as FA, amino acids, or compounds primarily related to structural functions, polysaccharides, xyloglucans (cell wall breakdown products), sterols, or waxes had a trait relative variability of 17.0. The group of 14 datasets of defense and signaling compounds expected to be evolutionarily and environmentally labile (terpenes and phenolics; specifically, monoterpenes, sesquiterpenes, iridoids, flavonoids, and anthocyanins) had a trait variability of 51.75. The “defense-signaling” compound trait variance is thus considerably (300%) higher than for the “structure-nutrition” data (Table 1).

*Normalization and Discretization of the Fatty Acid Data*

The "clade specificity" tests (CST, Table 2) and phylogenetic inference (Fig. 3) with the original 107 leaf fatty acid gymnosperm dataset of Mongrand *et al*. (2001) showed that discretized data out-performed the original continuous data (Table 2). Figure 3 is the phylogenetic tree inferred under the parsimony criterion, which graphically shows a high clade-specificity for the leaf fatty acid data alone. Genera within the Pinaceae as well as the majority of the Cycadales, and all taxa in the genus *Ephedra*, have taxa that are highly clade-specific. The data discretized into 4 states (0-3) produced trees with the highest clade specificity for most taxa of gymnosperms, and especially *Ephedra*, *Taxus-Cephalotaxus*, and the Cycadales, with a mean clade specificity (CS) of 4.27 (where a clade in the fatty acid tree with a CS of 1 is perfectly congruent with the *rbc*L tree). The continuous log-transformed data did not perform well, with a mean clade specificity of 5.64, and the difference between this method and 0-3 discretization of the data (CST = 0.427) is significant (p=0.0198). The continuous data have few identifiable gaps, and so performed poorly. The presence-absence (binary), 0-4, and 0-9 treatments all performed worse than 0-3 coding, but the difference was not significant (p = 0.134). The discretization of the data into 5 (0-4), or 10 categories (0-9) captured less signal, possibly by artificially imposing more gaps than occurred naturally in the data. Other methods of analysis, such as comparative methods, supported the 0-3 coding method as having the highest amount of phylogenetic signal, as detailed below.

*Phylogenetic Analysis of the Molecular, Fatty Acid, and Morphological Data: Acquisition, Alignment, and Phylogenetic Tree Reconstruction*

Compared with the *rbc*L tree, the tree from combined *rbc*L and FA data (tanglegram, Fig. 4) had higher support in some clades, especially in the Gnetales and Cycadales, and lower support in the Pinaceae. Increased congruence with recent phylogenies was also noted, particularly in the Cycadales (Chaw *et al*., 2005; Zgurski *et al.,* 2008), and in some clades in the Pinales (Ran *et al*., 2010; Lin *et al*., 2010) and Cupressaceae (Gadek *et al*. 2000; Wang *et al*., 2003; Yang *et al*., 2012; Mao *et al*., 2012).

Figure 5 shows a comparison of the ML trees from the combined and reduced data (39 taxa) available for FA (Mongrand *et al.*, 2001) and morphology (Hart, 1987). These trees show some significant changes in clade structure and support in the combined *rbc*L + morphology + FA trees vs. the *rbc*L tree alone for several clades. Results from phylogenetic reconstructions with MP are similar (data not shown).

Homoplasy was estimated in the MP trees (Table 3), comparing the *rbc*L tree with the tree inferred from the leaf FA data. The RI values show that increased homoplasy is evident in the FA data (0.832 vs. 0.603). These data show that the combined analysis with the *rbc*L and leaf FA data have comparable amounts of homoplasy to the *rbc*L data alone (0.809 vs. 0.832) in the 107 taxon alignment. According to these measures, combining the morphological data with the molecular data decreases the amount of homoplasy, and the morphological data alone have a level of homoplasy nearly identical to that of the *rbc*L data (0.743 vs. 0.747). The *rbc*L and leaf FA data combined have a level of homoplasy calculated by this method that is nearly identical to that of the *rbc*L and morphological data combined (0.721 vs. 0.717).

*Measurement of Phylogenetic Signal*

Significant phylogenetic signal could be detected in the gymnosperm FA data using two different measures with comparative analysis in R: lambda, as estimated by ML with phylogenetic generalized least squares (pgls) (Fig. 6), and Blomberg's K (Fig. 7). Table 4 shows lambda values and p values for both measures. Lambda values (from pgls) showed the presence of significant phylogenetic signal in 16/24 leaf FA characters, whereas K values were significant in 15/24 characters. This result can be interpreted as the evolution of the leaf FA being more tied to phylogeny and long-term selection pressures than to immediate environmental changes. However, when the 24 FA characters were mapped onto the phylogeny using parsimony, some traits were highly clade-specific. For instance, C18:4, coniferonic acid, occurs primarily in the Pinaceae, in the gnetophyte *Gnetum*, and rarely in Araucariaceae and Podocarpaceae, but not at all in the Cupressaceae or cycads (Fig. 8a). The unidentified very long-chain FA (UFA) is scattered throughout the phylogeny (Fig. 8b).

Measurement of phylogenetic signal for the morphological data of Hart (1987) showed that out of 123 morphological characters, 72 characters had a high correspondence to the *rbc*L phylogeny (lambda = 1) and 34 characters had very little or no signal (p = >0.05), with the remaining 16 characters with moderate signal >0.1 to 0.9 (Fig. 9). Phylogenetic signal in Hart's (1987) morphological dataset, as estimated here in comparison with *rbc*L and FA data, showed that vegetative characters had less signal overall than reproductive characters (62% to 72% measured with lambda; Fig. 9). Of 20 wood anatomy characters, 15 (75%), and 8/16 (50%) of the leaf characters had measurable signal, while of 16 ovulate strobilus characters, 14 (88%) had medium to high signal. See Hart (1987) for a detailed description of the morphological characters.

*Model-testing with FitContinuous*

Calculation of the likelihoods of the fit of 8 different evolutionary models to the 24 leaf FA characters show that lambda fit 10 characters best (8 are at or near 100% for lambda being most likely of all the models), and the other 14 were mixed—with varying likelihoods of 2-4 possible best model fits (Table 5). For these 14 characters, the Ornstein-Uhlenbeck model was most likely for 4 characters, the delta model for 2 characters, the white noise model for 5 characters, and no model was clearly the best for 3 characters. The lambda function collapses internal branches, leaving tip branches unchanged, which may be interpreted to imply that phylogenetic signal is lost mainly in the deeper nodes, while phylogenetic signal in recent history is still present.

*Tree-Congruence Test*

Results of the tree-congruence tests showed that the tree from the discretized 107 taxon leaf FA data (coded 0-3) was highly congruent with the *rbc*L tree (Table 6) under MP and ML, with an Icong Index of 1.32 (p = 0.001) under MP, whereas the tree derived from the gap-coded data was significantly less congruent (Icong, 1.25, p = 0.0046), and the binary FA data (1.18, p = 0.021) produced trees that were not more congruent than by chance. The MP tree of the FA data used for this test can be seen in Fig. 3.

The morphological dataset of Hart (1987) was highly congruent with the *rbc*L tree

(Icong index, 1.64, p = 2.43e-05) under ML, and less so with the FA trees because of the smaller dataset (1.20, p=0.046), as detailed in Table 6. However, the combined morphological and FA tree is highly congruent with the *rbc*L tree (Icong index, 2.19, p = 1.80e-09). Similar results can be seen between the trees under MP.

## DISCUSSION

Here I coin the term "phylochemistry," which is chemosystematics, i.e., the use of secondary metabolites (SM), applied to phylogenetic reconstruction. Analyses performed in this study have showed that the successful use of chemical data in phylogenetic reconstruction is dependent on the selection of the best-suited class of constituents for the group of organisms under study, as well as careful attention to SM character coding; that is, an individualized approach is warranted.

In this study, when multiple SM data sets were compared, it was found that data sets consisting of structural and nutritive compounds had much less variance and were more phylogenetically conserved than were datasets consisting of defensive and signaling compounds, which appear to have little association with phylogeny.

Although compounds associated with nutritional and structural functions, such as fatty acids (FAs), have been widely utilized for taxonomic studies in simple organisms--particularly bacteria, fungi, and lichens, few studies are available in more complex organisms such as gymnosperms and angiosperms, and particularly using recent phylogenetic procedures.

The uncombined FA data were used directly in phylogenetic analyses with standard methods, and the trees these data produced were remarkably concordant with the DNA tree as measured for example by the Clade Specificity Test, and the FA tree was statistically more concordant with the *rbc*L tree than by chance. When combined with *rbc*L data, the FA data increased branch support and resolution in some unresolved clades, particularly the gnetophytes and cycads, consistent with recent multigene analyses, but resulted in reduced branch support and resolution in the Pinaceae, perhaps due to reduced congruence of the *rbc*L and FA data as a result of hybridization affecting the former (Gernandt *et al.*, 2009).

The morphological dataset of representative gymnosperms in comparison with leaf FA data and cpDNA sequences showed that the leaf FA and morphological characters had significant phylogenetic signal, as measured by Pagel's lambda and Blomberg's K. Evolutionary modeling showed the FA characters were mostly highly constrained and did not follow a strict Brownian motion process. FA data combined with morphological data, and in turn with *rbc*L sequences typically showed an additive effect in support and resolution of clades.

Overall, among the 123 morphological characters, vegetative characters had less phylogenetic signal than reproductive characters. Other authors have found a low level of phylogenetic signal in plant vegetative characters. Zheng *et al.* (2009) determined that of 13 morphological characters studied in the genus *Manglietia*, 3 characters related to photosynthesis, 3 to leaf morphology, 4 to growth, and 3 to thermal tolerance, only two showed measurable phylogenetic signal, both being photosynthetic characters. Roncal *et al.* (2012) found that of the 17 vegetative traits in the genus *Geonoma* (Arecaceae) studied, 4 had phylogenetic signal (24%).

*Gymnosperm Phylogeny*

The positions of the gnetophytes, cycads, and ginkgos in the seed plant phylogeny have been difficult to settle (Matthews, 2009; Wu *et al*., 2013). In one of the largest recent studies, in terms of number of genes, Chumley *et al*. (2008) reported on an 83-gene plastid MP analysis that resolved a clade of *Ginkgo* + cycads, with ML and BI analyses instead supporting cycads branching off before ginkgos. Wu *et al.* (2013) consistently resolved a clade of *Ginkgo* and the cycads, as sister groups, based on amino acid sequences.

The ML and BI analyses of Chumley *et al*. (2008) also found support for the so-called gnepine hypothesis, i.e., that gnetophytes are sister to all conifers except Pinaceae. Zhong *et al.'s* (2011) recent study using an amino acid substitution matrix is in agreement with the gnepine hypothesis. Findings from the present study in combined analyses with *rbc*L + leaf FA data support *Ginkgo* and the cycads as a clade (ML) with 100% branch support), and are consistent with the gnepine hypothesis, albeit with only moderate support. The phylogenetic distribution of Δ-5 unsaturated FA, specifically, taxoleic acid in *Abies* and *Ephedra*, pinolenic acid in the Pinaceae and *Gnetum*, and coniferonic acid in Pinaceae and Gnetidae, as well as Araucariaceae and Podocarpaceae, is consistent with the gnepine hypothesis. These FA traits are among the ones with the highest measured phylogenetic signal based on lambda, pgls, and Blomberg's K (Fig. 6, 7, Table 3).

Comparing the *rbc*L tree with the *rbc*L + morphology + FA combined tree (Fig. 5), the most notable differences in the latter that lead to increased congruence with the multigene phylogeny of Yang *et al*. (2012), based on LFY, NLY, *mat*K, and *rbs*3 and focused primarily on Cupressaceae, are outlined below.

a) *Juniperus* and *Cupressus* constitute a clade to the exclusion of *Fokienia* (Ran *et al*., 2010; Mao *et al.,* 2012; Yang *et al.,* 2012), and this relationship is not resolved by *rbc*L alone;

b) *Platycladus* and *Microbiota* have a sister group relationship (Gadek *et al.,* 2000; Yang *et al*., 2012; Mao *et al.,*2012), and this result has significantly higher support in combined trees;

c) *Sequoiadendron* and *Sequoia* constitute a clade (Ran *et al*., 2010; Mao *et al*., 2012; Yang *et al*., 2012), and this result has higher support in the combined trees;

d) *Torreya* is sister to *Taxus*, and this pair is sister to *Cephalotaxus* (Ran *et al.,* 2010; but not Hao *et al.,* 2008), which is highly supported in combined trees, but not supported in the *rbc*L tree;

e) The placement of *Cedrus* in the Pinaceae has been controversial (Lin *et al.*, 2010). Neither the *rbc*L or the combined *rbc*L + morphology tree resolves relationships of *Cedrus* with other members of Pinaceae with significant support (>70 MBS). The *rbc*L + morphology + FA tree includes a clade with *Cedrus* sister to *Abies*, *Tsuga*, and *Pseudolarix* with a branch support of 72 MBS, and the latter 3 taxa with 81 MBS. Other research has *Cedrus* in a clade with *Abies* (Lin *et al.*, 2010; Mao *et al.*, 2012), with *Pinus* + *Picea* (Ran *et al.*, 2010), or basal to the rest of Pinaceae (Eckert *et al.*, 2006).

Extensive hybridization in Pinaceae (Kormutak *et al.,* 2008; Klaehn *et al.,* 1962) might explain increased incongruence between the FA and *rbc*L datasets, with less hybridization in wild populations detected in cycads (Chamberlin, 1926), but not *Ephedra* (Won & Renner, 2003; Ickert-Bond, 2004).

*Working with Continuous Chemical Data*

Chemical data are typically measured as percent of the total area under the curves from all chromatographic peaks, and thus are quantitative data. Values for individual compounds are the fraction under the curve of this compound compared with the total area. The use of continuous characters in phylogenetic analyses has been widely discussed, and often criticized for various reasons, such as by Farris (1990), Thiele (1993), Goloboff *et al.* (2006), and Cohen (2012). When the use of inherently continuous data is not workable, or considered undesirable because phylogenetic methods of ancestral reconstruction such as MP, ML, and BI require categorical data (or quantitative data with gaps), it is coded using a variety of often *ad hoc* methods that have also been frequently debated (Gift & Stevens, 1997; Swiderski *et al.*, 1998; Garcia-Cruz & Sosa, 2006; Farris, 2007). Most of the analyzed phenotypic datasets obtained from the literature in this present study are continuous data (20/28), including the Mongrand *et al*. (2001) leaf FA data used for intensive analysis, and so testing different coding methods is recommended. The treatment of leaf FA data as continuous characters resulted in the lowest phylogenetic signal and in the least useful addition to the *rbc*L data in combined phylogenetic analyses, likely because of the natural variation of chemical traits and lack of clear gaps in that variation. Continuous data are fine for comparative analysis, because the trait data can be compared with the branch lengths of a phylogenetic tree, but not for building phylogenetic trees in the first place.

The variety of tests used to examine the performance of various character coding methods for the FA data suggested that one method was best for phylogenetic purposes. Coding the continuous Mongrand *et al.* (2001) FA data into 4 states (0-3) carried the most phylogenetic signal for all analyses (see Tables 2, 5, and 9) based on the Clade Specificity Test, tree congruency tests, and phylogenetic inference using MP and ML analyses. This could be because coding the data as discrete characters can reduce the effect of natural variation in the amounts of FA in different populations, and reduce inherent noise (Kitching *et al.*, 1998; Wiens, 2001; Adams *et al.*, 2004) from a possible lack of precision in analytical methods attributable to slight variation in instrumentation, methodology, measurement, and extraction of the plant material. For example, in the 0-3 coding scheme used in this study, 0 is any value below 2.0, so the impact of instrumentation or methodology that didn't detect very small amounts of a particular FA is lessened. Based on results obtained here, choice of a coding method for continuous chemical data to use in phylogenetic analyses should be preceded by coding of the data using a variety of methods, testing the coded data for phylogenetic signal, for example, with fitDiscrete in the Geiger package of R (Harmon *et al.*, 2008), and comparing results of phylogenetic analyses using congruence tests and separate and combined datasets for the same set of taxa. Coding the traits with finer divisions, 0-4, 0-9, or gap-coding of Thiele (1993), did not perform well in most cases compared with 0-3 coding, likely because those codings imposing artificial or environmentally induced variation that did not correspond to genetically-controlled variation. The desirability of coding tests prior to utilizing chemical or morphological data in phylogenetic analyses has been previously noted (Rae, 1998; Laurin & Germain, 2011).

As with morphological data, SM data are available in published studies as presence-absence characters (binary data), but my results indicate that some evolutionary signal may be

lost in such datasets, possibly introducing noise and reducing resolution during phylogenetic estimation (Adams & Rosenberg, 1998; Adams *et al.*, 2004). As noted above, binary data performed poorly in all tests.

The independence of chemical characters has been questioned by Barkman (2001), who recommended coding biosynthetic pathways in a step matrix. However, that method had mixed success in phylogenetic reconstruction. Coding biosynthetic pathways has several distinct disadvantages. The major problem is that one enzyme can code for many different products; this is not uncommon in plants (Chen *et al.*, 2004, Roeder *et al.*, 2007). Steele *et al.* (1998) determined that in *Abies grandis*, one enzyme, δ-selinene synthase, can produce up to 34 different products, and γ-humulene synthase up to 51 products. Another problem arises because one compound, such as the common monoterpene 1,8-cineol can be produced by more than one pathway (Wise *et al.*, 1998). Since terpene synthases making different products from the same species have been shown to be more similar to each other than synthases making the same product from different species, convergent evolution is likely to be common (Chen *et al.*, 2004; Sharkey et al, 2005).

*Environmentally Labile and Highly Conserved Traits*

The results of tests for coefficient of variance (Table 1) showed that traits from datasets of SM that are likely to be produced by organisms to serve structural, nutritional, or metabolic functions, such as FAs, have a lower mean variance than traits thought to be used for defense or other ecological interactions like flavonoids, terpenes, and alkaloids, among others.

Results of comparative analyses that tested leaf FAs in gymnosperms for modes of evolution (Table 4), and phylogenetic signal tests (Figs. 6, 7), along with phylogenetic reconstruction (Fig. 4,5), are also consistent with this idea.

Nutritional and structural compounds such as FA and cell wall components are expected to be more conserved, in part based on the results of this study, and therefore useful for increasing resolution and support for phylogenetic studies at deeper levels, while chemicals that are labile to variable environmental conditions, produced on demand after predation, infection, and exposure to varying weather conditions, for example, are better suited to ecological studies or possibly fine-scale phylogenetic analyses. These are compounds like flavonoids, anthocyanidins, and alkaloidal or cyanogenic toxins.

Because of the demonstrated lability of SM used for defense**,** these traits may be suited for understanding rapidly evolving lineages undergoing adaptive radiations. Phenotypic characters from morphology and SM may provide phylogenetically informative characters in rapidly evolving lineages because it is likely that divergence of molecular characters in general lags behind divergence of phenotypic characters in such groups (Flagel *et al*., 2008).

FA datasets, the class of chemicals chosen for this study (Mongrand *et al*., 2001), are widely available in the phytochemical literature and are known to be essential for energy storage, plant growth, and seed development, and so are likely to be highly conserved (Bonaventure *et al.*, 2003). Lin and Oliver (2008) showed that fatty acid concentrations in

leaves were mostly consistent in composition and concentrations in 2 day-old, 5 month-old, and, to a lesser degree, senescing leaves. Extensive study of FA in all gymnosperm groups (Wolff *et al.*, 1996, 1997a, b, 1998, 2001) showed little variation due to environmental conditions (Wolff *et al*., 2001), consistent with the idea that chemicals associated with nutrition and structure are less environmentally labile than chemicals thought to be utilized for protection such as phenolic compounds and terpenes. Nasri *et al.* (2005) found that the seed FA content of 17 samples of *Pinus pinea* collected from different geographic locations in the Mediterranean had little variation. The coefficient of variation of the FA characters of Nasri *et al*. (2005) is 9.15 between the samples of each character, below the mean CV for all structure-nutrition compounds measured in this study, and well below the 55.7 mean measured for putative defense compounds. Analyses of individual taxa in *Picea* from different geographic regions showed that intraspecific variation in FA from *P. abies* was considerably lower than interspecific variation between other species in the genus (Wolff *et al.*, 2001).

*Models of Chemical Evolution*

None of the leaf FA characters analyzed in this study fits an evolutionary model of Brownian motion (BM) *per se*, which is not surprising since this is an idealized model of random trait evolution without regard to the selection pressures that bear on an organism. The Lambda model of trait evolution is the most likely fit for 12 out of the 24 FA characters in the FA data (Table 4). When lambda is highly correlated with organismal trait evolution across the phylogeny, it is widely held that trait evolution is concordant with the phylogeny, implying conservation of a trait that is under genetic control. Model testing using AIC Weights (Wagenmakers & Farrell, 2004) shows that all of the FA traits are a good fit to models that indicate strong to moderate evolutionary constraint, including the 12 that are the most likely fit to lambda (Table 4). When the measured value of lambda is greater than 0 and less than 1, traits have less covariance than expected by BM (Pagel, 1997, 1999). As can be seen in Fig. 6, 11 of the FA traits have lambda < 0.5, which means that these traits are showing the influence of increasing constraint as lambda approaches 0.

Alpha measurements for the 24 leaf fatty acid traits show that the 8 traits (bolded in Table 7) with the lowest alpha are all a highly likely fit to the lambda model (Table 4), which indicates lower selection pressure than the remainder. Alpha is a parameter that indicates the amount of stabilizing selection measured. Where $\alpha = 0$, selection pressure is not detected and trait evolution is closer to BM (drift), values >0 but <1 indicate weak-moderate pressure, >1 and <2 indicate moderate pressure, and >2 indicate strong selection pressure (Butler & King, 2004). Seven traits are a good fit to the White model of evolution and can be interpreted as having no concordance with the gene tree or BM due to significant constraint. Smith *et al.* (2011) noted that the White model describes characters that are drawn from the same normal distribution, independent of their evolutionary relationships, suggestive of very strongly constrained evolution acting (in other words, a model with a central tendency, due to evolutionary selective pressures with an infinitely strong constraint parameter, alpha). Note that these 7 FA characters all have high alpha values (Table 5), with all of the traits with the highest likelihood of a fit to the White model (bolded in Table 4) having very high values of alpha, except for juniperonic acid (20:4). Apparently, most of the leaf FA studied here, all of

which are common in gymnosperms, have been under moderate to strong selection pressures over evolutionary time and are likely highly conserved.

The results of this phylochemical study from a number of different lines of evidence show the value of SM for phylogenetic analysis is high when an individualized approach is used to select the most suitable chemical data for the taxonomic group under study. Tests of several coding methods for continuous chemical data should be performed *a priori*. The use of structural and nutritive constituents such as FA are likely to have more value than defense and signaling compounds.

Chapter 2 Illustrations

## Figures and Tables

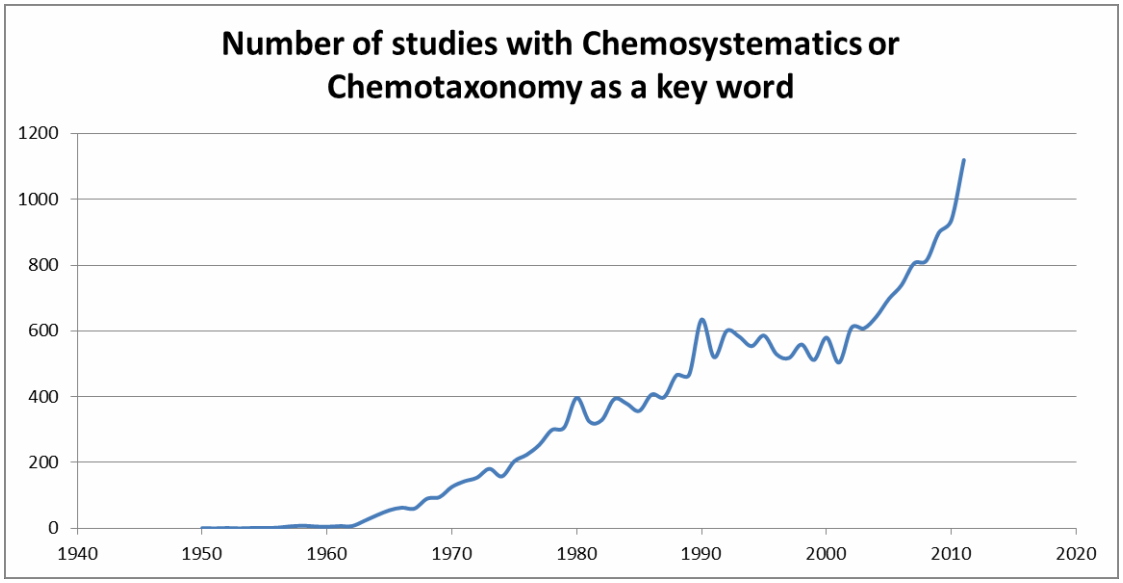



FIGURE 1. Searches of Google Scholar from 1950 to 2012.

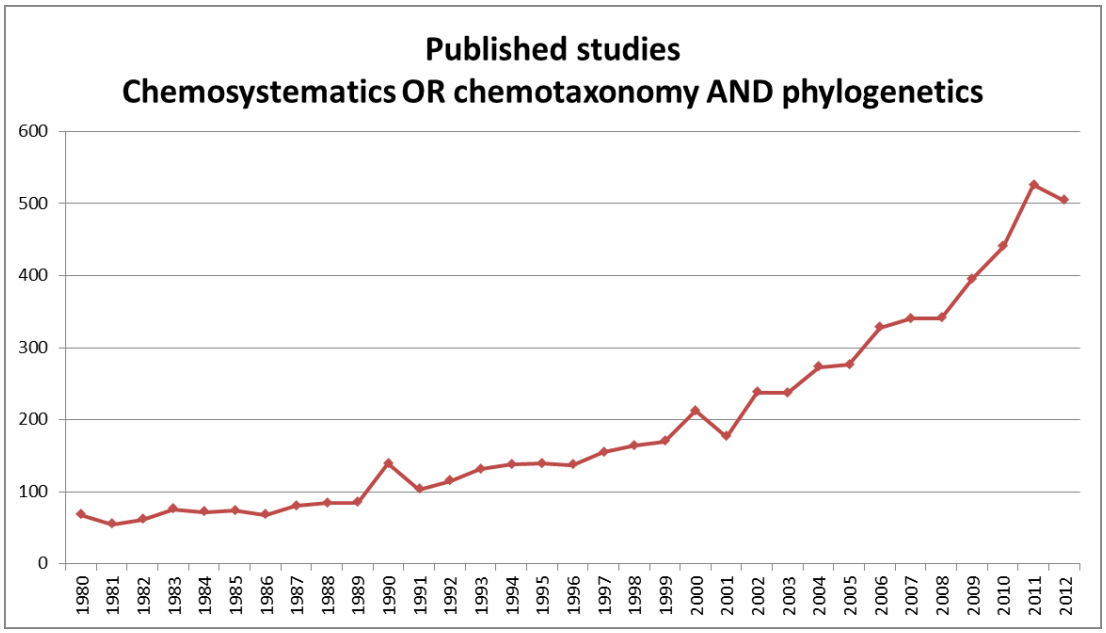



FIGURE 2. Searches of Google Scholar from 1980 to 2012—Chemosystematics.

TABLE 1. Summary of Chemical Datasets Chosen for Analysis, and Coefficients of Variance

| **Structure-Nutrition** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Chemical Group | Taxon | No. Taxa, Traits | Character CV | Outliers (No./ highest) | GB DNA alignment | Data Type | Reference |
| Waxes (from FA) | Cactaceae | 13/19 | 18.68 | 3/2.8 | *rbcL*, resolved | Continuous | Maffei et al, 1997 |
| Terpene, sterol | *Codonopsis* | 28/9 | 51.83 | 1/3.9 | 12 taxa, ITS, resolved | categorical | Wang et al, 1995 |
| Xyloglucans | Gymnosperms, monilophytes | 28/27 | 0.36 | 2/3.1 | *rbcL*, resolved | continuous | Hsieh-Harris & Harris, 2012 |
| Polysaccharides | Vochysiaceae | 16/8 | 5.8 | 2/2.7 | 5 taxa, *nadH*, resolved | Continuous | Mayworm & Salatino, 2000 |
| Fatty Acids | Rhodophylaceae | 7/22 | 3.11 | 1/2.6 | 18S, resolved | continuous | Dunstan et al, 2005 |
| Fatty acids | Brassicaceae | 43/11 | 25.8 | 6/3.0 | ITS, resolved | Continuous | Goffman et al, 1999 |
| Fatty acids | Nicotaneae | 14/17 | 0.89 | 1/2.4 | ITS, resolved; 8 taxa | Continuous | Maestri & Guzmán, 1995 |
| Fatty acids | Orobanchaceae | 21/11 | 0.2 | 0 | ITS, 20 taxa, resolved | Continuous | Velasco et al, 2000 |
| Fatty acids | Boraginaceae | 39/9 | 3.27 | 3/2.6 | ITS, 24 taxa, resolved | Continuous | Velasco & Goffman, 1999 |
| Fatty acids | *Equisetum* | 10/26 | 0.94 | 2/3.4 | *rbcL*, resolved | Continuous | Řezanka, 1998 |
| Fatty acids | Gymnosperms | 107/24 | 0.73 | 3/5.3 | *rbcL*, resolved | Continuous | Mongrand et al, 2001 |
| Fatty Acids | *Pinus* | 1/16 | 9.15 | 1/3.4 | Insufficient sequences | Continuous | Nasri et al, 2005 |
| Amino acids | Caesalpineae | 63/16 | 102.9 | 1/2.4 | Insufficient sequences | categorical | Evans & Bell, 1978 |
| Amino Acids | *Acacia* | 66/19 | 30.83 | 3/3.0 | Insufficient sequences | categorical | Evans & Bell, 1977 |
| Amino acids | *Allium* | 43/4 | 0.51 | 2/2.3 | ITS, resolved | Continuous | Fritsch & Keusgen, 2006 |
| **Mean** | | 33/14 | **17.0** | | | | |
| | | | | | | | |
| **Defense Compounds** | | | | | | | |
| Terpene, volatile | *Zingiber* | 15/67 | 42.91 | 0 | ITS, 6 taxa, poor | categorical | Jiang et al, 2006 |
| Terpene, volatile | Apiaceae | 9/168 | 8.2 | 0 | *rps*16, ITS, fair | continuous | Dev et al, 2010 |
| Terpene, volatile | *Lomatium* | 7/240 | 17.86 | 0 | *rps*16, ITS, fair | continuous | Beauchamp et al, 2005 |
| Terpene, volatile | *Stachys* | 9/191 | 17.44 | 6/2.7 | *trnLF*, 5 taxa, resolved | continuous | Skaltsa et al, 2003 |
| Phenolics | *Cuscuta* | 9/10 | 0.12 | 1/2.47 | ITS, resolved | continuous | Löffler et al, 1997 |
| phenolics, salicin | *Salix* | 20/13 | 57.64 | 2/2.3 | *rbcL* or ITS, fair | continuous | Julkunen-Tiitto, 1989 |
| Terpene, iridoids | *Plantago* | 32/39 | 54.53 | 1/2.0 | ITS, resolved | binary | Rønsted et al, 2000 |
| Terpene, iridoids | *Galium* | 31/23 | 33.65 | 1/2.5 | *rbcL*, 7 taxa, resolved | categorical | Mitova et al, 2001 |
| Phenolic, flavonoids | *Nothofagus* | 11/47 | 81.2 | 2/2.7 | ITS+*rbcL* resolved | binary | Wollenweber et al, 2003 |
| Phenolic, flavonoids | *Sedum* | 97/11 | 324 | 4/8.0; 4.2; 2.2 | *matK*, resolved | continuous | Stevens et al, 1996 |
| Terpene, volatile | *Pinus* | 5/50 | 2.63 | 0 | *matK*+*rbcL*, not resolved | continuous | Roussis et al, 1995 |
| Phenolic, flavonoids | *Hypericum* | 6/9 | 63.55 | 1/2.3 | ITS resolved | continuous | Umek et al, 1999 |
| Phenolic, flavonoids | *Dionysia* | 35(29)/ 43 | 47.84 | 6/2.5 | ITS, resolved | categorical | Valant-Veschera et al, 2010 |
| Phenolic, Anthocyanins | *Rosa* | 74/16 | 28.05 | 7/4.3 | ITS, 18 taxa, fair | categorical | Mikanagi et al, 1995 |
| **Mean** | | 26/65 | **55.7** | | *Phenolic CV=125.15 | | |

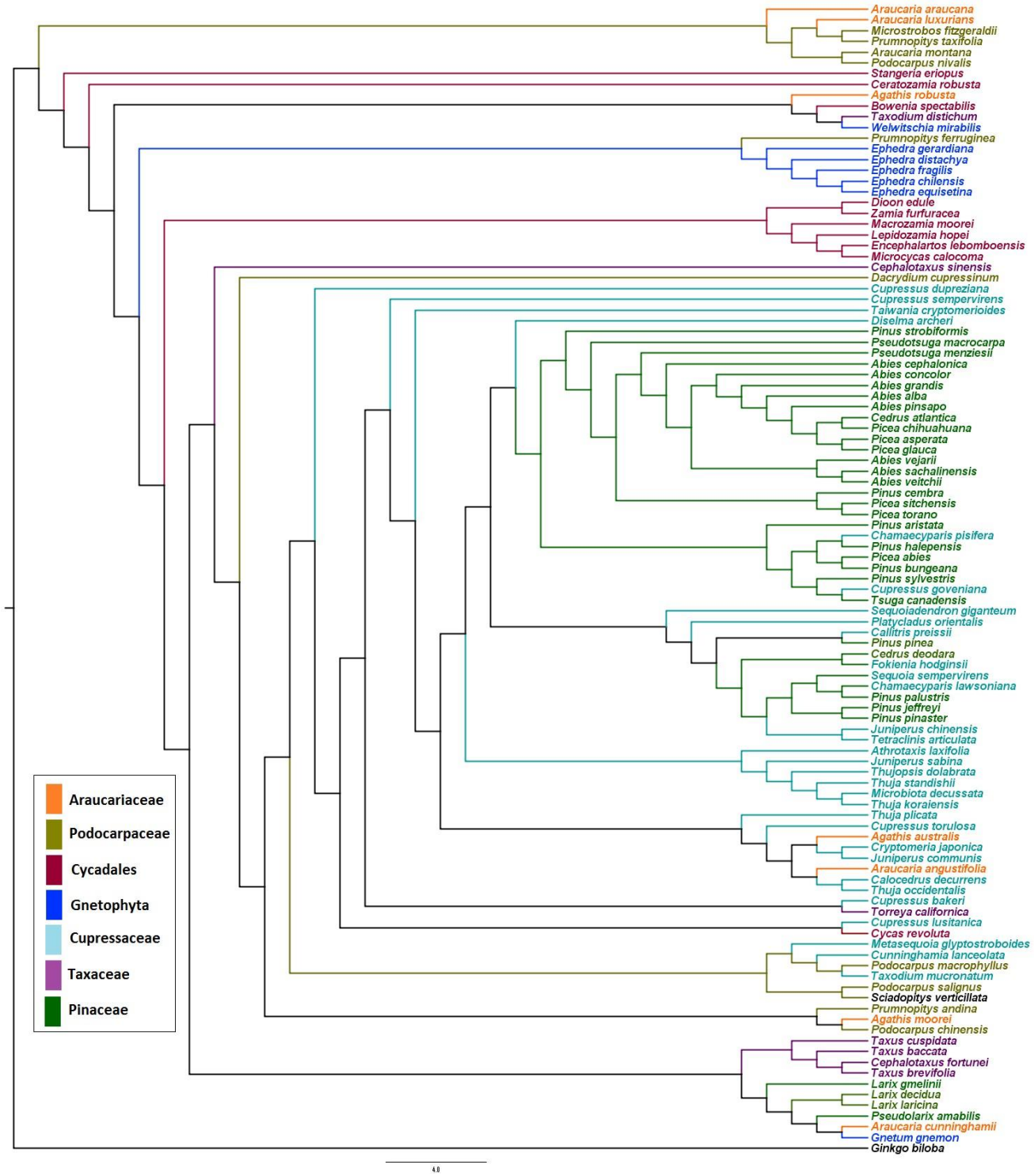



FIGURE 3. Tree produced with parsimony ratchet in TNT (1200 replicatons) from 24 fatty acid characters (Mongrand *et al.* (2001), showing clade specificity. Clades are colored according to the figure key.

TABLE 2 Clade Specificity Test: number of separate clades in which each genus falls

| | *Abies* | *Picea* | *Pinus* | *Junip.* | *Larix* | *Cupres.* | *Ephedra* | *Tax./Ceph.* | *Araucaria* | *Thuja* | *Cycads* | **mean** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Continuous | 7 | 4 | 9 | 3 | 3 | 6 | 4 | 3 | 4 | 4 | 9 | 5.09 |
| Continuous log | 8 | 6 | 9 | 3 | 3 | 6 | 4 | 4 | 5 | 4 | 10 | 5.64 |
| Cont. log RAxML | 6 | 6 | 9 | 3 | 3 | 6 | 4 | 3 | 5 | 4 | 10 | 5.36 |
| Binary | 6 | 6 | 8 | 3 | 2 | 6 | 4 | 4 | 5 | 4 | 8 | 5.09 |
| Disc., 0-3, Parsimony | 6 | 4 | 9 | 3 | 3 | 6 | 1 | 2 | 5 | 4 | 4 | 4.27 |
| Disc., 0-3, RAxML | 5 | 4 | 10 | 3 | 2 | 6 | 5 | 5 | 5 | 4 | 10 | 5.36 |
| Disc., 0-3, Bayes | 5 | 4 | 9 | 3 | 3 | 6 | 5 | 5 | 5 | 4 | 9 | 5.27 |
| Disc., 0-4 | 6 | 6 | 10 | 3 | 2 | 5 | 3 | 3 | 4 | 4 | 10 | 5.09 |
| Disc., 0-9 | 5 | 5 | 8 | 3 | 2 | 6 | 5 | 5 | 4 | 4 | 9 | 5.09 |
| Gap-weighted (RAxML) | 6 | 5 | 9 | 3 | 1 | 5 | 3 | 5 | 4 | 4 | 7 | 4.73 |
| Gap-weighted TNT parsimony | 5 | 5 | 9 | 3 | 3 | 6 | 5 | 5 | 5 | 4 | 5 | 5.00 |
| gap-weighted (Bayes) | 4 | 6 | 8 | 2 | 2 | 5 | 4 | 4 | 5 | 4 | 4 | 4.36 |

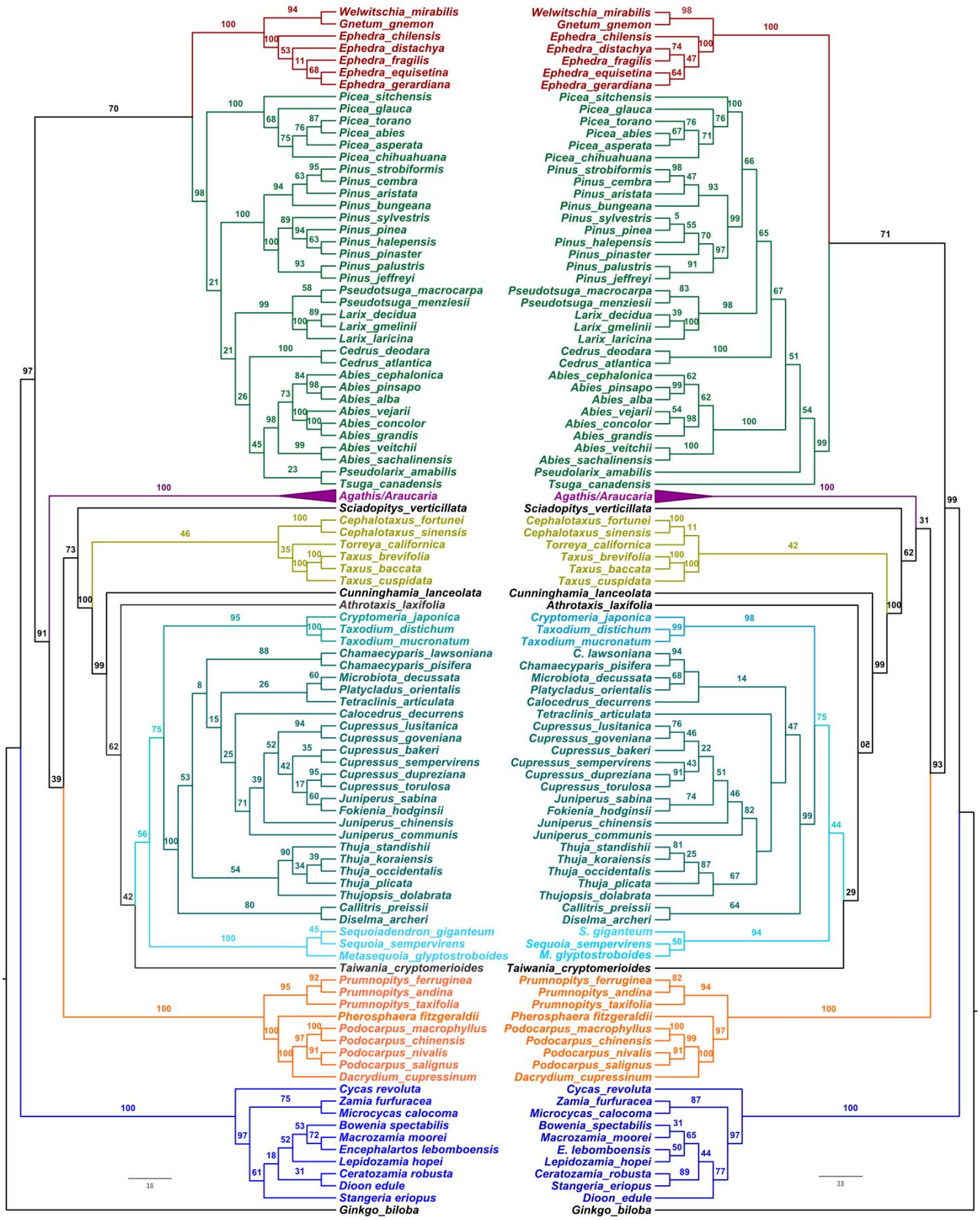



FIGURE 4. Tanglegram ofML rbcL tree and rbcL + 24 fatty acid traits demonstrating differences in resolution and branch support in combined tree, compared with rbcL tree alone on the right. See table 4 for specific results.

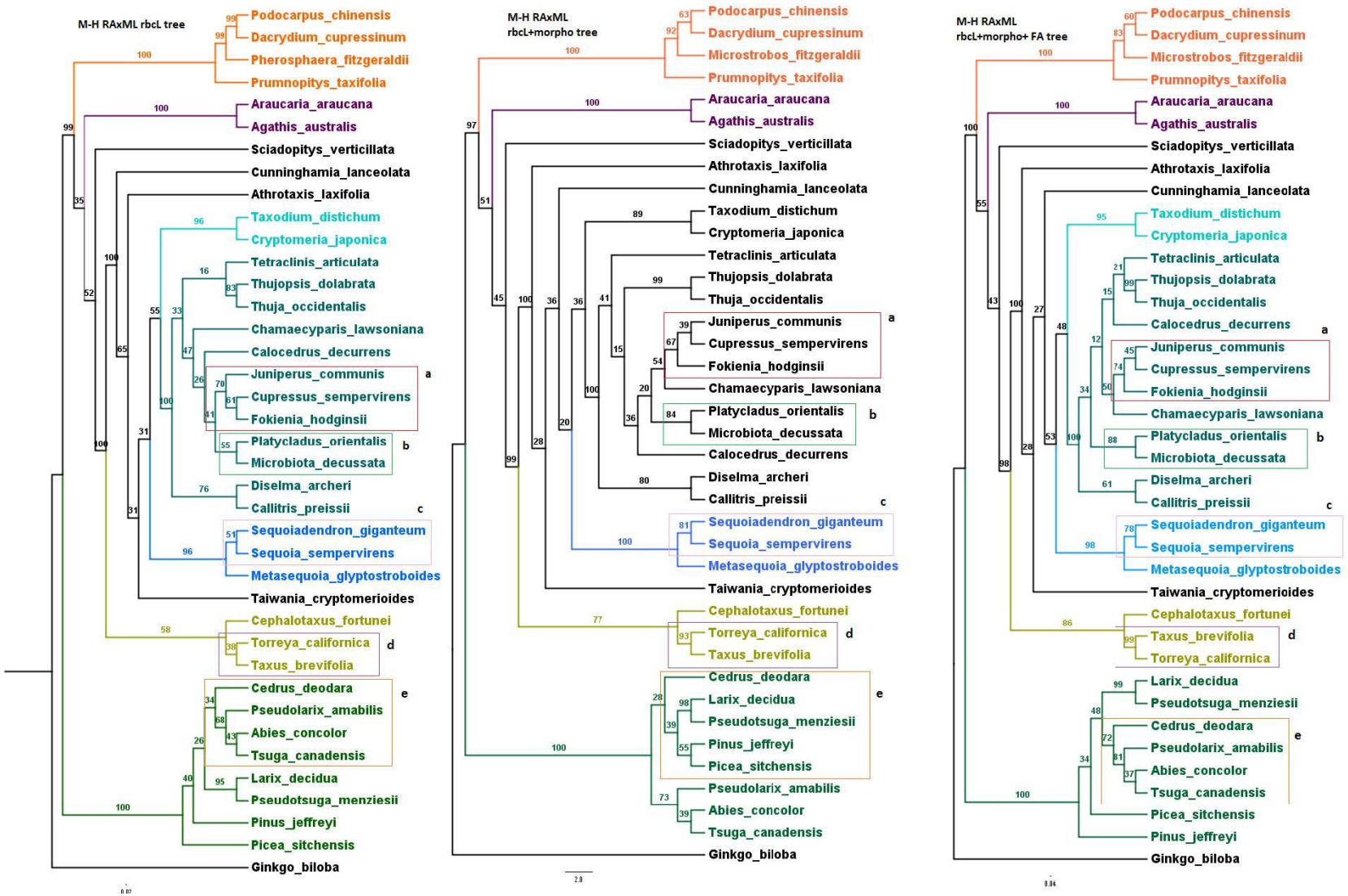



FIGURE 5. ML trees from Mongrand *et al.* (2001) and Hart (1987) combined data; rbcL, rbcL+morphological, and rbcL+morphological+leaffatty acid tree Comparison; mean branch support (MBS) is 63.58 for the rbcL tree, 65.94 for the combined rbcL + morphological tree, and 67.25 for the rbcL + morphological + leaffatty acid tree.

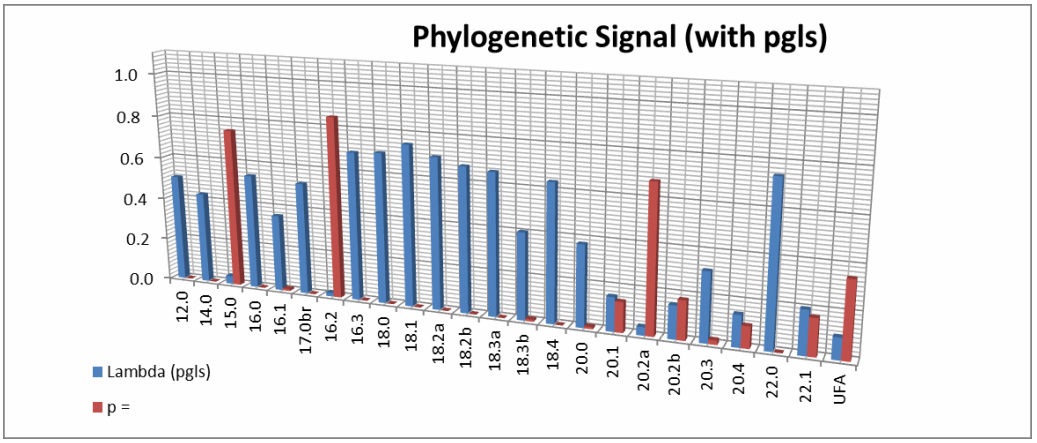



FIGURE 6. Phylogenetic signal (measured with pgls) of 24 leaf fatty acid traits, C12:0 to an unidentified long-chain fatty acid (UFA), where lambda = 1 is complete concordance with the phylogeny. The p value is the probability that lambda is significantly different than 0.

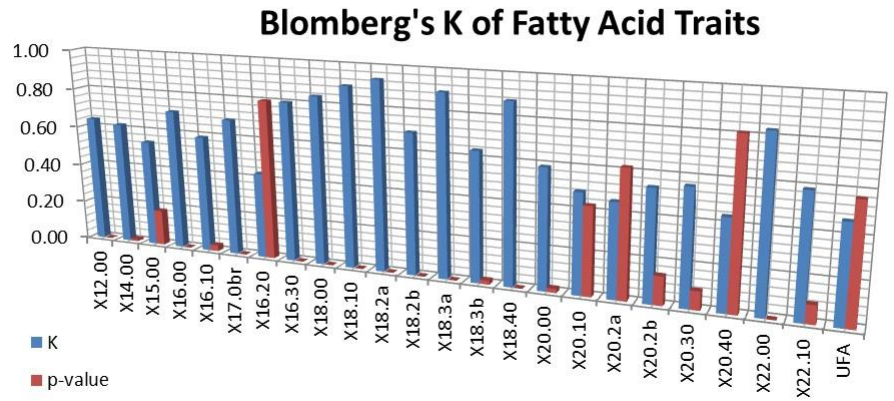



FIGURE 7. Blomberg's K measurement of phylogenetic signal of fatty acid traits, where 16 of the 24 leaf fatty acid traits show strong concordance with the phylogeny. The p value is the probability that lambda is significantly different than 0.

TABLE 3. Parsimony tree statistics for Mongrand *et al.* (2001) leaf fatty acid data, 24 traits, compared with the rbcL tree

| **TNT Parsimony Tree Statistics** | | | | |
|---|---|---|---|---|
| **No. Characters** | **Best Tree Score** | **Parsimony Informative Characters** | **CI** | **RI** |
| **Mongrand *et al.* (2001), 107 taxa rbcL + FA Combined analyses** | | | | |
| FA only (discrete, 0-3) | 247 | 24 | 0.166 | 0.603 |
| DNA Only | 3819 | 1075 | 0.467 | 0.832 |
| DNA + FA24 0-3 (parsimony, Bayes, 5M) | 4231 | 1099 | 0.432 | 0.809 |
| DNA + FA24 gap (parsimony, Bayes5M) | 4985 | 1099 | 0.369 | 0.767 |
| | | | | |
| **Mongrand *et al.* (2001) + Hart (1987); 39 taxa x 123 characters** | | | | |
| TNT Parsimony ratchet | | | | |
| DNA only (rbcL) | 1962 | 427 | 0.423 | 0.747 |
| Morphological only | 409 | 104 | 0.516 | 0.743 |
| FA only, 0-3 | 110 | 21 | 0.30 | 0.552 |
| rbcL + morpho | 2492 | 534 | 0.522 | 0.717 |
| rbcL + FA (0-3) | 2132 | 448 | 0.526 | 0.721 |
| rbcL + morpho + FA | 2661 | 555 | 0.501 | 0.697 |
| Morpho + FA (0-3) | 568 | 128 | 0.43 | 0.656 |

Notes: Homoplasy indices are given for trees derived from the combined Mongrand *et al.* (2001) and Hart (1987) data, comparing trees derived from the rbcL, morphological, and fatty acid data, and trees derived from combined rbcL + FA, rbcL +morpho, rbcL + FA, and rbcL + morpho + FA data.

TABLE 4. Blomberg's K measured with Phylosignal (Picante)

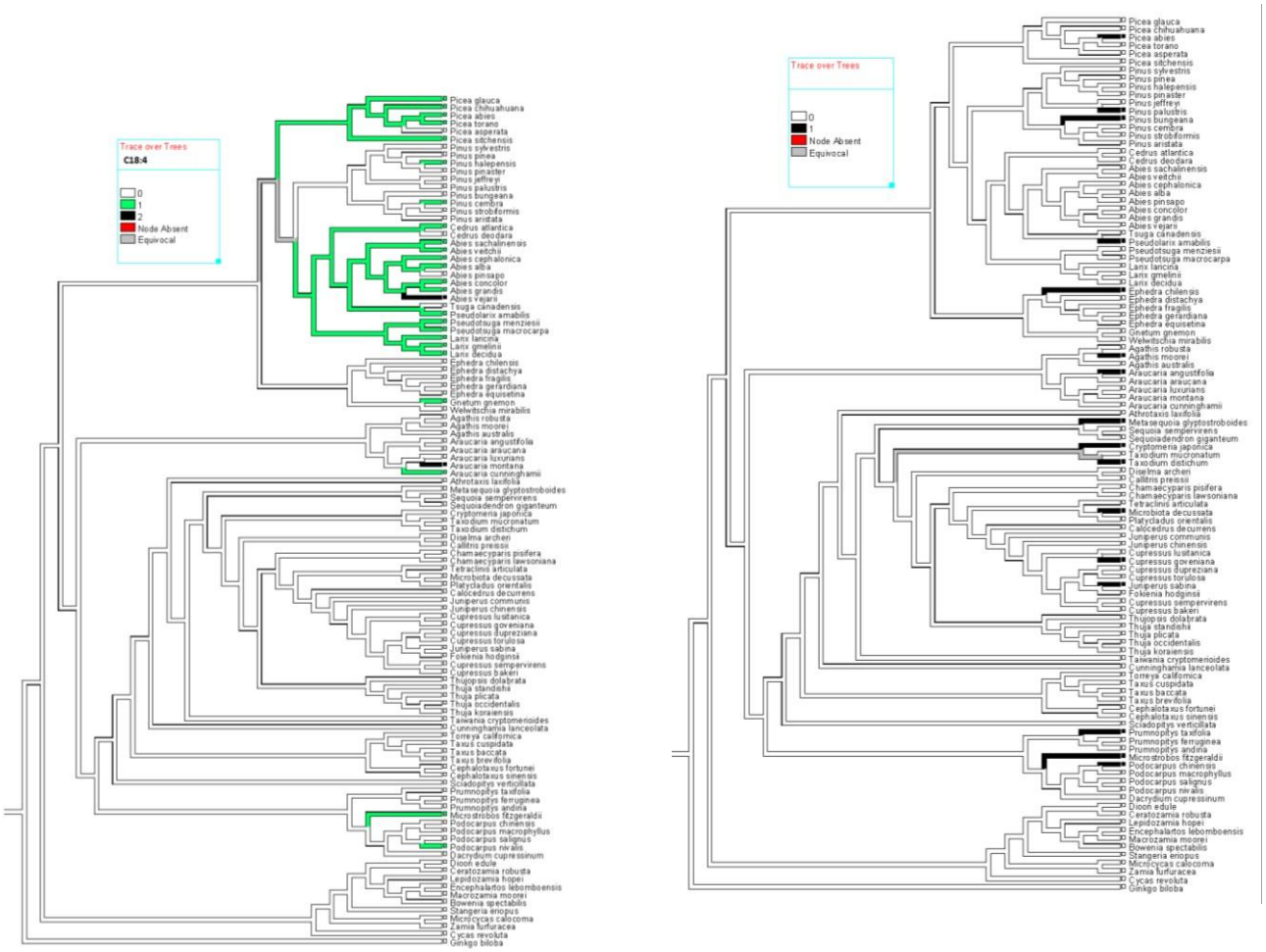
| FA | K | p-value |
|---|---|---|
| X12.00 | 0.64 | 0.00 |
| X14.00 | 0.62 | 0.01 |
| X15.00 | 0.54 | 0.18 |
| X16.00 | 0.71 | 0.00 |
| X16.10 | 0.59 | 0.03 |
| X17.0br | 0.69 | 0.00 |
| X16.20 | 0.43 | 0.80 |
| X16.30 | 0.80 | 0.00 |
| X18.00 | 0.84 | 0.00 |
| X18.10 | 0.90 | 0.00 |
| X18.2a | 0.94 | 0.00 |
| X18.2b | 0.70 | 0.00 |
| X18.3a | 0.90 | 0.00 |
| X18.3b | 0.64 | 0.02 |
| X18.40 | 0.88 | 0.00 |
| X20.00 | 0.59 | 0.02 |
| X20.10 | 0.49 | 0.43 |
| X20.2a | 0.46 | 0.62 |
| X20.2b | 0.54 | 0.14 |
| X20.30 | 0.56 | 0.09 |
| X20.40 | 0.44 | 0.81 |
| X22.00 | 0.83 | 0.00 |
| X22.10 | 0.59 | 0.09 |
| UFA | 0.47 | 0.57 |

Notes: p value reflects P-value of observed vs. random variance of PICs.

TABLE 5. Akaike Weights for the relative likelihood of 8 evolutionary models for leaf fatty acid traits

| | Brownian | Lambda | Delta | Kappa | OU | EB | trend | white |
|---|---|---|---|---|---|---|---|---|
| 12:0lauric acid (S) | 0.000 | **0.943** | 0.025 | 0.000 | 0.035 | 0.000 | 0.000 | 0.001 |
| 14:0myristic (S) | 0.000 | **0.971** | 0.012 | 0.000 | 0.013 | 0.000 | 0.000 | 0.005 |
| 15:0pentadecyclic (S) | 0.000 | 0.159 | 0.228 | 0.000 | 0.259 | 0.000 | 0.000 | **0.607** |
| 16:0palmitic (S) | 0.000 | **0.997** | 0.001 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 |
| 16:1palmitoleic (mono) | 0.000 | **0.752** | 0.112 | 0.000 | 0.129 | 0.000 | 0.000 | 0.066 |
| 17:0br 14-Mepalmitic (U) | 0.000 | 0.007 | 0.441 | 0.058 | **0.664** | 0.000 | 0.000 | 0.000 |
| 16:2hexadecadienoic (U) | 0.000 | 0.195 | 0.191 | 0.000 | 0.210 | 0.000 | 0.000 | **0.627** |
| 16:3cis-7,10,13 h-decatr. U) | 0.000 | **0.913** | 0.039 | 0.000 | 0.055 | 0.000 | 0.000 | 0.000 |
| 18:0 stearic (S) | 0.000 | **0.981** | 0.006 | 0.003 | 0.010 | 0.000 | 0.000 | 0.000 |
| 18:1 oleic (M) | 0.000 | **0.991** | 0.000 | 0.008 | 0.001 | 0.000 | 0.000 | 0.000 |
| 18:2ataxoleic (U) | 0.000 | **0.999** | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| 18:2blinoleic (U) | 0.000 | 0.012 | 0.483 | 0.000 | 0.668 | 0.000 | 0.000 | 0.001 |
| 18:3apinolenic (U) | 0.000 | **0.991** | 0.001 | 0.006 | 0.002 | 0.000 | 0.000 | 0.000 |
| 18:3blinolenic (U) | 0.000 | 0.007 | 0.507 | 0.000 | 0.650 | 0.000 | 0.000 | 0.004 |
| 18:4coniferonic (U) | 0.000 | **0.999** | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| 20:0arachidic (S) | 0.000 | **0.428** | 0.331 | 0.000 | **0.420** | 0.000 | 0.000 | 0.090 |
| 20:1gondoic | 0.000 | 0.226 | 0.279 | 0.000 | 0.343 | 0.000 | 0.000 | **0.541** |
| 20:2aeicosadienoic (U) | 0.000 | 0.202 | 0.190 | 0.000 | 0.209 | 0.000 | 0.000 | **0.625** |
| 20:2bbishomo-linoleic (U) | 0.000 | 0.303 | 0.225 | 0.000 | 0.258 | 0.000 | 0.000 | **0.560** |
| 20:3podocarpic (U) | 0.000 | 0.369 | 0.339 | 0.000 | **0.453** | 0.000 | 0.000 | 0.164 |
| 20:4juniperonic (U) | 0.000 | **0.781** | 0.056 | 0.000 | 0.056 | 0.000 | 0.000 | 0.153 |
| 22:0behenic (S) | 0.000 | 0.282 | 0.352 | 0.007 | **0.609** | 0.000 | 0.000 | 0.000 |
| 22:1 (erucic (U) | 0.000 | 0.310 | 0.180 | 0.000 | 0.198 | 0.000 | 0.000 | **0.592** |
| UFA (unidentified) (U) | 0.000 | 0.255 | 0.185 | 0.000 | 0.203 | 0.000 | 0.000 | **0.608** |

Notes: Bolded values are the most likely model. U= unsaturated, S= saturated. Shaded FA are delta 5-olefinic acids, universally found in the Gymnospermae, but rare in the Angiospermae.




(a) (b)

FIGURE 8. Reconstructed ancestral states for Fatty Acids C18:4, UFA showing clade specificity (a), and homoplasy (b)

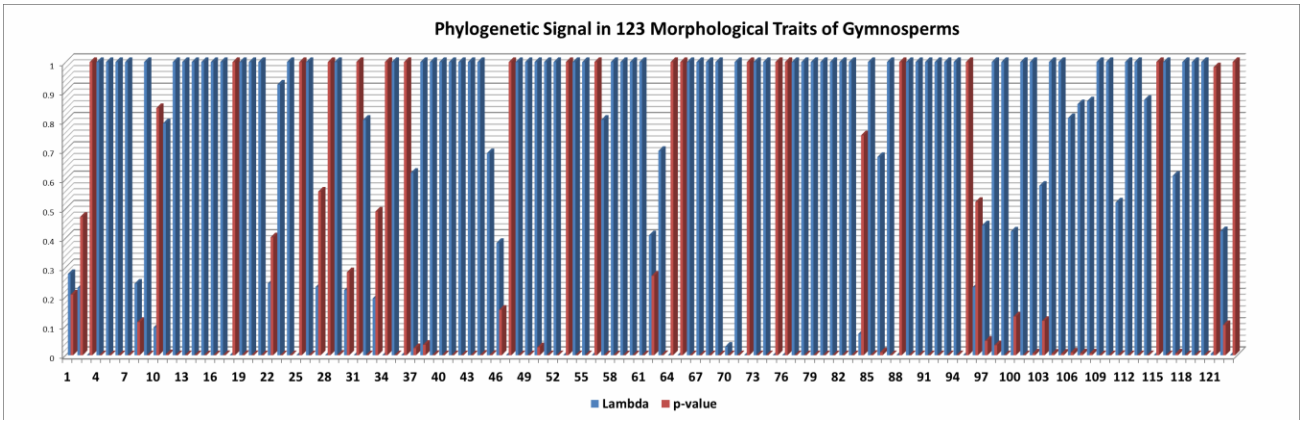



FIGURE 9. Phylogenetic signal of 123 morphological traits in Gymnosperms; characters 1-4, branching and growth patterns; 5-6, stem anatomy; 7-26, wood anatomy; leaves, 27-42; chemistry, 43-47; sex distribution, 48; microsporangiate strobilus, 49-55; microgametophyte, 56-70; megagametophyte and embryo, 71-97; ovulate strobilus, 98-113; ovules and seeds, 114-123 (see Hart, 1987 for complete details).

TABLE 6. Tree Congruence Tests showing the effect of different coding methods on the congruence of trees from FA alone, FA + morphological data compared with the rbcL tree

| **Mongrand *et al.* (2001); 107 taxa FA dataset** | | |
|---|---|---|
| Parsimony (TNT) Trees | **Icong Index** | **p value** |
| **DNA only (rbcL) compared with:** | | |
| FA (0-3) | 1.32 | 0.001* |
| FA (gap-coded) | 1.25 | 0.0046 |
| FA (binary) | 1.18 | 0.021 |
| **CADM.global matrix comparison tree congruence** | Kendall's W | p-value |
| rbcL/FA (continuous data), 9,999 permutations | 0.532 | 0.09 |
| **Mongrand *et al.* (2001) and Hart (1987) combined dataset; rbcL compared with:** | **Icong values** | |
| morpho | 1.64 | 2.43e-05* |
| FA (0-3) | 1.20 | 0.046 |
| FA (gap-coded) | 1.095 | 0.303 |
| FA (binary) | 1.20 | 0.046 |
| morpho + FA (0-3) | 2.19 | 1.80e-09* |
| | | |
| **PAUP ratchet 1500 iterations rbcL tree compared with:** | | |
| Comparison of parsimony trees with other trees: | ICong | p-value |
| Morpho | 1.423 | 0.0011* |
| FA (continuous data) | 1.18 | 0.021 |
| FA only (0-3) | 1.20 | 0.046 |
| morpho + FA (0-3) | 2.19 | 1.94e-09* |
| Morpho + FA (gap-coded) | 1.642 | 0.000024* |
| *Highly significantly congruent with rbcL tree, compared with random trees (<0.001)<br>*IcongIndex (higher values = more congruence) | | |

TABLE 7. Alpha (constraint parameter) measured in 24 leaf fatty acids from the Mongrand *et al.* (2001) data using comparative methods. (shaded values indicate strong selection pressure).

| | Alpha for leaf fatty acids | *Stabilizing selection | |
|---|---|---|---|
| **Fatty Acid** | **alpha*** | **Fatty acid** | **alpha*** |
| 12:0 | **0.359** | 18:3a | **0.167** |
| 14:0 | **0.543** | 18:3b | 0.324 |
| 15:0 | 0.678 | 18:4 | **0.187** |
| 16:0 | **0** | 20:0 | **0.467** |
| 16:1 | **0.481** | 20:1 | 0.822 |
| 17:0br | 0.303 | 20:2a | 16.831 |
| 16:2 | 16.537 | 20:2b | 0.784 |
| 16:3 | **0.258** | 20:3 | 0.466 |
| 18:0 | **0.197** | 20:4 | 15.238 |
| 18:1 | **0.17** | 22:0 | 0.206 |
| 18:2a | **0.198** | 22:1 | 17.109 |
| 18:2b | 0.293 | UFA | 17.288 |

TABLE S1. Genbank Accession Numbers

| **Supplementary Table 1** | **Genbank** |
|---|---|
| *Abies alba* Mill. | AB029652 |
| *Abies cephalonica* Loudon | FR831931 |
| *Abies concolor* (Gordon & Glend.) Lindl. ex Hildebr. | AB029648 |
| *Abies grandis* (Douglas ex D. Don) Lindl. | AB029646 |
| *Abies pinsapo* Boiss. | AB029656 |
| *Abies sachalinensis* (F. Schmidt) Mast. | AB015651 |
| *Abies veitchii* Lindl. | AB015649 |
| *Abies vejarii* Martínez | JN935622 |
| *Agathis australis* (D. Don) Steud | AF362993 |
| *Agathis moorei* (Lindl.) Mast. | U96480 |
| *Agathis robusta* (C. Moore ex F. Muell.) F.M. Bailey | U96484 |
| *Araucaria angustifolia* (Bertol.) Kuntze | U96470 |
| *Araucaria cunninghamii* Aiton ex D. Don | U96469 |
| *Araucaria luxurians* (Brongn. & Gris) de Laub. | U96464 |
| *Araucaria araucana* (Molina) K. Koch | AF249664 |
| *Araucaria montana* Brongn. & Gris | U96457 |
| *Athrotaxis laxifolia* Hook. | L25754 |
| Bowenia spectabilis Hook. Ex Hook. f. | AF394372 |
| *Callitris preissii* Miq. | JF725940 |
| *Calocedrus decurrens* (Torr.) Florin | L12569 |
| *Cedrus atlantica* (Endl.) Manetti ex Carrière | AF145457 |
| *Cedrus deodara* (Roxb.) G. Don | AF456381 |
| *Cephalotaxus fortunei* Hook. | AY450863 |
| *Cephalotaxus sinensis* (Rehder & E.H. Wilson) H.L. Li | AY450864 |
| *Ceratozamia robusta* Miq. | AF394346 |
| *Chamaecyparis lawsoniana* (A. Murray) Parl. | HM024272 |
| *Chamaecyparis pisifera* (Siebold & Zucc.) Endl. | HM024274 |
| *Cryptomeria japonica* (L.f.) D. Don | L25751 |
| *Cunninghamia lanceolata* (Lamb.) | L25757 |
| *Cupressus bakeri* Jeps. | AY988237 |
| *Cupressus dupreziana* A. Camus | AY988243 |
| *Cupressus goveniana* Gordon | AY380888 |
| *Cupressus lusitanica* Mill. | AY380889 |
| *Cupressus sempervirens* L. | HM024278 |
| *Cupressus torulosa* D. Don | AY988257 |
| *Cycas revolute* Thunb. | AY056556 |
| *Dacrydium cupressinum* Sol. Ex Lamb. | AF249634 |
| *Dioon edule* Lindl. | AF531203 |
| *Diselma archeri* Hook. f. | L12572 |
| *Encephalartos lebomboensis* I. Verd. | JQ025575 |
| *Ephedra chilensis* C. Presl. | AY755786 |
| *Ephedra distachya* L. | AY755793 |
| *Ephedra equisetina* Bunge | AY056572 |
| *Ephedra fragilis* Desf. | AY755784 |
| *Ephedra gerardiana* Wallich ex C.A. Meyer | AY755792 |
| *Fokienia hodginsii* (Dunn) A. Henry & H.H. Thomas | EF053219 |
| *Ginkgo biloba* L. | DQ069500 |
| *Gnetum gnemon* L. | L12680 |
| *Juniperus chinensis* L. | JQ512542 |
| *Juniperuscommunis* L. | AY988260 |
| *Juniperus Sabina* A. | HM024331 |
| *Larix decidua* Mill. | AB019826 |
| *Larix gmelinii* (Rupr.) Rupr. | AB303668 |
| *Larix laricina* (Du Roi) K. Koch | AF479878 |
| *Lepidozamia hopei* Regel | AF394342 |
| *Macrozamia moorei* F. Muell | AF394343 |
| *Metasequoia glyptostroboides* Hu & W.C. Cheng | JQ512562 |
| *Microbiota decussata* Kom. | L12575 |
| *Microcycas calocoma* (Miq.) A. DC. | AF531214 |
| *Pherosphaera fitzgeraldii* (F. Muell.) F. Muell ex Hook. f. | AF249646 |
| *Picea abies* (L.) H. Karst. | AB019825 |
| *Picea asperata* Mast. | AY056578 |
| *Picea chihuahuana* Mart. | EF440575 |
| *Picea glauca* (Moench) Voss | AY611035 |
| *Picea sitchensis* (Bong.) Carr. | EF440602 |
| *Picea torano* (Siebold ex K. Koch) Koehne | EF440605 |
| *Pinus aristata* Engelm. | AB019805 |
| *Pinus bungeana* Zucc. Ex Endl. | AB019808 |
| *Pinus cembra* L. | AB019795 |
| *Pinus halepensis* Mill. | AB019819 |
| *Pinus jeffreyi* Balf. | AY497235 |

| | |
|---|---|
| *Pinus pinaster* Ait. | AB019818 |
| *Pinus pinea* L. | AB019822 |
| *Pinus strobiformis* Engelm. | AB455588 |
| *Pinus sylvestris* L. | AB019809 |
| *Pinus wallichiana* A.B. Jacks. | AB019801 |
| *Platycladus orientalis* (L.) Franco | JQ512598 |
| *Podocarpus chinensis* Roxb. Wall. Ex Forbes | AF249602 |
| *Podocarpus macrophyllus* (Thunb.) Sweet | AF249616 |
| *Podocarpus nivalis* Hook. | AF249619 |
| *Podocarpus salignus* D. Don | HM593672 |
| *Prumnopitys andina* (Poepp. Ex Endl.) de Laub. | AF249655 |
| *Prumnopitys ferruginea* (G. Benn. Ex D. Don) de Laub. | AF249656 |
| *Prumnopitys taxifolia* (Banks & Sol. Ex D. Don) de Laub. | AF249658 |
| *Pseudolarix amabilis* (J. Nelson) Rehder | AB019829 |
| *Pseudotsuga macrocarpa* (Vasey) Mayr | GU457445 |
| Pseudotsuga menziesii (Mirb.) Franco | AY664856 |
| *Sciadopitys verticillata* (Thunb.) Siebold & Zucc. | L25753 |
| *Sequoia sempervirens* (D. Don) Endl. | L25755 |
| *Sequoiadendron giganteum* (Lindl.) J. Buchh. | AY056580 |
| *Stangeria eriopus* (Kunze) Baill. | AF394354 |
| *Taiwania cryptomerioides* Hayata | L25756 |
| *Taxodium distichum* (L.) Rich. | AF127427 |
| *Taxodium mucronatum* Ten. | JF725913 |
| *Taxus baccata* L. | AF456388 |
| *Taxus brevifolia* Nutt. | AF249666 |
| *Taxus chinensis* (Pilg.) Rehd. | AY450855 |
| *Taxus cuspidata* Siebold & Zucc. | DQ478793 |
| *Tetraclinis articulate* (Vahl) Masters | L12576 |
| *Thuja koraiensis* Nakai | JQ512617 |
| *Thuja occidentalis* L. | L12578 |
| *Thuja plicata* Donn ex D. Don. | AF127428 |
| *Thuja standishii* (Gordon) Carr. | HM024350 |
| *Thujopsis dolabrata* (Thunb. Ex L. f.) Siebold & Zuzz. | JQ512621 |
| *Torreya californica* Torr. | EF660732 |
| *Tsuga canadensis* (L.) Carriére | AY056581 |
| *Welwitschia mirabilis* Hook. f. | AF394335 |
| *Zamia furfuracea* L.f. in Aiton | AF394349 |

# Chapter 3

## Taming the sagebrushes: Island life results in loss of chemical defenses

### INTRODUCTION

Animals living in insular habitats and isolated from predators over evolutionary time are likely to modify or eliminate costly anti-predator behaviour such as startle and flight reflexes. This phenomenon is known as "island tameness," and reduced flocking in birds is a well-studied example (Beauchamp, 2004; Magurran, 1999; Vitousek *et al*., 2010) Defense reactions can be costly because they distract from other productive behaviour such as mate selection and foraging (Ydenberg & Dill, 1986).

While mechanisms of island tameness are starting to be understood in animals, such research with plants has been lacking, especially for island endemics, despite Carlquist's long-standing hypothesis of loss of competitiveness, poisonous compounds, and scented oils in island plants (Carlquist, 1970). Island tameness in plants is highly probable, because like animals, they normally initiate and maintain costly production of tens of thousands of individual chemicals from many different biosynthetic pathways known to act as toxins against microbes, other plants, and animals, or as olfactory repellents (Walters, 2011; Mithöfer and Boland, 2012; Wittstock and Gershenzon, 2002).

Of the many kinds of defense compounds produced, volatile organic constituents (VOCs) are the most numerous and versatile. VOCs consist of a complex mixture of simple volatile chemicals, primarily green leaf volatiles (GLV), hydrocarbons, oxygenated hydrocarbons (alcohols, aldehydes and ketones), monoterpenes, sesquiterpenes, and some volatile phenolic compounds like eugenol and thymol. These complex mixtures function primarily for communication and defense. Many plants have a palette of several hundred individual compounds to work with, allowing the ability to quickly tailor their response to potential threats. VOCs can be thought of as "semiochemicals,"(Allman *et al*., 2013; Law and Regnier, 1971; Maffei *et al.*, 2011) because they act as signals and mediators of myriad plant-plant (Ishizaki *et al.*, 2012), plant-animal (especially plant-insect), and plant-pathogen interactions (Heil and Karban, 2010). The bioactivity of VOCs is also context-specific and often species-specific. For defense, VOCs from numerous plants are also known to inhibit bacterial (Hendry *et al*., 2009) and fungal growth (Sokovic *et al*., 2002), seed germination (Halligan, 1975), seedling shoot and root growth through allelopathy (Yoshimura *et al.*, 2011), and herbivore feeding (Perez, 2013).

Here, a global phylogenetic and biochemical analysis of the VOC-rich sagebrush genus *Artemisia* (350–450 species) is conducted to assess the fate of its arsenal of terpenes and other

defense and signalling compounds following colonization of oceanic archipelagos. The ecologically well-characterised, rich chemistry of *Artemisia* and its occurrence in a diversity of continental and island settings make it an ideal system for examining hypotheses on the evolution of tameness in island plants.

## METHODS (in brief)

The number, range, complexity, and identity of individual compounds from the VOC profiles of 74 species of the world-wide genus *Artemisia were* investigated previously, representing all 5 subgenera (Hobbs and Baldwin, 2013). Five highly-supported clades in the genus were identified using a 4-gene dataset (2 nrDNA, 2 cpDNA), wherein biogeographical signal was also present. The recently described subg. *Pacifica*, with 4 island endemics, was part of the focus of this study.

Gas chromatography—mass spectrometry (GC/MS) was used to separate and identify from 120 to 180 compounds in each species using freshly collected plant material. Solid-phase microextraction (SPME) techniques were used to obtain samples of the total volatile mixtures of compounds directly from the plant samples for injection into the GC, without using heat or solvents in order to reduce methodological variation and solvent-volatile compound interactions. The evolutionary history and modes of trait evolution of the 1,227 chemical characters were inferred using comparative methods with the 2-, or 4-gene phylogeny (Hobbs and Baldwin, 2013), and then ancestral trait reconstruction was undertaken to determine where the most likely loss of defense compounds occurred in the island lineages.

## RESULTS

World-wide *Artemisia* volatile terpene diversity includes 1,227 identified compounds and 300-400 compounds that were visualized, but not identified. Two Hawaiian species, *A. australis* and *A. mauiensis*, compared with the non-Hawaiian member of subg. *Pacifica*, *A. chinensis (Hobbs and Baldwin, 2013)*, and a Canary island endemic, *A. canariensis*, were remarkably devoid of the often abundant and most frequently researched antimicrobial, phyotoxic, and anti-herbivore monoterpenes 1,8-cineol (Halligan, 1975; Mclean *et al.*, 2007; Perez, 2013; Yoshimura *et al.*, 2011), camphor (Barney *et al*., 2005; Sinclair *et al*., 1988; Negahban and Moharramipour, 2007) Pinocarvone (Goralka *et al.*, 1996; Panseri *et al*., 2011), thujone (Satyal *et al*., 2012; Vourc'h *et al.*, 2002), artemisia ketone (Verdeguer *et al.*, 2009), borneol (Fischer *et al.*, 1994), fenchene (Giorgi *et al*.,2010), *p*-cymene (Bray *et al.*, 1991) and the known bark beetle pheromone myrtenol (Zhang and Sun, 2006). These compounds were not detected in multiple accessions of both Hawaiian species, with the detection level to < 0.01% of the total peak area. The less isolated island endemic *A. chinensis* (Southeast Asia) was devoid of all 10 compounds except 1,8-cineol and the distantly related, less isolated *A. canariensis* (Canary Islands) was determined to have lost all these compounds except for fenchene.

Of the 74 species sampled, 96% (67 species) of those not endemic to islands had detectible levels of either 1,8-cineol or camphor, the two best-studied anti-herbivore and phytotoxic VOCs in *Artemisia*, and in 79% (58 species) both could be detected (Table 1). Besides the 4 island endemic species (with the exception of 1,8-cineol in *A. chinensis*), only 2 other sampled species had undetectable levels of both compounds. No species other than the island endemics was missing both compounds and α-thujone, another well-researched phytotoxin and anti-herbivory compound. The probability of any given species missing these 10 major defense compounds is 1.76E-06 (calculated with $(1-x_1)(1-x_2)(1-x_3)...$, where x = proportion of plants that have each compound). Independence of these traits is likely because they are produced by different terpene synthases, except α-, and β-thujone, but where an additional and separate biosynthetic step is required (Grausgruber-Gröger *et al.*, 2012).

Phylogenetic analysis of the 1227 character VOC data alone, or combined with a 4-gene (2 cpDNA and 2 nuclear DNA) dataset, showed a low level of phylogenetic signal, as measured by phylogenetic least squares (pgls) and Blomberg's K in ~9% of VOCs. These results coupled with comparative analysis in R, testing 4 evolutionary models, showed that >91% of VOCs do not follow Brownian motion (random walk), but are evolving conservatively under evolutionary constraint (unpublished results). Much biogeographical signal here was also seen throughout these data.

Based on the context of a 2-gene (nrDNA) phylogeny representative of every *Artemisia* subgenus, ancestral reconstruction showed a high likelihood (Table 2, fig. 1) that the 10 major defense compounds were lost gradually in the Hawaiian *Artemisia* lineage on the islands themselves and not in earlier ancestors.

## DISCUSSION

Despite the wide geographic distance between them, Pacific and Atlantic island lineages of *Artemisia* lost 9 or 10 of the most intensively-studied VOC compounds (except 1,8-cineol in *A. chinensis* and fenchene in *A. canariensis*), in a genus that widely contains them. Comparative phylogenetic analysis clearly shows a high likelihood of 9 of 10 traits occurring in the Hawaiian and Canary island ancestor, with their gradual disappearance after further generations on the islands (Fig. 2, Table 2). These results provide a novel phylogenetic and biochemical demonstration of Carlquist's (1974) principle of loss of competitiveness and defensiveness in island biota, as applied to plants.

The chemical arms race between plants and would-be pathogens and herbivorous animals, including insects, is often cited as the impetus for plants to produce an increasingly complex array of biologically active small molecular weight compounds. Becerra *et al.* (2009) used a molecular phylogeny from the genus *Bursera* to show that as new species diverged, they produced more compounds per species from an increasing number of biosynthetic pathways, increasing chemical diversity in the genus. It is also known that plants respond to insect damage

with a different and often more complex profile of volatiles than when undamaged or after mechanical injury (Karban and Baldwin, 1997). The tendency appears to be more complexity in defense compounds, both specifically and globally.

While some toxins are produced on demand, minimizing costs, some levels of toxic volatile compounds are constitutive(Wittstock and Gershenzon, 2002; Dicke and Van Loon, 2000). Many defense chemicals have to be stored for immediate use in compartments that protect plant cells and processes from deleterious effects from the toxins, also incurring a cost. A balance must be struck because lowered levels of protective compounds can leave the plant open to attack by pathogens and browsers, and encroachment by other species that may sequester and utilize local resources.

The Hawaiian endemics, which showed the greatest loss of chemical armament, are exceptionally isolated geographically, with an estimated arrival and onset of diversification in the archipelago based on fossil-calibrated molecular trees (Hobbs & Baldwin, 2013) occurring after the formation of Kaua`i (ca. 4.7 Ma) or less probably after the emergence of O`ahu about 3.0 Ma (Clague, 1996). For herbivores, shifts between host plants have been shown to be most likely when the old and new hosts are closely related genetically and chemically (Connor, *et al*., 1980). The absence of any native Hawaiian taxa of the biochemically distinctive tribe Anthemideae other than those studied here from the sole Hawaiian clade of *Artemisia* is consistent with a lack of herbivore or encroachment pressure on the original Hawaiian colonizing ancestor, as evidenced by the success of *A. australis*, which is abundant on every modern Hawaiian island.

Darwin theorized that introduced plant species, particularly on oceanic islands, have a higher likelihood of becoming invasive when closely-related plants are not present ("Darwin's naturalization hypothesis"), supported by some recent research of Schaefer *et al.*, 2011 (but see Park & Potter, 2013). The most likely source of the Hawaiian *Artemisia* ancestor is from tiny sticky cypselae with hooked barbs transported 8,000 km from the vicinity of Taiwan to Kaua`I (Hobbs and Baldwin, 2013), so the likelihood that insects adapted to *Artemisia* arrived at the same time is also low. A phylogenetic comparative analysis (Brown, 2013), based on a 2-gene tree for the endemic tephritid fruit flies, including *Trupanea artemisae*, a seed-feeding species reported on *A. mauiensis*, placed the arrival of a single colonizing ancestor for the entire lineage at 1.25 Ma, at least 2 Ma after the arrival of the *Artemisia* ancestor. *Trupanea artemisia* is embedded in the phylogeny such that its appearance may be about half that age (~0.6 Ma). Evidence for cospeciation between insects and plant hosts was not found. Competitive interference with other plants at that relatively early period in the history of the modern high islands is also presumed to have been less than in the modern flora, especially in light of evidence for a pre-Kaua`i gap in presence of high islands that could have served as sources of colonists (Price and Clague, 2002). The presumed long timeframe for such geographically and

phylogenetically isolated plant lineages to develop an insect fauna can be seen in *Quercus suber*, a European oak introduced to Australia in 1917, with no leaf-miners, and other examples from the genus cited by Connor *et al.*, (1980).

## METHODS

### Sample acquisition, processing, analysis

The fresh or fresh-frozen plant material (200 mg) was crushed with a mortar and pestle under liquid nitrogen to a fine power and added to 10 mL of saturated salt solution in 20 mL amber tubes. The GC/MS system consisted of an Agilent Technologies 6890 gas chromatograph coupled to an Agilent Technologies 5973 Network MSD (Agilent Technologies, Palo Alto, CA). A 60 m X 0.25 mm DB-1ms fused silica capillary column was used ($d_f$= 0.25 μm). A CTC Analytics AG Combi PAL (Zwingen, Switzerland) was used to perform the solid-phase microextraction (SPME) samplings and injections with a divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) 23-gauge, 50/30 □m SPME fiber assembly (Supelco). The incubation temperature was 35$^o$ C, incubation time was 5.00 min., agitator on time was 10 s, incubator off time was 1 s, agitator speed was 250 rpm, vial penetration was 35.00 mm, extraction time was 2.50 min, injector penetration was 48.00 mm, and desorption time was 60 s. The incubation temperature was 35$^o$ C, and the extraction time, i.e., the exposure of the fiber to the headspace containing the sample and volatile compounds in the 20 mL tube, was 90 seconds.

### Comparative analysis

The R platform RStudio (2012) was used for all comparative analyses. The phylogenetic tree (Hobbs and Baldwin, 2013) and matrix of volatile terpene trait values (original continuous data of 828 traits) was loaded for use in all analyses performed in the packages fitContinuous, pgls, and Phylosig. Fitting 4 different likelihood models with parametric comparative methods using the Geiger package fitContinuous (Harmon *et al.*, 2008) determined the best-fitting model of evolution for individual chemical characters. The Ornstein-Uhlenbeck (OU) model is a random walk with a central tendency that measures how stabilizing selection and random genetic drift act on phenotypic characters(Lande, 1976; Felsenstein, 1988). White is a model that corresponds to no BM-like evolution (no signal). Both of these models measure the amount of evolutionary constraint. Smith *et al*. (2011) reported that a high likelihood of the white model for a trait indicates very strong constrained evolution is acting. Geiger also estimated alpha (selection-strength parameter) during model testing. Low values of α are consistent with no selection and random drift (α =0), and higher values imply stronger selection pressure (Butler and King, 2004). The ace function of the R package ape (Paradis *et al.*, 2004) was used for all ancestral character estimations.

Table 1 Volatile defense compounds from 74 species of Artemisia (every subg. represented)

| | C70 | C400 | C309 | C320 | C373 | C354 | C519 | C651 | C688 | C701 |
|---|---|---|---|---|---|---|---|---|---|---|
| model | OU / lambda | OU | white | OU | white / OU | white | white | OU | white | white |
| | 1,8-Cineole | Camphor | Artemisia ketone | a-Thujone | b-Thujone | Borneol | Fenchene | Myrtenol | Para-cymene | Pinocarvone |
| A_abrotanum | 22.73 | 4.06 | 0.01 | 0.04 | 0 | 0.01 | 4.58 | 0.13 | 4.45 | 0.13 |
| A_absinthium | 1.3 | 0 | 0.01 | 8.43 | 0 | 0 | 0.03 | 0.03 | 0.07 | 0 |
| A_afra | 0.66 | 0 | 0 | 70.73 | 16.86 | 0 | 0.01 | 0.01 | 0 | 0.1 |
| A_alba | 2.05 | 0.02 | 16.59 | 0.17 | 0 | 0 | 0 | 0 | 0.12 | 0 |
| A_annua | 23.43 | 0.04 | 2.67 | 0.01 | 0.04 | 0 | 0.05 | 0.03 | 0.01 | 0.57 |
| A_anomala | 0 | 5.27 | 0 | 0.06 | 0 | 0.28 | 5.47 | 0.03 | 1.38 | 0.02 |
| A_arborescens | 0 | 0 | 0.12 | 2.57 | 70.19 | 0 | 0.02 | 0 | 0.8 | 0 |
| A_arbuscula | 54.87 | 1.94 | 0.54 | 0.21 | 0.08 | 0.11 | 0.22 | 0.07 | 4.42 | 0.33 |
| A_argentea | 4.22 | 0.26 | 0.01 | 0.11 | 0 | 0 | 0 | 0.06 | 19.61 | 7.75 |
| A_argyi | 22.66 | 0.27 | 0.14 | 0.02 | 0 | 3.51 | 1.56 | 0 | 1 | 0.02 |
| A_australis | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A_barrelieri | 10.16 | 1.47 | 0 | 0.01 | 0.01 | 0.63 | 2.34 | 0.03 | 1.8 | 0.1 |
| A_biennis | 0.33 | 0.22 | 0.01 | 0 | 0 | 0.01 | 0 | 0 | 0.14 | 0 |
| A_caerulescens_2 | 0 | 0.06 | 0.04 | 1.56 | 15.02 | 0 | 0.11 | 0.03 | 0.42 | 0.11 |
| A_californica_9 | 10.03 | 0.01 | 0.19 | 64.08 | 5.12 | 0 | 0.05 | 0.01 | 1.27 | 0.13 |
| A_campestris_caud | 0.32 | 0 | 0.01 | 0 | 0 | 0.05 | 2.58 | 0.01 | 0.09 | 0.01 |
| A_cana | 11.78 | 14.23 | 2.02 | 19.62 | 0.03 | 0 | 0.02 | 0.13 | 0.93 | 0 |
| A_canariensis | 0 | 0 | 0 | 0 | 0 | 0 | 0.27 | 0 | 0 | 0 |
| A_capillaris | 0 | 0 | 0.03 | 0.07 | 0.59 | 0 | 6.64 | 0 | 0.16 | 0.02 |
| A_carruthii | 7.43 | 0.04 | 0 | 0.02 | 0.01 | 0.26 | 0.21 | 0.01 | 13.39 | 0.01 |
| A_chamaemelifolia | 27.36 | 0.43 | 0.02 | 0.13 | 0.36 | 0 | 0.03 | 0 | 1.06 | 0 |
| A_chinensis | 5.58 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A_cretacea | 3.14 | 11.35 | 0.02 | 0 | 0 | 0.6 | 4.28 | 0.04 | 2.97 | 1.25 |
| A_crithmifolia | 4.07 | 0.63 | 1.24 | 1.17 | 1.11 | 0 | 0.05 | 0 | 0 | 0.03 |
| A_douglasiana | 2.05 | 2.47 | 0 | 14.12 | 0.6 | 0 | 0.62 | 0 | 1.31 | 0.26 |
| A_dracunculus | 0 | 0.36 | 0 | 0 | 0 | 0.02 | 0.37 | 0 | 0 | 0.01 |
| A_ferganensis | 18.28 | 6.54 | 0.01 | 0.04 | 0.01 | 2.2 | 2.45 | 0 | 0 | 0.11 |
| A_fillifolia | 7.88 | 0.7 | 0.5 | 0.16 | 0 | 0.13 | 0.27 | 0 | 3.95 | 0.33 |
| A_franserioides | 18.57 | 55.24 | 2.22 | 0.04 | 0 | 1.29 | 8.85 | 0 | 0.56 | 1.28 |
| A_frigida | 24.7 | 34.1 | 0 | 0.01 | 0 | 2.85 | 7.32 | 0.05 | 3.9 | 0.5 |
| A_gmelinii | 5.28 | 0 | 0.03 | 0.17 | 0.17 | 0 | 0.01 | 1.18 | 0.66 | 15.35 |
| A_gorgonum | 0 | 33.51 | 0 | 0.08 | 0 | 0.14 | 4.78 | 0.01 | 12.82 | 0.01 |
| A_herba-alba | 23.21 | 0.24 | 0 | 0.11 | 0 | 0.23 | 0.1 | 0 | 0.05 | 0.03 |
| A_inaequifolia | 0.08 | 63.4 | 0 | 0 | 0 | 6.95 | 15.79 | 0.29 | 0.17 | 1.11 |
| A_japonica | 0 | 3.18 | 0.07 | 0.06 | 0.04 | 0 | 0.06 | 0 | 0.05 | 0 |
| A_kawakamii | 0 | 0.01 | 0.01 | 0.01 | 0 | 0 | 0.17 | 0.01 | 0.36 | 0.02 |
| A_laciniata | 0 | 0 | 0.04 | 0.01 | 0 | 0 | 0.01 | 0 | 0.19 | 0.02 |
| A_lactiflora | 0.13 | 0.02 | 0.08 | 0.14 | 0.12 | 0 | 2.32 | 0 | 0.09 | 0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A_lagocephala | 14.25 | 2.05 | 0 | 2.6 | 0 | 6.11 | 11.67 | 0 | 1.06 | 0.55 |
| A_leucophylla | 6.37 | 0.14 | 0.01 | 0 | 0 | 1.05 | 0.89 | 0.07 | 3.93 | 0.25 |
| A_ludoviciana_A8 | 4.85 | 2.86 | 15.85 | 0 | 0 | 0 | 0.99 | 0.03 | 0.28 | 0.04 |
| A_maritima | 0.89 | 0.09 | 0.03 | 3.55 | 0 | 0 | 0.58 | 0.01 | 9.26 | 0.29 |
| A_mauiensis | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A_michauxiana | 10.85 | 12.68 | 0.01 | 0 | 0 | 0.05 | 3.27 | 0.01 | 6.36 | 0 |
| A_molinieri | 3.36 | 0.01 | 0.01 | 0.08 | 0.03 | 0 | 0.1 | 0 | 11.22 | 0.02 |
| A_momiyamae | 6.73 | 0.55 | 0.04 | 0.04 | 0 | 6.61 | 9.01 | 0 | 1.32 | 0.1 |
| A_montana | 13 | 11.86 | 0.01 | 0.46 | 0.04 | 10.89 | 9.06 | 0.06 | 1.05 | 0.49 |
| A_nesotica | 24.14 | 29.93 | 0.03 | 0.02 | 0.01 | 1.02 | 3.59 | 0.06 | 2.25 | 0.33 |
| A_nitida | 1.6 | 0.45 | 0.06 | 0.51 | 0.24 | 1.62 | 0.92 | 1.21 | 0 | 0.71 |
| A_norvegica | 0 | 0.11 | 0.26 | 0 | 0 | 0.03 | 0.36 | 0 | 1.04 | 0.07 |
| A_nova | 20.42 | 12.63 | 0 | 0 | 0 | 0.55 | 0 | 0.02 | 12.13 | 0.07 |
| A_palmeri | 26.36 | 0.04 | 0.01 | 0.27 | 0.93 | 0 | 0.05 | 0.02 | 0 | 0.31 |
| A_pontica | 22.46 | 0.02 | 33.44 | 16.63 | 1.84 | 0 | 0.01 | 0.04 | 0.49 | 0.07 |
| A_potentilloides | 4.31 | 0.13 | 0 | 0.38 | 6.01 | 2.43 | 1.6 | 0.02 | 0 | 0.08 |
| A_princeps | 12.22 | 0.04 | 45.72 | 0.1 | 0 | 0 | 0.01 | 0 | 1.67 | 0.04 |
| A_pycnocephala | 0.33 | 0.06 | 0.18 | 0.17 | 0.07 | 0 | 0.59 | 0 | 9.17 | 0.01 |
| A_pygmaea | 38.44 | 23.47 | 0 | 0 | 0 | 2.68 | 0 | 0 | 5.49 | 0.21 |
| A_rothrockii | 9.12 | 14.01 | 0.01 | 55.61 | 7.06 | 0.79 | 4.06 | 0.05 | 1.13 | 0.24 |
| A_roxburghiana | 1.96 | 0.02 | 0.01 | 0.01 | 0 | 4.81 | 14.69 | 0.86 | 3.37 | 0.66 |
| A_ruthiae | 0.17 | 57.67 | 0 | 0 | 0 | 0.77 | 5.28 | 0.01 | 0 | 0.22 |
| A_rutifolia | 10.13 | 0.01 | 0 | 56.09 | 21.71 | 0 | 0.02 | 0.05 | 0.13 | 0.1 |
| A_schmidtiana | 7.83 | 0.17 | 0.01 | 0.2 | 0 | 0 | 1.61 | 0.09 | 4.28 | 0.31 |
| A_scoparia | 0.35 | 0.56 | 0 | 0.85 | 0.76 | 0 | 0.09 | 0 | 14.05 | 0.01 |
| A_spiciformis | 23.62 | 15.41 | 0 | 0 | 0 | 1.08 | 4.13 | 0 | 0 | 0.2 |
| A_spinescens | 0.01 | 13.75 | 0 | 12.72 | 1.4 | 0.37 | 2.32 | 0 | 4.55 | 0 |
| A_stelleriana | 25.92 | 0 | 0.02 | 0.02 | 0 | 0.01 | 0.27 | 0.02 | 0.28 | 1.88 |
| A_stolonifera | 3.25 | 0.33 | 0.01 | 37.53 | 4.74 | 0 | 0.09 | 0.01 | 0 | 0.08 |
| A_suksdorfii | 13.3 | 32.75 | 0.1 | 0.2 | 0.03 | 0.86 | 7.63 | 0.02 | 20.74 | 0.04 |
| A_thuscula | 0.43 | 18.53 | 0.15 | 0.15 | 0 | 0.76 | 3.94 | 0 | 7.83 | 0 |
| A_tridentata | 3.84 | 0.29 | 2.02 | 1.71 | 0 | 0 | 0.32 | 0 | 1.28 | 0 |
| A_velesciarca | 34.03 | 24.74 | 0.01 | 0.05 | 0 | 8.2 | 6.4 | 0.01 | 0.23 | 0.67 |
| A_verlotiorum | 0.69 | 1.68 | 0 | 63.87 | 14.89 | 0 | 0.02 | 0.01 | 0.1 | 0.08 |
| A_vulgaris | 58.24 | 3.95 | 0.08 | 0.03 | 0.01 | 0.55 | 0.94 | 0.06 | 1.24 | 0.31 |
| A_wurzelii | 4.07 | 0 | 0 | 0.09 | 0.05 | 0.02 | 0.15 | 0 | 8.79 | 0.03 |
| Achillea_millefolium | 1.36 | 3.01 | 0 | 0.3 | 0 | 0.13 | 0.11 | 0.02 | 0.1 | 0.05 |
| | 721.2 | 524.07 | 124.71 | 438.2 | 170.18 | 70.66 | 166.41 | 4.92 | 213.02 | 38.13 |

*Likelihood of a presence of the compound in the MRCA

Table 2 Ancestral State Reconstruction

**Defense phytochemicals of Interest**

Variable Traits

| Trait | 1 | 6 | 2 | 3 | 5 | 4 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| | C70 | C400 | C309 | C320 | C373 | C354 | C519 | C651 | C688 | C701 |
| **Node** | Mean Ancestral Values | | | | | | | | | |
| Hawaiian Ancestors | 1,8-cineole | Camphor | Artemisia Ketone** | α-thujone* | β-thujone | Borneol | Fenchene* | Myrtenol* | p-cymene* | Pinocarvone |
| 103 | 8.878 | 7.258 | 0.261/0.621 | 1.217 | 2.1018 | 0.747 | 1.381 | 0.557 | 1.393 | 0.42982 |
| 139 | 7.174 | 4.76698 | 0.937/0.052 | 1.351 | 3.0328 | 0.615 | 1.367 | 0.574 | 1.554 | 0.28769 |
| 140 | 8.110 | 2.6249 | 0.476/04.76 | 1.685 | 5.273 | 0.426 | 1.394 | 0.673 | 1.543 | 0.24587 |
| 144 (Canary Is. anc.) | 13.465 | 3.745 | 0.047 | 1.717 | 2.238 | 0.623 | 1.314 | 0.814 | 0.452 | 0.772 |
| **Tips** | 0.000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A. australis | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A. chinensis | 5.58 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A. mauiensis | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A. canariensis | 0 | 0 | 0 | 0 | 0 | 0 | 0.27 | 0 | 0 | 0 |
| | | | | | | | | | | |
| Binary Traits | | | | | | | | | | |
| **D-statistic (binary)** | 0.214 | 0.560 | 0.864 | 0.660 | 0.867 | 0.475 | 0.130 | 1.040 | 0.710 | 0.660 |
| BM structure (p=) | 0.328 | 0.750 | 0.001 | 0.030 | 0.000 | 0.043 | 0.420 | 0.000 | 0.020 | 0.034 |
| no structure (p=) | 0.002 | 0.032 | 0.234 | 0.066 | 0.219 | 0.006 | 0.005 | 0.579 | 0.121 | 0.075 |

*Traits are coded 0-4, as follows: xx **Note: try model-testing in fitDiscrete with 0-4-coded characters; can use BM for ASR and mapping?

**Trait coded as presence-absence (binary); D=0.862, Probability of D resulting from BM phylogenetic structure (p= 0.001)

**Likelihood of state 0/1 (ace function of ape, type = discrete)

Figure 1　Ancestral reconstruction of p-cymene on the nDNA phylogeny

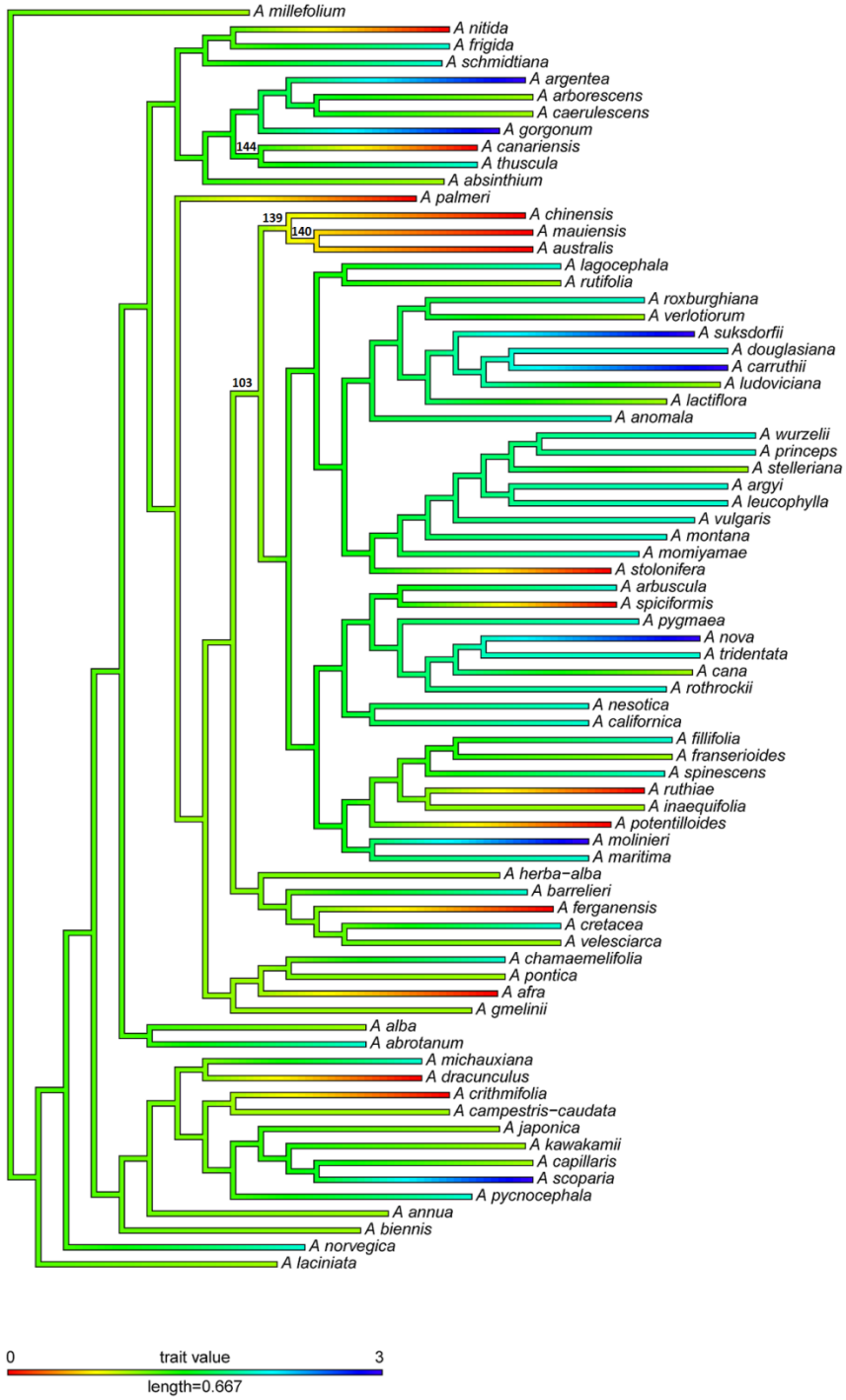
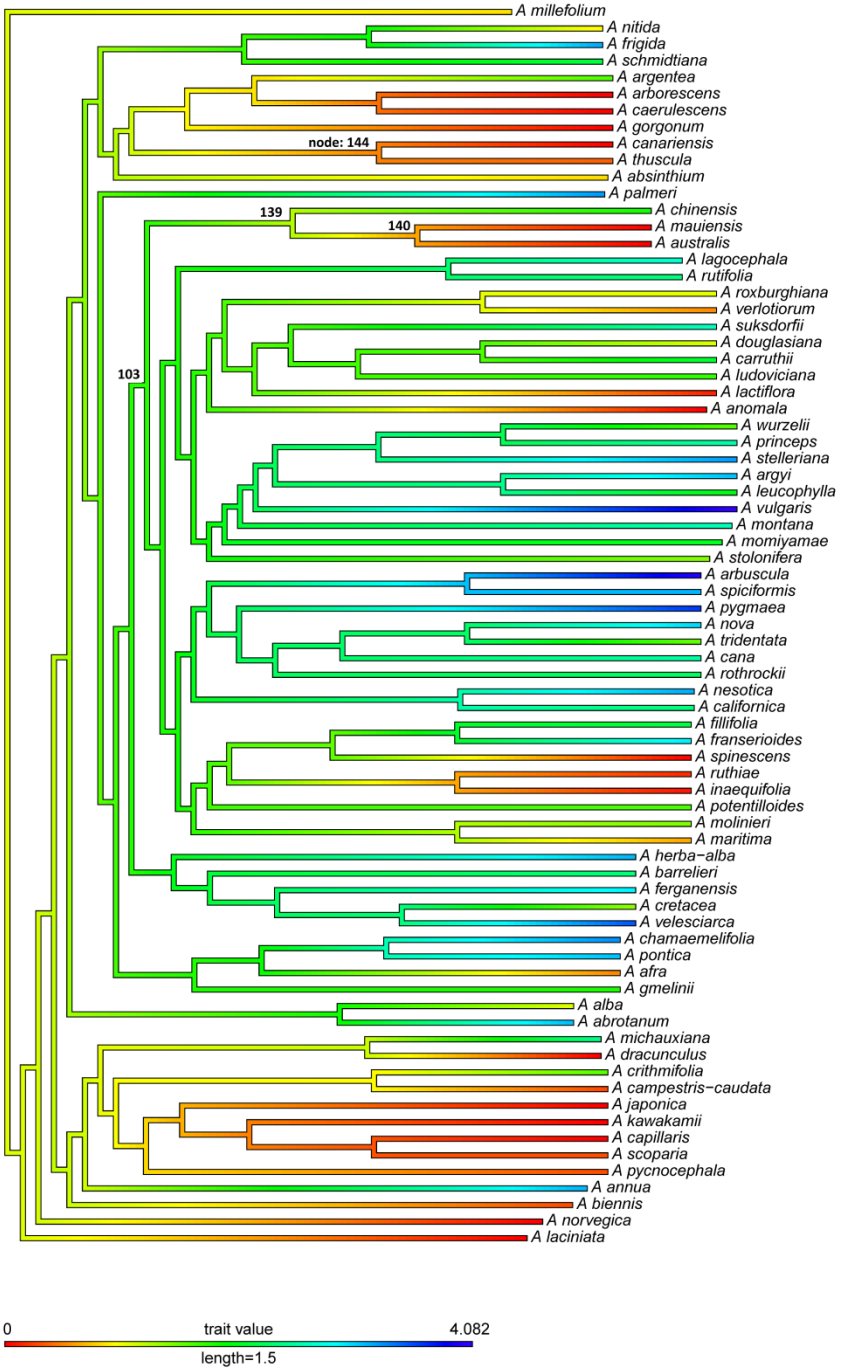


Figure 2　Ancestral reconstruction of 1,8 cineol on the nDNA phylogeny



References Cited


Ackerly, D.D. (2009) Evolution, origin and age of lineages in the Californian and Mediterranean floras. *Journal of Biogeography*, **36**, 1221–1233.

Ackman, R.G. (2009) Simplification of analyses of fatty acids in fish lipids and related lipid samples. *Acta Medica Scandinavica,* **222**, 99-103.

Adams, D.C. & Rosenberg, M.S. (1998) Partial warps, phylogeny, and ontogeny: a comment on Fink and Zelditch (1995). *Systematic Biology,* **47**, 168-173.

Adams, D.C., Rohlf, F.J., & Slice, D.E. (2004) Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology,* **71**, 5-16.

Agrawal, A.A., Salminen, J.-P., & Fishbein, M. (2009) Phylogenetic trends in phenolic metabolism of milkweeds (*Asclepius*): evidence for escalation. *Evolution,* **63**, 663-673.

Ahlholm, J.U., Helander, M., Lehtimäki, P.W., & Saikkonen, K. (2002) Vertically transmitted fungal endophytes: different responses of host-parasite systems to environmental conditions. *Oikos,* **99**, 173-183.

Allmann, S., Späthe, A.; Bisch-Knaden, S.; Kallenbach, M.; Reinecke, A.; Sachse, S.; Baldwin, I.T. & Hansson, B.S. (2013) Feeding-induced rearrangement of green leaf volatiles reduces moth oviposition. *eLife, 2:e00421*.

Armbruster, W.S., Lee, J., & Baldwin, B.G. (2009) Macroevolutionary patterns of defense and pollination in Dalechampia vines: Adaptation, exaptation, and evolutionary novelty. *PNAS,* **106**, 18085-18090.

Arnold, A.E. (2007) Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. *Fungal Biology Reviews,* **21**, 51-66.

Assis, L.C.S. (2009) Coherence, correspondence, and the renaissance of morphology in phylogenetic systematics. *Cladistics,* **25**, 528-544.

Austin, J. J., Soubrier, J.; Prevosti, F.J.; Prates, L.; Trejo, V.; Mena, F. & Cooper, A. (2013) The origins of the enigmatic Falkland Islands wolf. *Nature Communications* **4**, doi:10.1038/ncomms2570.

Axelrod, D.I. (1987) The late Oligocene Creede flora, Colorado. *University of California Publications in Geological Sciences*, **130**, 1–235.

Baldwin, B.G. & Markos, S. (1998) Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae). *Molecular Phylogenetics and Evolution,* **10**, 449–463.

Baldwin, B.G. & Robichaux, R.H. (1995) Historical biogeography and ecology of the Hawaiian silversword alliance (Asteraceae): New molecular phylogenetic perspectives. *Hawaiian biogeography: Evolution on a hot spot archipelago* (ed. by W. L. Wagner and V. A. Funk), pp. 259–287. Smithsonian Institution Press, Washington.

Baldwin, G. & Wagner, W.L. (2010) Hawaiian angiosperm radiations of North American origin. *Annals of Botany,* **105**, 849–879.

Ballard, H.E. & Sytsma, K.J. (2000) Evolution and biogeography of the woody Hawaiian violets (*Viola*, Violaceae): Arctic origins, herbaceous ancestry and bird dispersal. *Evolution,* **54**, 1521–1532.

Barney, J. N., Hay, A. G., & Weston, L. A. (2005) Isolation and characterization of allelopathic volatiles from mugwort. *Journal of Chemical Ecology* **31**, 247-265.

Beauchamp, G. (2004) Reduced flocking by birds on islands with relaxed predation. *Proceedings of the Royal Society B* **271**, 1042.

Beauchamp, P.S., Chea E., Dimaano J.G., Dev V., Ly B. Mirand A.E. (2009) Essential oil composition of six *Lomatium* species attractive to Indra swallowtail butterfly (*Papilio indra*): principal component analysis against essential oil composition of *Lomatium dissectum* var. *multifidum*. *Journal of Essential Oil Research* **21**, 535-542.

Becerra, J. X., Noge, K., & Veable, L. (2009) Macroevolutionary chemical escalation in an ancient plant-herbivore arms race. *Proceedings of the National Academy of Science* **106**, 18062-18066.

Benito, B.M., Montserrat, M.-O., Muñoz, J.L., Jorite, J. & Peñas, J. (2009) Assessing extinction-risk of endangered plants using species distribution models: A case study of habitat depletion caused by the spread of greenhouses. *Biodiversity and Conservation,* **18**, 2509–2520.

Bentham, G. (1873) Notes on the classification, history, and geographical distribution of Compositae. *Journal of the Linnean Society (Botany) London*, **13**, 335–557.

Biffin, E., Conran, J. G., & Lowe, A. J. (2010) Podocarp Evolution: A Molecular Phylogenetic Perspective. *Ecology of the Podocarpaceae in Tropical Forests* pp. 1-20. Smithsonian Institution Scholarly Press, Washington, DC.

Birding Hawaii (2002) Hawai`i bird sightings for 1998. http://www.birdinghawaii.co.uk/Sightings19982.htm (Accessed Dec., 2011).

Blomberg, S.P., Garland, T.Jr., & Ives, A.R. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution,* 57, 717-745.

Bonaventure, G., Salas, J.J., Pollard, M.R., & Ohlrogge, J.B. (2003) Disruption of the FATB Gene in Arabidopsis Demonstrates an Essential Role of Saturated Fatty Acids in Plant Growth. *The Plant Cell,* **15**, 1020-1033.

Bowen, L. & Van Vuren, D. (1997) Insular endemic plants lack defenses against herbivores. *Conservation Biology,* **11**, 1249-1254.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling,* **157**, 281–300.

Bray, R.O., Wambolt, C.L., & Kelsey, R.G. (1991) Influence of sagebrush terpenoids on mule deer preference. *Journal of Chemical Ecology,* **17**, 2053-2062.

Bremer, K. & Humphries, C.J. (1993) Generic monograph of the Asteraceae-Anthemideae. *Bulletin of the Natural History Museum, Botany series,* **23**, 71–177.

Brown, F.B.H. (1922) The secondary xylem of Hawaiian trees. *Occasional Papers of Bernice P. Bishop Museum*, **8**, 217–371.

Brown, J.M. (2013) Personal Communication from author: Pacific Entomological Conference, 2012.

Burney, D.A., James, H.F., Burney, L.P., Olson, S.L., Kikuchi, W., Wagner, W.L., Burney, M., McCloskey, D., Kikuchi, D., Grady, F.V., Gage, R. & Nishek, R. (2001) Fossil evidence for a diverse biota from Kaua`i and its transformation since human arrival. *Ecological Monographs*, **71**, 615–641.

Butler, M.A. & King, A.A. (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist,* **164**, 683-695.

Butler, M.A. & King, A.A. (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist,* **164**, 683-695.

Cappers, R.T.J. (1993) Seed dispersal by water: A contribution to the interpretation of seed assemblages. *Vegetation History and Archaeobotany,* **2**, 173–186.

Carlquist, S. (1966) The biota of long-distance dispersal. II. Loss of dispersibility in Pacific Compositae. *Evolution*, **20**, 30–48.

Carlquist, S. (1967) The biota of long-distance dispersal. V. Plant dispersal to Pacific Islands. *Bulletin of the Torrey Botanical Club,* **94**, 129–162.

Carlquist, S. (1970) *Hawaii: a Natural History*, 2nd edn. The Natural History Press, Garden City, NY.

Carlquist, S. (1974) *Island biology*. Columbia University Press, New York.

Carlquist, S. (1980). *Hawaii, a natural history: Geology, climate, native flora and fauna above the shoreline, second edition*. Pacific Tropical Botanical Garden, Honolulu.

Carlquist, S. (1996) Plant dispersal and the origin of Pacific island floras. *The origin and evolution of Pacific Island biotas, New Guinea to Eastern Polynesia: Patterns and processes* (ed. by A. Keast and S.E. Miller), pp. 153–164. SPB Academic Publishing, Amsterdam.

Carstens, B.C. & Knowles, L.L. (2007) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from Melanoplus grasshoppers. *Systematic Biology,* **56**, 411.

Chamberlin, C.J. (1926) Hybrids in cycads. *Botanical Gazette* 81, 401-418.

Chaw, S.-M.,Walters, T.W., Chang, C.-C., Hu, S.-H., Chen, S.-H. (2005) A phylogeny of cycads (Cycadales) inferred from chloroplast *matK* gene, *trnK* intron, and nuclear rDNA ITS region. *Molecular and Phylogenetics and Evolution*, **37**, 214-234.

Chen, F., Ro, D.-K., Petri, J, Gershenzon, J, Bohlmann, J, Pickersky, E. Tholl, D. (2004) Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the

formation of the volatile monoterpene 1,8-cineole. *Plant Physiology* **135**, 1956-1966.

Chumley, T. W., McCoy, S. K. R. Raubeson, L. A. (2008) Gnedeep: exploring Gnetalean affinities in seed plant phylogeny with83 plastid genes. Botany 2008: Joint Annual Meeting of Canadian Botanical Association, American Fern Society, American Society of Plant Taxonomists, and the Botanical Society of America, Vancouver, British Columbia, Canada [online abstract, http://2008.botanyconference.org/engine/search/index.php?func=detail & aid=770] .

Clague, D. A. (1996) The growth and subsidence of the Hawaiian-Emperor volcanic chain. *The origin and evolution of Pacific Island biotas, New Guinea to Eastern Polynesia: patterns and processes* (ed. by A. Keast and S. E. Miller), pp. 35-50. A. Keast and S.E. Miller, Amsterdam.

Clague, D.A. & Dalrymple, G.B. (1987) The Hawaiian-Emperor volcanic chain. *Volcanism in Hawaii* (eds. R.W. Decker, T.L. Wright & P.H. Stauffer), pp. 1–54. U.S. Geological Service Professional Paper 1350. U.S. Government Printing Office, Washington.

Connor, E. F., Faeth, S. H., Simberloff, D., & Opler, P. A. (1980) Taxonomic isolation and the accumulation of herbivorous insects: a comparison of introduced and native trees. *Ecological Entomology* **5**, 205-211.

Cox, J.G. & Lima, S.L. (2006) Naiveté and an aquatic-terrestrial dichotomy in the effects of introduced predators. *Trends in Ecology and Evolution,* **21**, 674-680.

Cross, T. & Taggart, R.E. (1982) Causes of short-term sequential changes in fossil plant assemblages: Some considerations based on a Miocene flora of the northwest United States. *Annals of the Missouri Botanical Garden,* **69**, 679–734.

de Vienne, D.M., Giraud, T., & Martin, O.C. (2007) A congruence index for testing topological similarity between trees. *Bioinformatics,* **23**, 3119-3124.

Degener, O. & Degener, I. (1946) Compositae, Family 344, *Artemisia*. *Flora Hawaiiensis or new illustrated flora of the Hawaiian Islands, Books 1–4*.

Degnan, J.H. & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution,* 24, 332-340.

Dev, V., Whaley, W.H., Bailey, S.R., Chea, E., Dimaano, J.G., Jogani, D.K., Ly, B., & Eggett, De. (2010) Essential oil composition of nine Apiaceae species from western United States that attract the female Indra Swallowtail butterfly (Papilio indra). *Biochemical Systematics and Ecology,* **38**, 538-547.

Dicke, M. & Van Loon, J.J.A. (2000) Multitroophic effects of herbivore-induced plant volatiles in an evolutionary context. *Entomologia Experimentalis et Applicata,* **97**, 237-249.

Doel, R.R. & Dalrymple, G.B. (1973) Potassium-argon ages and paleomagnetism of the Waianae and Koolau volcanic series. *Geological Society of America Bulletin,* **84**, 1217–1242.

Doyle, J.J. (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany,* **17**, 144-163.

Drake D.R. (1992) Seed dispersal of *Metrosideros polymorpha* (Myrtaceae): A pioneer tree of Hawaiian lava flows. *American Journal of Botany*, **79**, 1224–1228.

Dreyfuss, M. M. & Chapela, I. H. (1994) Potential of fungi in the discovery of novel, low-molecular weight pharmaceuticals. (ed. by V. P. Gullo), p. 49. Butterworth-Heinemann, Boston.

Drummond, A.J. & Rambaut, A. (2007) BEAST, ver. 1.54. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. & Wilson, A. (2010) Geneious v5.5, Available from http://www.geneious.com.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G. & Solomon, W. (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics,* **161**, 1307–1320.

Drummond, A.J., Simon, Y.W., Ho, N.R. & Rambaut, A. (2007) *A rough guide to BEAST 1.4*. http://beast-mcmc.googlecode.com/files/BEAST14_Manual_6July2007.pdf.

Dunstan, G.A., Brown, M.R., & Volkman, J.K. (2005) Cryptophyceae and rhodophyceae; chemotaxonomy, phylogeny, and application. *Phytochemistry,* **66**, 2557-2570.

Dupont-Nivet, G., Hoorn, C. & Konert, M. (2008) Tibetan uplift prior to the Eocene-Oligocene climate transition: Evidence from pollen analysis of the Xining Basin. *Geology*, **40**, 987–990.

Dupont-Nivet, G., Krijgsman, W., Langereis, C.G., Abels, H.A., Dai, S. & Fang, X. (2007) Tibetan plateau aridification linked to global cooling at the Eocene–Oligocene transition, *Nature*, **445**, 635–638.

Eder, K. (1995) Gas chromatographic analysis of fatty acid methyl esters. *Journal of Chromatography B: Biomedical Sciences and Applications*, **671**, 113-131.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Suisan, A., Hijmans, R., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography,* **29**, 129–151.

Engler, A. (1879) *Versuch einer Entwicklungsgeschichte der Pflanzenwelt insbesondere der Florengebiete seit der Tertiäperiode*. Wilhelm Engelmann, Leipzig.

Erpenbeck, D. & van Soest, R.W.M. (2006) Status and perspective of sponge chemosystematics. *Marine Biotechnology,* **9**, 2-19.

Evans , C.S. & Bell, E.A. (1978) 'Uncommon' amino acids in the seeds of 64 species of Caesalpinieae. *Phytochemistry* , **17**, 1127-1129.

Evans, C.S., Qureshi, Y., Bell, E.A. (1977) Free amino acids in the seeds of *Acacia* species.

*Phytochemistry* **16**, 565-570.

Farris, J.S. (1970) Methods for computing Wagner trees. *Systematic Zoology,* **19**, 83-92.

Farris, J.S. (1990) Phenetics in camouflage. *Cladistics,* **6**, 91-100.

Farris, J.S. (2007) Coding of continuous characters, revisited. *Darwiniana,* 45, 9-10.

Felsenstein, J. (1988) Phylogenies and quantitative characters. *American Naturalist,* **125**, 1-15.

Felsenstein, J. (1988) Phylogenies and quantitative characters. *American Naturalist,* **125**, 1-15.

Felsenstein, J. (2008) Consense, ver. 3.69. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Fischer, N.H., Williamson, G.B., Weidenhamer, J.D., & Richardson, D.R. (1994) In search of allelopathy in the Florida scrub: the role of terpenoids. *Journal of Chemical Ecology,* **20**, 1355-1380.

Flagel, L.E., Rapp, R.A., Grover, C.E., Widrlechner, M.P., Hawkins, J., Grafenberg, J.L., Alvarez, I., Chung, G.Y., & Wendel, J.F. (2008) Phylogenetic, morphological, and chemotaxonomic incongruence in the North American endemic genus Echinacea. *American Journal of Botany,* **95**, 756-765.

Fosberg, F.R. (1948) Derivation of the flora of the Hawaiian islands. *Insects of Hawaii, vol. 1*, pp. 107–119. University of Hawaii Press, Honolulu.

Freckleton, R.P., Harvey, P.H., & Pagel, M. (2002) Phylogenetic analysis and comparative data: A test and review of evidence. *American Naturalist,* **160**, 712-726.

Fritsch, R.M. & Keusgen, M. (2006) Occurrence and taxonomic significance of cysteine sulphoxides in the genus Allium L. (Alliaceae). *Phytochemistry,* **67**, 1127-1135.

Fritz, S.A. & Purvis, A. (2010) Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology,* **24**, 1042-1051.

Funk, V.A. & Wagner, W.L. (1995) Biogeographic patterns in the Hawaiian Islands. *Hawaiian biogeography: Evolution on a hot spot archipelago* (ed. by W.L. Wagner & V.A. Funk), pp. 379–419. Smithsonian Institution Press, Washington.

Gadek, P.A., Alpers, D.L., Heslewood, M.M., Quinn, C.J. (2000) Relationships within Cupressaceae *sensu lato*: a combined morphological and molecular approach. *American Journal of Botany*, **87**, 1044-1057.

Galtier, N., Gouy, M. & Gautier, C. (1996) SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. *Computer Applications in the Biosciences,* **12**, 543–548.

Garcia, S., Garnatje, T., McArthur, E.D., Pellicer, J., Sanderson, S.C. & Vallès, J. (2011) Taxonomic and nomenclatural rearrangements in *Artemisia* subgen. *Tridentatae*, including redefinition of *Sphaeromeria* (Asteraceae, Anthemideae). *Western North American Naturalist,* **71**: 158–163.

Garcia, S., McArthur, E.D. & Pellicer, J. (2010) A molecular phylogenetic approach to western North American endemic *Artemisia* and allies (Asteraceae): Untangling the sagebrushes. *American Journal of Botany*, **98**, 638–653.

Garcia-Cruz, J. & Sosa, V. (2006) Coding quantitative character data for phylogenetic analysis: a comparison of five methods. *Systematic Botany,* **31**, 302-309.

Garzione, C.N., Quade, J., DeCelles, P.G. & English, N.B. (2000) Predicting paleoelevation of Tibet and the Himalaya from $\delta^{18}$O vs. altitude gradients in meteoric water across the Nepal Himalaya. Earth and Planetary Science Letters 183: 215–229.

GBIF (2011) GBIF: www.gbif.org.

Gershenzon, J. & Dudareva, N. (2007) The function of terpene natural products in the natural world. *Nature Chemical Biology,* **3**, 408-414.

Gibson, L., Barrett, B. & Burbidge, A. (2007) Dealing with uncertain absences in habitat modelling: A case study of a rare ground-dwelling parrot. *Diversity and Distributions,* **13**, 704–713.

Gillespie, R.G., Baldwin, B.G., Waters, J.M., Fraser, C.I., Nikula, R. & Roderick, G.K. (2012) Long-distance dispersal: A framework for hypothesis testing. *Trends in Ecology and Evolution* 27, 47–56.

Giorgi, A., Panseri, S.; Nanayakkarawasam, C.; Lozzia, G.; Chiesa, L.; Soncin, S.; Biondi, P. & Cocucci, M. (2010) HS-SPME method to monitoring "in vivo" volatile compounds emitted from *Achillea collina* infested by aphids. *Planta Medica* **76**, (2010).

Givnish, T.J., Millam, K.C., Mast, A.R., Paterson, T.B., Theim, T.J., Hipp, A.L., Henss, J.M., Smith, J.F., Wood, K.R. & Sytsma, K.J. (2009) Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society B, Biological Sciences,* **276**, 407-416.

Goffman, F.D., Thies, W., Velasco, L. (1999) Chemotaxonomic value of tocopherols in Brassicaceae. *Phytochemistry*, **50**, 793-798.

Goloboff, P.A., Farris, J.S., & Nixon, K.C. (2008) TNT, a free program for phylogenetic analysis. *Cladistics,* **24**, 774-786.

Goloboff, P.A., Mattoni, C.I., & Quiteros, A.S. (2006) Continuous characters analyzed as such. *Cladistics,* **22**, 589-601.

Goralka, R.J.L., Schumaker, M.A., & Langenheim, J.H. (1996) Variation in Chemical and Physical Properties During Leaf Development in California BayTree *(Umbellularia californiaa):* Predictions Regarding Palatability for Deer. *Biochemical Systematics and Ecology,* **24**, 93-103.

Graham, S.W., Kohn, J.R., Morton, B.R., Eckenwalder, J.E., & Barrett, C.H. (1998) Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. *Systematic Biology,* **47**, 545-567.

Grausgruber-Gröger, S., Schmiderer, C., Steinborn, R., & Novak, J. (2012) Seasonal influence on

gene expression of monoterpene synthases in *Salvia officinalis* (Lamiaceae). *Journal of Plant Physiology,* **169**, 353-359.

Gray, A. (1884) *Synoptical Flora of North America.* Ivison, Blakeman, Taylor, and Co., New York.

Gray, J. (1958) *Plant microfossils from the Miocene of the Columbia Plateau, Oregon*. Ph.D. Thesis, University of California, Berkeley.

Grayer, R.J., Chase, M.W., & Simmonds, M.S.J. (1999) A comparison between chemical and molecular characters for the determination of phylogenetic relationships among plant families: an appreciation of Hegnauer's "Chemotaxonomie der Pflanzen". *Biochemical Systematics and Ecology,* **27**, 369-393.

Guan, Z., Chen, S., Chen, F., Yin, D., Liu, Z., Tang, J. & Yang, F. (2010) Salt tolerance screening of 32 taxa from *Chrysanthemum* and its relative genera. *Scientia Agricultura Sinica,* **43**, 4063–4071.

Guindon, S. & O. Gascuel (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology,* **52**, 696–704.

Guo, B., Wang, Y., Sun, X., & Tang, K. (2008) Bioactive natural products form endophytes: a review. *Applied Biochemistry and Microbiology,* **44**, 136-142.

Guppy, H. B. (1903) *Observations of a naturalist in the Pacific between 1896 and 1899, vol*. *2: Plant dispersal*. Macmillan and Co., Ltd., London.

Halligan, J. P. Toxic Terpenes from Artemisia-Californica. *Ecology (Washington D C)* **56**, 999-1003 (1975).

Hanslin, H.M. & Eggen, T. (2005) Salinity tolerance during germination of seashore halophytes and salt-tolerant grass cultivars. *Seed Science Research,* **15**, 43–50.

Hao, D.C., Xiao, P.G., Huang, B.L., Ge, G.B., Yang, L. (2008). Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned Bayesian analyses of chloroplast and nuclear DNA sequences. *Plant Systematics and Evolution*, **276**, 89-104.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E., & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics,* **24**, 129-131.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E., & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics,* **24**, 129-131.

Harris, E.S.J. (2009) Phylogenetic and environmental lability of flavonoids in a medicinal moss. *Biochemical Systematics and Ecology,* **37**, 180-192.

Hart, J.A. (1987) A cladistic analysis of conifers: preliminary results. *Journal of the Arnold Arboretum,***68**, 269-305.

Hawksworth, D. L. (1976) Lichen chemotaxonomy. *Lichenology: Progress and Problems* (ed. by D. H. Brown, D. L. Hawksworth, and R. H. Bailey), pp. 139-184. Academic Press, New York.

Heath, T.A., Hedtke, S.M., & Hillis, D.M. (2008) Taxon sampling and the accuracy of

phylogenetic analyses. *Journal of Systematics and Evolution,* **46**, 239-257.

Heethoff, M., Laumann, M., Weigmann, G., & Günther, R. (2011) Integrative taxonomy: Combining morphological, molecular and chemical data for species delineation in the parthenogenetic Trhypochthonius tectorum complex (Acari, Oribatida, Trhypochthoniidae). *Frontiers in Zoology,* **8**, 1-10.

Hegnauer, R. (1962) *Chemotaxonomie der Planzen*. Birkhäuser-Verlag, Basel.

Heil, M. & Karban, R. (2010) Explaining evolution of plant communication by airborne signals. *Trends in Ecology and Evolution,* **25**, 137-144.

Hendry, E. R., Worthington, T., Conway, B. R., & Lambert, P. A. (2009) Antimicrobial efficacy of eucalyptus oil and 1,8-cineole alone and in combination with chlorhexidine digluconate against microorganisms grown in planktonic and biofilm cultures. *Journal of Antimicrobial Chemotherapy* **64**, 1219-1225.

Hijmans, R.J., Cameron, S.E., Parra, J., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* **25**, 1965–1978.

Hillebrand, W. (1888) *Flora of the Hawaiian Islands*. Williams & Norgate, New York.

Ho, S.Y.W. (2007) Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology,* **38**, 409–414.

Hobbs, C.R. & Baldwin, B.G. (2013) Asian origin and upslope migration of Hawaiian *Artemisia* (Compositae-Anthemideae). *Journal of Biogeography,* **40**, 442-454.

Hollingsworth, P.M. Forresta, L.L., Spougeb, L.L., Hajibabaeic, M., Ratnasingham, S. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Science*, **106**, 12794-12797.

Hsieh, Y.S.Y. & Harris, P.J. (2012) Structures of xyloglucans in primary cell walls of gymnosperms, monilophytes (ferns sensu lato) and lycophytes. *Phytochemistry,* **79**, 87-101.

Huang, W.Y., Cai, Y.Z., Hyde, K.D., Corke, H., & Sun, M. (2010) Biodiversity of endophytic fungi associated with 29 traditional Chinese medicinal plants. *Fungal Diversity,* **33**, 61-75.

Huang, Y., Xia, G., Zheng, B., Huang, J. & Shao, X. (2007) Morphological changes and physiological responses of *Crossostephium chinense* to salt stress. *Acta Agriculturae Universitatis Jiangxiensis,* **29**, 389–394.

Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics,* **17**, 754–755.

Ickert-Bond, S.M., Wojciechowski, M.F. (2004) Phylogenetic relationships in *Ephedra* (Gnetales): evidence from nuclear and chloroplast DNA sequence data. *Systematic Botany*, **29**, 834-849.

Ishizaki, S., Shiojiri, K., Karban, R., & Ohara, M. (2012) Clonal growth of sagebrush (*Artemisia tridentata*) (Asteraceae) and its relationship to volatile communication. *Plant Species*

*Biology,* **27**, 69-76.

Jiang H., Xie Z., Koo H.J., McLaughlin S.P., Timmermann B.N., Gang D.R. (2006) Metabolic profiling and phylogenetic analysis of medicinal *Zingiber* species: tools for authentication of ginger (*Zingiber officinale* Rosc.). *Phytochemistry*, **67**, 1673-1685.

Johns, R.B., Nichols, P.D., Perry, G.J. (1979) Fatty acid composition of ten marine algae from Australian waters. *Phytochemistry*, **18**, 789-802.

Julkunen-Tiitto, R. (1989) Phenolic constituents of *Salix*: a chemotaxonomic survey of further Finnish species. *Phytochemistry*, **28**, 2115-2125.

Karban, R. & Baldwin, I. T. (1997) *Induced Responses to Herbivory*. Chicago University Press, Chicago.

Katoh, A. (2009) Multiple alignment of DNA sequences with MAFFT. *Methods in Molecular Biology,* **537**, 39–64.

Kitching, I.J., Forey, P.L., Humphries, C.J., Williams, D.M. (1998). *Cladistics: The Theory and Practice of Parsimony*. Oxford, Oxford University Press.

Klaehn, F.U., Winieski, J.A. (1962) lnterspecific hybridization in the genus Abies. *Silvae Genetica*, **11**, 130-142.

Kluge, A.G. (1989) A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Biology*, **38**, 7-25.

Kormutak, A., Demankova, B., Gömöry, D. (2008) Spontaneous hybridization between *Pinus sylvestris* L. and *P. mugo* Turra in Slovakia. *Silvae Genetica*, **57**, 76-82.

Lande, R. (1976) Natural selection and random genetic drift in henotypic evolution. *Evolution,* **30**, 314-334.

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution,* **30**, 314-334.

Laurin, M, Germain, D. (2011) Developmental characters in phylogenetic inference and their absolute timing information. *Systematic Biology*, **60**, 630-644.

Law, J.H. & Regnier, F.E. (1971) Pheromones. *Annual Review of Biochemistry,* **40**, 533-548.

Lecointre, G. & Delaporte, P. (2004) Total evidence requires exclusion of phylogenetically misleading data. *Zoologica Scripta,* **34**, 101-117.

Leopold, E.B. & Denton, M.F. (1987) Comparative age of grassland and steppe east and west of the northern Rocky Mountain. *Annals of the Missouri Botanical Garden,* **74**, 841–867.

Levin, R.A., McDade, L.A., & Raguso, R.A. (2003) The systematic utility of floral and vegetative fragrance in two genera of nyctaginaceae. *Systematic Biology,* **52**, 334-351.

Lin, C.-P., Huang, J.-P., Wu, C.-S., Hsu, C.-Y., Chaw, S.-M. (2010) Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biology and Evolution*, **2**, 504-517.

Lin, W. & Oliver, D.J. (2008) Role of triacylglycerols in leaves. *Plant Science,* **75**, 233-237.

Ling, Y.-R., Humphries, C.J. & Gilbert, M.G. (2008) *Artemisia. Flora of China* at eFloras.org. Published on the Internet http://www.efloras.org [accessed 2008–2011]. Missouri Botanical Garden, St. Louis, MO & Harvard University Herbaria, Cambridge, MA.

Linnaeus, C. (1735) *Species Plantarum*. Laurentius Salvius, Stockholm.

Liu, Z., Wang, C. & Yi, H. (2001) Evolution and mass accumulation of the Cenozoic Hoh Xil Basin, northern Tibet. *Journal of Sedimentary Research,* **71,** 971-984.

Liu, Z., Zhao, X., Wang, C., Liu, S. & Yi, H. (2003) Magnetostratigraphy of Tertiary sediments from the Hoh Xil Basin: Implications for the Cenozoic tectonic history of the Tibetan Plateau. *Geophysical Journal International,* **154**, 233–252.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography,* **17**, 145–151.

Löffler, C., Czygan, F.-C., & Proksch, P. (1997) Phenolic Constituents as Taxonomic Markers in the Genus *Cuscuta* (Cuscutaceae). *Biochemical Systematics and Ecology,* **25**, 297-303.

Lu, Y., Harrington, P.B. (2010) Classification of bacteria by simultaneous methylation—solid phase microextraction and gas chromatography/mass spectrometry analysis of fatty acid methylesters. *Analytical and Bioanalytical Chemistry*, **397**, 2959-2966.

Lutzoni, F., Vilgalys, R. (1994) Integration of morphological and molecular data sets in estimating fungal phylogenies. *Canadian Journal of Botany*, **73**, S649-S659.

Maddison, W.P. & Maddison, D.R. (2011) Mesquite: a modular system for evolutionary analysis.

Maestri, D.M., Guzman, C.A. (1995) A comparative study of seed lipid components of Nicotianeae (Solanaceae). *Biochemical Systematics and Ecology*, **23**, 201-207.

Maffei, M., Meregalli, M., & Scannerini, S. (1997) Chemotaxonomic significance of surface was n-alkanes in the Cactaceae. *Biochemical Systematics and Ecology,* **25**, 241-253.

Maffei, M.E., Gertsch, J., & Appendino, G. (2011) Plant volatiles: production, function and pharmacology. *Natural Products Reports,* **28**, 1359-1380.

Magurran, A.E. (1999) The causes and consequences of geographic variation in antipredator behavior: perspectives from fish populations. In *Geographic variation in behavior: perspectives on evolutionary mechanisms* (eds. Foster, S. A. & Endler, J. A.). Oxford University Press, New York.

Mao K., Milne R.I., Zhang L., Peng Y., Liu J., Thomas P., Mill R.R., Renner S.S. (2012) Distribution of living Cupressaceae reflects the breakup of Pangea. *Proceedings of the National Academy of Science* USA, **109**, 7793-7798.

Markos, S. & Baldwin, B.G. (2001) Higher-level relationships and major lineages of *Lessingia* (Compositae, Astereae) based on nuclear rDNA internal and external transcribed spacer (ITS and ETS) sequences. *Systematic Botany,* **26**, 168–183.

Martín, J., Torrell, M. & Vallés, J. (2001) Palynological features as a systematic marker in *Artemisia* L. and related genera (Asteraceae, Anthemideae). *Plant Biology*, **3**, 372–378**.**

Martín, J., Torrell, M., Korobkov, A.A. & Vallés, J. (2003) Palynological features as a systematic marker in *Artemisia* L. and related genera (Asteraceae, Anthemideae) – II: implications for subtribe Artemisiinae delimitation. *Plant Biology*, **5**, 85–93.

Matthews, S. (2009) Phylogenetic relationships among seed plants: persistent questions and the limits of molecular data. *American Journal of Botany*, **96**, 228-236.

Mayworm, M.A.S., Salatino, A. (2000) Distribution of seed fatty acids and the taxonomy of Vochysiaceae. *Biochemical Systematics and Ecology* **30**, 961-972.

McCune, B. & Mefford, M.J. (2011) PC-ORD. Multivariate analysis of ecological data. MjM Software, Gleneden Beach, OR.

McDougall, I. (1979) Age of shield-building volcanism of Kauai and linear migration of volcanism in the Hawaiian island chain. *Earth and Planetary Science Letters,* **46**, 31–42.

Mclean, S., Boyle, R.R., Brandon, S., Davies, N.W., & Sorensen, J.S. (2007) Pharmacokinetics of 1,8-cineole, a dietary toxin, in the brushtail possum (Trichosurus vulpecula): Significance for feeding. *Xenobiotica,* **37**, 903-922.

Microsoft. 2010. Microsoft Excel. Redmond, Washington: Microsoft. Computer Software.

Mikanagi, Y., Yokoi, M., Ueda, Y., Saito, N. (1995) Flower flavonol and anthocyanin distribution in subgenus *Rosa*. *Biochemical Systematics and Ecology*, **23**, 183-200.

Miller, M.A., Holder, M.T., Vos, R., Midford, P.E., Liebowitz, T., Chan, L., Hoover, P. & Warnow, T. (2009) The CIPRES Portals. CIPRES. 2009-08-04.

**Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1–8.**

Mishler, B.D. (2000) Deep Phylogenetic Relationships among "Plants" and Their Implications for Classification. *Taxon*, **49**, 661-683.

Mishler, B.D. (2005) The logic of the data matrix in phylogenetic analysis. In V.A. Albert (ed.), Parsimony, Phylogeny, and Genomics, pp. 57-70. Oxford: Oxford University Press.

Mithöfer, A. & Boland, W. (2012) Plant defense against herbivores: chemical aspects. *Annual review of plant biology,* **63**, 431-450.

Mitova, M.I., Anchev, M.E., Handjieva, N.V. (2001) Iridoid patterns in *Galium* L. and some phylogenetic considerations. *Zeitschrift für Naturforschung*, **57c**, 226-234.

Mongrand, S., Badoc, A., Patouille, B., Lacomblez, C., Chavent, M., Cassagne, C., & Bessoule, J.-J. (2001) Taxonomy of gymnospermae: multivariate analyses of leaf fatty acid composition. *Phytochemistry,* **58**, 101-115.

Mysore, K. S. & Ryu, C.-M. (2004) Nonhost resistance: how much do we know? *Trends in Plant Science* **9**, 97-104.

Nasri, N., Khaldi, A., Fady, B., & Triki, S. (2005) Fatty acids from seeds of Pinus pinea L.: Composition and population profiling. *Phytochemistry,* **66**, 1729-1735.

Naughton, J.J., Macdonald, G.A., & Greenberg, V.A. (1980) Some additional potassium-argon ages of Hawaiian rocks: The Maui volcanic complex of Molokai, Maui, Lanai and Kahoolawe. *Journal of Volcanology and Geothermal Research,* **7**, 339–355.

Negahban, M. & Moharramipour, S. (2007) Fumigant toxicity of essential oil from *Artemisia sieberi* Besser against three stored-product insects. *Journal of Stored Product Research,* **43**, 123-128.

Nixon, K.C. (1999) The Parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics,* **15**, 407–414.

Nylander, J. A. A. (2004) MrModeltest 2.3. Computer program and documentation distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Ohlrogge, J.B. (1997) Regulation of fatty acid synthesis. *Annual Review of Plant Physiology and Plant Molecular Biology,* **48**, 109-136.

Ortega-Huerta, M.A. & Peterson, A.T. (2008) Modeling ecological niches and predicting geographic distributions: A test of six presence-only methods. *Revista Mexicana de Biodiversidad*, **79**, 205–216.

Pagel, M. (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London* B, **244**, 37-45.

Pagel, M. (1997) Inferring evolutionary processes from phylogenies . *Zoologica Scripta,* **26**, 331-348.

Pagel, M. (1999) Inferring the historical patterns of biological evolution. *Nature,* **401**, 877-884.

Panseri, S., Nanayakkarawasam, N., Nanayakkara, M. C., Mattara, M., & Giorgi, A. (2011) HS-SPME method to monitoring *in vivo* volatile compounds emitted from *Achillea collina* infested by aphids. 280. *International Symposium on Medicinal, Aromatic and Nutraceutical Plants from Mountainous Areas* **955**, 275.

Paradis, E., Claude, J., & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics,* **20**, 289-290.

Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T. (2007) Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography,* **34**, 102–117.

Perez, M.A.B. (2013) Roles of a volatile terpene in plant-herbivore interactions: a foraging odour cue as well as a toxin. *Australian Mammal Society 59th Scientific and Annual Meeting*.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2011) Maxent software for species distribution

modeling. *Ecological Modelling,* **190**, 259.

Poljakov P.P. (1961) *Artemisia* L. *Flora of the U.S.S.R.*, Volume 25 (*English Translation*), (ed. by E. K. Shishkin and E. G. Bobrov), pp. 404–600. Koeltz Scientific Books, Koenigstein, Germany.

Poljakov, P.P. (2007) *Artemisia* L. *Flora of Siberia* (English). Vol. 13. Asteraceae (ed. by I.M. Krasnoborov). N.H. Enfield: Science Publishers.

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

Price, J.P. & Clague, D.A. (2002) How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proceedings of the Royal Society B, Biological Sciences*, **269**, 2429–2435.

Price, J.P. & Clague, D.A. (2002) How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proceedings of the Royal Society B: Biological Sciences,* **269**, 2429-2435.

Price, J.P. & Wagner, W.L. (2004) Speciation in Hawaiian angiosperm lineages: Cause, consequence, and mode. *Evolution,* **58**, 2185–2200.

Printzen, C. (2010) Lichen Systematics: The Role of Morphological and Molecular Data to Reconstruct Phylogenetic Relationships. *Progress in Botany* pp. 233-275.

R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0, URL http://www.Rproject.org.

Rae, T.C. (1998) The logical basis for the use of continuous characters in phylogenetic systematics. *Cladistics*, **14**, 221-228.

Rambaut, A. & Drummond, A.J. (2003) Tracer v1.3. Available from: <http://tree.bio.ed.ac.uk/software/tracer/>.

Rambaut, A. (2008) FigTree v1.3.1. Available from: <http://tree.bio.ed.ac.uk/software/FigTree/>.

Ran, J.-H.*,* Gao H., Wang X.-Q. (2010) Fast evolution of the retroprocessed mitochondrial *rps*3 gene in conifer II and further evidence for the phylogeny of gymnosperms. *Molecular Phylogenetics and Evolution*, **54**, 136-149.

Randell, R. & Morden, C. (1999) Hawaiian Plant DNA Library ll: Endemic, indigenous, and introduced species. *Pacific Science,* **53**, 401–417.

Rattan, R. S. (2010) Mechanism of action of insecticidal secondary metabolites of plant origin. *Crop Protection* **29**, 913-920.

Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by

dispersal, local extinction, and cladogenesis. *Systematic Biology,* **57**, 4–14.

Ree, R.H., Moore, B.R., Webb, C.O. & Donoghue, M.J. (2005) A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution,* **59**, 2299–2311.

Revell, L.J. (2012) An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution,* **3**, 217-223.

Řezanka, T. (1998) Branched and very long-chain dicarboxylic acids from Equisetum species. *Phytochemistry,* **47**, 1539-1543.

Ridley, H. (1930) *The dispersal of plants throughout the world*. L. Reeve & Co., Ashford, England.

Rödl, T., Berger, S., Romero, L.M., & Wikelski, M. (2007) Tameness and stress physiology in a predator-naive island speies confronted with novel predation threat. *Proceedings of the Royal Society B,* **274**, 577-582.

Roeder, S., Hartmann A.-M., Effmert U., Piechulla B. (2007) Regulation of simultaneous synthesis of floral scent terpenoids by the 1,8-cineole synthase of *Nicotiana suaveolens*. *Plant Molecular Biology*, **65**, 107-124.

Roncal, J., Henderson, A., Borchsenius, F., Cardoso, S.R.S., & Balslev, H. (2012) Can phylogenetic signal, character displacement, or random phenotypic drift explain the morphological variation in the genus *Geonoma* (Arecaceae)? *Biological Journal of the Linnean Society,* **106**, 528-539.

Rønsted, N., Göbel, E., Franzyk, H., Jensen, S.R., & Olsen, C.E. (2000) Chemotaxonomy of Plantago. Iridoid glucosides and ca€eoyl phenylethanoid glycosides. *Phytochemistry*, **55**, 348.

Roussis, V., Petrakis, P.V., Ortiz, A., Mazomenos, B.E. (1995) Volatile constituents of needles of five *Pinus* species grown in Greece. *Phytochemistry*, **39**, 357-361.

RStudio (2012) RStudio: Integrated development environment for R. RStudio, Boston.

Sakai, A., Wagner, W., Ferguson, D. & Herbst, D. (1995) Origins of dioecy in the Hawaiian flora. *Ecology,* **76**, 2517-2529.

Sang, T., Crawford, D.J. & Stuessy, T.F. (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany*, **84**, 1120-1136.

Sanz, M., Vilatersana, R., Hidalg, O., Garcia-Jacas, N., Susanna, A., Schneewei, S.S., Gerald, M., & Vallès, J. (2008) Molecular phylogeny and evolution of floral characters of *Artemisia* and allies (Anthemideae, Asteraceae): Evidence from nrDNA ETS and ITS sequences. *Taxon,* **57**, 66-78.

Satyal, P., Paudel, P., Kafle, A., Pokharel, S.K., Lamichhane, B., Dosoky, N.S.; Moriarity, D.M. & Setzer, W.N. (2012) Bioactivities of volatile components from Nepalese *Artemisia* species. Natural Products Communication **7**, 1651-1658.

Sax, D.F. & Gaines, S.D. (2008) Species invasions and extinction: the future of native biodiversity

on islands. *PNAS,* **105**, 11490-11497.

Schaefer, H., Hardy, O.J., Silva, L., Barraclough, T.G., & Savolainen, V. (2011) Testing Darwin's naturalization hypothesis in the Azores. *Ecology Letters,* **doi: 10.1111/j.1461-0248.2011.01600.x**, 1-8.

Schleifer, K.H. (2009) Classification of Bacteria and Archaea: Past, present and future. *Systematic and Applied Microbiology,* **32**, 533-542.

Schols, P., D'hondt, C., Geuten, K., Merckx, V., Janssens, S., & Smets, E. (2004) MorphoCode: coding quantitative data for phylogenetic analysis. *PhyloInformatics,* **4**, 1-4.

Scotland, R.W., Olmstead, R.G., & Bennett, J.R. (2003) Phylogeny reconstruction: the role of morphology. *Systematic Biology,* **52**, 539-548.

Shaffer, S.A., Tremblay, Y., Weimerskirch, H., Scott, D., Thompson, D.R., Sagar, P.M., Moller, H., Taylor, G.A., Foley, D.G., Block, B.A. & Costa, D.P. (2006) Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. *Proceedings of the National Academy of Sciences USA*, **103**, 12799-12802.

Sharkey, T.D., Yeh, S. Wiberley, A.E., Falbel, T.G., Gong, D.M., Fernandez, D.E. (2005) Evolution of the isoprene biosynthetic pathway in kudzu. *Plant Physiology*, **137**, 700-712.

Shultz, L. (2009) *Monograph of* Artemisia *Subgenus* Tridentatae *(Asteraceae–Anthemideae)*. *Systematic Botany Monographs*, **89**, 1–129.

Shultz, L. M. (1999) *Artemisia* L. Wormwood. *Manual of the flowering plants of Hawai`i* (ed. by W. L. Wagner, D. R. Herbst, and S. H. Sohmer), pp. 262–265. University of Hawai`i Press, Bishop Museum Press, Honolulu.

Shultz, L. M. (2006) *Artemisia*. *The Flora of North America North of Mexico, Vol. 19: Asterales, part 1* (ed. by Flora of North America Editorial Committee), pp. 503–534. Oxford University Press, New York.

Silvestro, D. & Michalak, I. *(*2010) raxmlGUI: A graphical front-end for RAxML*. Available at: http://sourceforge.net/projects/raxmlgui/.*

Silvestro, D. & Michalak, I. (2012) raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution,* **12**, 335-337.

Simmons, M. P. (2004) Independence of alignment and tree search. *Molecular Phylogenetics and Evolution*, **31,** 874–879.

Simmons, M.P. (2004) Independence of alignment and tree search. *Molecular Phylogenetics and Evolution*, **31**, 874-879.

Sinclair, A. R. E., Madhu, K. J., & Andersen, R. J. (1988) Camphor from juvenile white spruce as an antifeedant for snowshoe hares. *Journal of Chemical Ecology* **14**, 1505-1514.

Skaltsa, H.D., Demetzos, C., Lazari, D., Sokovic, M. (2003) Essential oil analysis and antimicrobial activity of 8 *Stachys* species from Greece. *Phytochemistry*, **64**,743-752.

Skottsberg, C. (1925) Juan Fernandez and Hawaii. *Bernice P. Bishop Museum Bulletin*, **16**, 1–47.

Skottsberg, C. (1927) *Artemisia, Scaevola, Santalum,* and *Vaccinium* of Hawaii. *Bernice P. Bishop Museum Bulletin*, **43**, 1–89.

Skottsberg, C. (1931) Remarks on the flora of the high Hawaiian volcanoes. *Acta Horti Gothoburgensis,* **1**, 47–65.

Smith, K.L., Harmon, L.J., Shoo, L.P., & Melville, J. (2011) Evidence of constrained phenotypic evolution in a cryptic species complex of Agamid lizards. *Evolution,* **65**, 976-992.

Smith, K.L., Harmon, L.J., Shoo, L.P., & Melville, J. (2011) Evidence of constrained phenotypic evolution in a cryptic species complex of Agamid lizards. *Evolution,* **65**, 976-992.

Sokovic, M., Tzakou, O., Pitarokili, D., & Couladis, M. (2002) Antifungal activities of selected aromatic plants growing wild in Greece. *Molecular Nutrition and Food Research* **46**, 317-320.

Song, Z.-C., Wang, W.-M., & Huang, F. (2004) Fossil pollen records of extant angiosperms in China. *The Botanical Review,* **70**, 425–458.

Stadler, M., Rournier, J., Gardt, S., & Peršoh, D. (2010) The phylogenetic positionof *Rhopalostroma* as inferred from a polythetic approach. *Persoonia,* **25**, 11-21.

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

Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology,* **75**, 758–771.

Steadman D.W. (1999) Prehistoric extinctions of Pacific island birds: biodiversity meets zooarchaeology. *Science* **267**, 1123-1131.

Steele, C.L., Crock, J, Bohlmann, J., Croteau, R. (1998) Sesquiterpene synthases from grand fir (*Abies grandis*): comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of δ-selinene synthase and γ-humulene synthase. Journal of Biological Chemistry, **273**, 2078-2089.

Stevens, J.F., Hart, E.T.E., & Bolck, A. (1996) Flavonoid variation in Eurasian *Sedum* and *Sempervivum*. *Phytochemistry,* **41**, 503-512.

Swofford, D.L. (2003) Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Syring, J., Farrell, K., Businský, R., Cronn, R., & Liston, A. (2007) Widespread Genealogical Nonmonophyly in Species of Pinus Subgenus Strobus. *Systematic Biology,* **56**, 163-181.

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

Taggart, R. E. & Cross, T. (1980) Vegetation change in the Miocene Succor Creek flora of Oregon and Idaho: A case study in paleosuccession. *Biostratigraphy of Fossil Plants- Successional and Paleoecological Analyses* (ed. by D. Dilcher and T.M. Taylor), pp. 185–210. Dowden, Hutchinson & Ross, Inc., Stroudsburg, PA.

Tate, J.A. & Simpson, B.B. (2003) Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany,* **28**, 723–737.

Thiele, K. (1993) The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics,* **9**, 275-304.

Tkach, N.V., Hoffmann, M.H., Roeser, M., Korobkov, A.A. & von Hagen, K.B. (2008a) Parallel evolutionary patterns in multiple lineages of arctic *Artemisia* L. (Asteraceae). *Evolution,* **62**, 184–198.

Tkach, N.V., Hoffmann, M.H., Röser, M. & von Hagen, K.B. (2008b) Temporal patterns of evolution in the Arctic explored in *Artemisia* L. (Asteraceae) lineages of different age. *Plant Ecology & Diversity*, **1**, 161–169.

Tutin, T. G., Persson, K., & Gutermann, W. (1976) *Artemisia*. *Flora Europaea* (ed. by T. G. Tutin, V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters, and D. A. Webb), pp. 178–186. Cambridge University Press, London.

Umek, A., Kreft, S., Kartnig, T., & Heydel, B. (1999) Quantitative Phytochemical Analyses of Six Hypericum Species Growing in Slovenia. *Planta Medica,* **65**, 388-390.

Urbatsch, L. E., B. G. Baldwin, and M. J. Donoghue (2000) Phylogeny of the coneflowers and relatives (Heliantheae: Asteraceae) based on nuclear rDNA internal transcribed spacer (ITS) sequences and chloroplast DNA restriction site data. *Systematic Botany*, **25**, 539–565.

Valant-Vetschera, K.M., Bhutia, T.D., & Wollenweber, E. (2010) Chemodiversity of exudate flavonoids in Dionysia (Primulaceae): A comparative study. *Phytochemistry,* **71**, 937-947.

Vallès, J. & McArthur, E.D. (2001) *Artemisia* systematics and phylogeny: Cytogenetic and molecular insights. *USDA Forest Service Proceedings,* **RMRS-P-21**, 67–74.

Vallès, J., Torrell, M., Garnatje, T., Garcia-Jacas, N., Vilatersana, R. & Susanna, A. (2003) The genus *Artemisia* and its allies: Phylogeny of the subtribe Artemisiinae (Asteraceae, Anthemideae) based on nucleotide sequences of nuclear ribosomal DNA internal transcribed spacers (ITS). *Plant Biology (Stuttgart),* **5**, 274–284.

Velasco, L., Goffman, F.D. (1999) Chemotaxonomic significance of fatty acids and tocopherols in Boraginaceae. *Phytochemistry*, **52**, 423-426.

Verdeguer, M., Blázquez, M.A., & Boira, H. (2009) Phytotoxic effects of *Lantana camara*, *Eucalyptus camaldulensis*, and *Eriocephalus africanus* essential oils in weeds of the Mediterranean summer crops. *Biochemical Systematics and Ecology,* **37**, 362-369.

Vitousek, M. N., Romero, L. M., Tarlow, E., Cyr, N. E., & Wikelski, M. (2010) Island tameness: an altered cardiovascular stress response in Galápagos marine iguanas. *Physiology Behavior* **99**, 544-548.

Vourc'h, G., Russell, J., & Martin, J.-L. (2002) Linking deer browing and terpene production among genetic identities in *Chamaecyparis nootkatensis* and *Thuja plicata* (Cupressaceae). *Journal of Heredity,* **93**, 370-376.

Vynne, N.G., Månsson, M., Nielsen, K.F., & Gram, L. (2011) Bioactivity, Chemical Profiling, and 16S rRNA-Based Phylogeny of Pseudoalteromonas Strains Collected on a Global Research Cruise. *Marine Biotechnology,* **13**, 1062-1073.

Wagenmakers, E.-J. & Farrell, S. (2004) AIC model selection using Akaike weights. *Psychonomic Bulletin and Review,* 11, 192-196.

Wagner, W. L., Herbst, D. R., & Sohmer, S. H. (1999) *Manual of the flowering plants of Hawai`i*, Revised edn. University of Hawai`i Press, Bishop Museum Press, Honolulu.

Wagner, W.L., Herbst, D.R. & Lorence, D.H. (2005-). *Flora of the Hawaiian Islands website*. http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/index.htm [accessed November 2011].

Wallace, A.R. (1881) *Island life: or, the phenomena and causes of insular faunas and floras, including a revision and attempted solution of the problem of geological climates*. Harper, New York.

Walters, D. (2011) *Plant Defense: Warding off attack by pathogens, herbivores and parasitic plants*. Wiley-Blackwell, Chichester, West Sussex, UK.

Wang, W.-M. (2004) On the origin and development of *Artemisia* (Asteraceae) in the geological past. *Botanical Journal of the Linnean Society,* **145**, 331–336.

Wang, W.P., Hwang, C.Y., Lin, T.P., Hwang, S.Y. (2003) Historical biogeography and phylogenetic relationships of the genus *Chamaecyparis* (Cupressaceae) inferred from chloroplast DNA polymorphism. *Plant Systematics and Evolution*, 241, 13-28.

Watson, L.E., Bates, P.L., Evans, T.M., Unwin, M.M. & Estes, J.R. (2002) Molecular phylogeny of Subtribe Artemisiinae (Asteraceae), including *Artemisia* and its allied and segregate genera. *BMC Evolutionary Biology,* **2**, 17.

Wen, J. & Ickert-Bond, S.M. (2009) Evolution of the Madrean-Tethyan disjunctions and the North and South American amphitropical disjunctions in plants. *Journal of Systematics and Evolution*, **47**, 331–348.

White, T. J., Bruns, T., Lee, S. & Taylor, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: A guide to methods and applications* (ed. by M. Innis, D. Gelfand, J. Sninsky, and T. White), pp. 315–322. Academic Press, San Diego.

Wiens, J.J. (2001) Character analysis in morphological phylogenetics: problems and solutions. *Systematic Biology*, 50, 689-699.

Wiens, J.J. (2004) The role of morphological data in phylogeny reconstruction. *Systematic Biology,* 53, 653-661.

Wiens, J.J., Chippindale, P.T., & Hillis, D.M. (2003) When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders. *Systematic Biology,* 52, 501-514.

Wilgenbusch, J. C., Warren, D. L. & Swofford, D. L. (2004) AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference.

http://ceb.csit.fsu.edu/awty [Accessed November, 2010].

Williams, N.H. & Whitten, W.M. (1999) Molecular phylogeny and floral fragrances of male euglossine bee-pollinated orchids: A study of *Stanhopea* (Orchidaceae). *Plant Species Biology,* 14, 129-136.

Wink, M. (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*, **64**, 3-19.

Wink, M., Botschen, F., Gosmann, C., Schäfer, H., & Waterman, P.G. (2010) Chemotaxonomy Seen from a Phylogenetic Perspective and Evolution of Secondary Metabolism. *Annual Plant Reviews,* **40**, 364-433.

Wise, M.L., Savage, T.J., Katahira, E., Croteau, R. (1998) Monoterpene synthases from common sage (*Salvia officinalis*). *Journal of Biological Chemistry*, **273**, 14891-14899.

Wittstock, U. & Gershenzon, J. (2002) Constitutive plant toxins and their role in defense against herbivores and pathogens. *Current Opinion in Plant Biology,* **5**.

Wolff, R.L., Comps, B., Marpeau, A.M., & Deluc, L.G. (1997a) Taxonomy of Pinus species based on the seed oil fatty acid compositions. *Trees,* **12**, 113-118.

Wolff, R.L., Deluc, L.G., & Marpeau, A.M. (1996) Conifer seeds: oil content and fatty acid composition. *Journal of the American Oil Chemists Society,* **73**, 765-771.

Wolff, R.L., Deluc, L.G., Marpeau, A.M., & Comps, B. (1997b) Chemotaxonomic differentiation of conifer families and genera based on the seed oil fatty acid compositions: multivariate analyses. *Trees,* **12**, 57-65.

Wolff, R.L., Lavialle, O., Pedrono, F., Pasquier, E., Deluc, L.G., Marpeau, A.M., & Aitzetmuller, K. (2001) Fatty acid composition of Pinaceae as taxonomic markers. *Lipids,* **36**, 439-451.

Wolff, R.L., Pédrono, F., Marpeau, A.M., Christie, W.W., & Gunstone, F.D. (1998) The seed fatty acid composition and the distribution of 5-olefinic acids in the triacylglycerols of some Taxaceae (Taxus and Torreya). *Journal of the American Oil Chemists Society,* **75**, 1637-1641.

Wollenweber, E., Stevens, J.F., Dörr, M., & Rozefelds, A.C. (2003) Taxonomic significance of flavonoid variation in temperate species of Nothofagus. *Phytochemistry,* **62**, 1125-1131.

Won, H., Renner, S.S. (2003) Horizontal gene transfer from flowering plants to *Gnetum*. *Proceedings of the National Academy of Science*. USA, **100**, 10824-10829.

Wu, C.-H., Chaw, S.-M., Huang, Y.-Y. (2013) Chloroplast phylogenomics indicates *Ginkgo biloba* is sister to cycads. *Genome Biology and Evolution*, **5**, 243-254.

Wu, J. & Baldwin, I. T. (2010) New insights into plant responses to the attack from insect herbivores. *Annual Review of Genetics* **44**, 1-24.

Yang, Z.-Y., Ran, J.-H., & Wang, X.-Q. (2012) Three genome-based phylogeny of Cupressaceae s.l.: Further evidence for the evolution of gymnosperms and Southern Hemisphere biogeography. *Molecular Phylogenetics and Evolution,* **64**, 452-470.

Ydenberg, R.C. & Dill, L.M. (1986) The economics of fleeing from predators. In *Advances in the Study of Behavior, v. **16*** (eds. Rosenbladd, J. S. & Bok, H.). Academic Press, Orlando.

Yoshimura, H., Sawai, Y., Tamotsu, S., & Sakai, A. (2011) 1,8-cineole inhibits both proliferation and elongation of BY-2 cultured tobacco cells. *Journal of Chemical Ecology,* **37**, 320-328.

Zaklinskaya, E.D. & Leopold, E.B. (1957) Outline of the geology and paleogeography of the Pavlodar section of the Irtysh Valley, the northern Aral Region and the Turgai Lowland (English). *International Geology Review*, **4**, 310–335.

Zgurski, J.M., Rai, H.S., Fai, Q.M., Bogler, D.J., Francisco-Ortega, J. (2008) How well do we understand the overall backbone of cycad phylogeny? New insights from a large, multigene plastid data set. Molecular Phylogenetics and Evolution, **47**, 1232-1237.

Zhang, L. & Sun, J. (2006) Electrophysiological and Behavioral Responses of *Dendroctonus valens* (Coleoptera: Curculionidae: Scolytinae) to Candidate Pheromone Components Identified in Hindgut Extracts. *Environmental Entomology,* **35**, 1232-1237.

Zheng, L., Ives, A.R., Garland Jr., T., Larget, B.R., Yu, Y., & Cao, K. (2009) New multivariate tests for phylogenetic signal and trait correlations applied to cophysiological phenotypes of nine Manglietia species. *Functional Ecology,* **23**, 1059-1069.

Zhong, B., Yonezawa, T., Zhon, Y., Hasegawa, M. (2010) The position of Gnetales among seed plants: overcoming pitfalls of chloroplast phylogenomics. Mol. Biol. Evol., **27**, 2855-2863.

Zhu, Z., Wu, L., Song, P. & Zhang, Y. (1985) *Research on Tertiary palynology from the Qaidam Basin*. Petroleum Industry Press, Beijing.