Novel nuclear markers inform the systematics and the evolution of serpentine use in Streptanthus and allies (Thelypodieae, Brassicaceae)

N. Ivalú Cacho a,*, A. Millie Burrell b,†, Alan E. Pepper b, Sharon Y. Strauss a

a Center for Population Biology, and Department of Evolution of Ecology, University of California, Davis, CA 95616, United States
b Department of Biology, Texas A&M University, College Station, TX 77843-3258, United States

1. Introduction

In the last decade, great advances have been made in identifying main lineages in Brassicaceae and their evolutionary relationships (Al-Shehbaz et al., 2006). However, despite containing some of the best-studied model systems, the single best-annotated plant genome, and many important crops, the phylogenetic relationships within some tribes of Brassicaceae remain elusive. Intrinsic factors such as rapid radiations, whole genome duplications, and introgression events have contributed to the challenges identifying relationships in this family (Warwick et al., 2009; Fränzke et al., 2011). This lack of resolution at shallow phylogenetic levels in some groups of the Brassicaceae, as with plants in general, has restricted investigations addressing fundamental questions in plant evolutionary biology (Hughes et al., 2006). The Thelypodieae tribe, which contains taxa with some of the most unusual floral adaptations within the family, has remained among the most difficult tribes from the standpoint of resolving phylogenetic relationships (Warwick et al., 2009).

Streptanthus is a genus of ca. 35 species in the tribe Thelypodieae (Brassicaceae) that has remarkable morphological and ecological diversity, a large number of species in the group being edaphic specialists endemic to unusual soils such as serpentine. While ecological research has shed some light on adaptation to serpentine in Streptanthus, there have been few insights on the origins and evolution of serpentine tolerance in this group, largely due to limited success in resolving the phylogenetic relationships among Streptanthus and allied genera of the Thelypodieae (Streptanthoid complex).

We present a well-resolved phylogenetic hypothesis for the Streptanthoid complex, based on three newly identified and highly variable single copy nuclear regions (AT4G34700, AT1G61620, and AT1G56590, and three others that are widely used (ITS, phyA, and PEPC). We also include data for two chloroplast regions (trnl and trnH-psbA). Collectively, our new markers provide 75% of the nuclear parsimony informative characters in our data. Taxonomically, our sampling is the most inclusive of any study of the Streptanthoid Complex to date, including 46 out of the 53 species of Streptanthus and Caulanthus, as well as representatives of several closely allied genera in the Thelypodieae.

Our results reveal that Streptanthus, Caulanthus, and Thelypodium are not reciprocally monophyletic as currently defined. The species of Streptanthus form two rather distantly related clades. One clade (SC-I) is comprised of species with bilateral flowers and urn-shaped calyces that occur mainly within the California Floristic Province (CFP) hotspot; the other clade (SC-II) is composed of species with extant ranges mainly outside the CFP. Our data indicate that serpentine tolerance has evolved between eight and ten times in this group, of which between four and five have resulted in endemism. While serpentine endemism has been rarely lost, large and diverse clades composed mainly of serpentine endemics indicate that serpentine endemics in this group are more than mere ‘dead-ends’.

© 2014 Elsevier Inc. All rights reserved.
several authors clearly recognize that while the tribe appears to be monophyletic, the taxonomic delimitation of species and genera in this complex is still debated (Al-Shehbaz et al., 2006; Warwick et al., 2009; Burrell et al., 2011). This situation has led some authors to treat a set of ca. 60 taxa within the Thelypodieae as the 'Streptanthoid Complex' (Burrell and Pepper, 2006; Burrell, 2010; Burrell et al., 2011), based on affinity with the North American genus *Streptanthus* and a largely conserved chromosome number of \( n = 14 \) (Pepper and Norwood, 2001; Warwick et al., 2002, 2009; Burrell et al., 2011).

The center of diversity of the Streptanthoid Complex is southwestern North America (Al-Shehbaz, 2010, 2012). In California, members of the complex occur in all three California desert regions (Mojave, Sonoran COLORADO and Great Basin) and all major mountain ranges and foothills. Beyond California, some species within the Streptanthoid Complex extend north to Oregon, east to Louisiana, and south to Baja California, Mexico (Al-Shehbaz, 2010). While some species in the complex have extremely narrow ranges (endemic to a few or even a single serpentine outcrop, e.g., *Streptanthus vernalis*), others have wide ranges that span several states (e.g., *S. hyacinthoides, S. cordatus*).

Ecologically, *Streptanthus* exhibits a high degree of edaphic specialization (Fig. 1), particularly associated with ultramafic or serpentine soil: about 30% of the recognized species are restricted to serpentine, hereafter referred to as serpentine endemics (Safford et al., 2005; Al-Shehbaz, 2010; Baldwin et al., 2012). A few species have populations growing on serpentine or non-serpentine soils (Kruckeberg, 1984, 2006), while others appear incapable of growing in ultramafic soils when assayed (Kruckeberg, 1957) (Strauss and Cacho, unpublished data). These latter species occur on a variety of other substrates like gabbro, granite, chert, and basalt, typically on exposed, dry, rocky, loose-soiled slopes (Baldwin et al., 2012). Also, several southern species are associated with limestone [e.g., *S. bracteatus, S. platycarpus; (Al-Shehbaz, 2010)] and sandy substrates.

![Fig. 1. Floral and ecological diversity in the Streptanthoid Complex. From top to bottom, left to right: Caulanthus coulteri; Streptanthus farnsworthianus; C. flavescens; S. glandulosus; C. amplexicaulis var. barbara; S. polygaloides; C. major; S. barbiger; C. hallii; S. tortuosus; C. glaucus; S. fenestratus; C. cordatus; S. hesperidis; C. anceps; C. inflatus; S. callistus; S. hispidus; serpentine outcrop home of S. breweri; granite outcrop habitat of S. diversifolius; scree where C. amplexicaulis var. amplexicaulis occurs; basalt mesa inhabited by S. tortuosus and S. diversifolius; rocks in sandy desert washes home to C. hallii and C. cooperi; and rocky outcrop home to S. tortuosus.](image-url)
Serpentine soils represent highly stressful environments for plants owing to low Ca:Mg ratios, high concentrations of heavy metals and other toxic elements (e.g., Cr, Cd, Co, Ni), low nutrient availability, and poor water retention (Brady et al., 2005; Kruckenberg, 2006). The ability to tolerate and become endemic to serpentine soils has long fascinated plant evolutionary ecologists (Morrison, 1941; Stebbins, 1942; Walker, 1948; Kruckenberg, 1957, 1986). In California, serpentine soils harbor a disproportionately high number of endemics relative to the land area they occupy and thus contribute to species diversity from local to regional scales (Harrison, 1999; Harrison and Inouye, 2002).

Ecological specialization has been thought to have resulted in evolutionary ‘dead ends’ (Simpson, 1953) but see (Nosil and Mooers, 2005). Consistent with this idea, an analysis of 23 Californian genera containing at least one serpentine endemic, and for which phylogenies were available, found a low number of transitions out of the serpentine endemic state (Anacker et al., 2011). The wide variation in the degree of edaphic adaptation observed in Streptanthus makes this an ideal group with which to study the evolution of serpentine specialization (Kruckenberg, 1984). However, efforts to carry out such studies in Streptanthus have been hampered by the lack of phylogenetic resolution within the Thelypodieae. One of the hurdles faced in elucidating relationships among these taxa is that gene regions that are typically informative for phylogenetic inference at this level (e.g., ITS, matK, trnLF, trnH) are not sufficiently informative in the tribe Thelypodieae (Warwick et al., 2010). In this study, we use nuclear and chloroplast gene regions, including three nuclear regions that we identified de novo, to develop a much-refined phylogeny of the Streptanthoid Complex, and we use this phylogeny as a framework to explore the acquisition and loss of serpentine endemism and tolerance in this alliance.

2. Materials and methods

2.1. Sampling and outgroups

For the taxa that are native to California, we follow the Index to California Plant Names (http://ucjeps.berkeley.edu/about_ICPN.html) for species nomenclature and generic delimitation, and for the rest, we follow the Flora of North America (Al-Shehbaz, 2010). We sequenced a total of 143 accessions representing 46 of the 53 species of Streptanthus and Caulanthus (including Guillenia), as well as other Californian Thelypodieae that have been hypothesized to be closely allied with the Streptanthoid Complex, or whose ranges overlap with species in the complex, such as Thelypodium, Stanleya, Streptanthella and Sibaropsis (Boyd, 1997; Al-Shehbaz, 2010; Baldwin et al., 2012). When possible, we sampled several populations per species and included the closely allied genus Thysonocerus as an outgroup (Warwick et al., 2008). Also, we included Sisymbrium accessions from California and Spain as more distant outgroups (Warwick et al., 2009; Beilstein et al., 2010). Locality and voucher information is provided in Appendix 1.

2.2. Identification of single copy nuclear genes and primer design

To identify several novel nuclear genomic regions that would present useful amounts of phylogenetically informative variation, we employed a combined approach utilizing previous studies, published ESTs, and low depth coverage Illumina genome sequencing [described in detail in (Cacho and Strauss, 2013)]. Briefly, we first cross-referenced the results of three previous studies that implemented algorithmic approaches to identifying putative single-copy nuclear genes (SCNGs; (Wu et al., 2006; Yuan et al., 2009; Duarte et al., 2010) with published comparative EST data from Brassica napus and A. thaliana (Ilut and Doyle, 2012). We targeted genes that were hypothesized to be SCNGs by multiple of these sources, and verified that they were present as a single copy in the most recently annotated A. thaliana genome assembly (TAIR v.10). We then guided our marker selection for primer design as follows (Rodriguez et al., 2009), using a subset of about 14 taxa spanning members of several hypothesized clades and Sisymbrium, a distant outgroup: (1) we selected regions from all five A. thaliana linkage groups in order to minimize the potential for linkage; (2) we designed primers targeting an amplicon length of 600–1200 bp that would amplify readily from genomic DNA and could be sequenced with a single pass of Sanger sequencing, and; (3) we targeted regions with 40–60% intron content in order to provide a balanced mix of more conserved exonic sequences and less conserved intronic sequences.

 Reads from low depth coverage Illumina sequencing of genomic DNA from two Brassica species (B. rapa, B. oleracea; L. Comai, unpublished data) were mapped onto A. thaliana’s genome TAIR v.10 using Burrows-Wheeler Aligner (as implemented by Li and Durbin, 2009) and visualized in IGV v. 1.5 (Thorvaldsdottir et al., 2012). We designed primers manually based on visual inspection of individual alignments for the selected SCNGs (see Cacho and Strauss, 2013 for more details). Multiple primer pairs for each selected region were tested for single band amplification in the subset of 14 taxa and sequenced. Regions with high amplification and sequencing success were selected. Here we present results derived from sequencing three newly identified SCNGs: AT4G34700, AT1G56390, and AT1G61620 (which we refer to as G34, G56, and G61 for brevity), in combination with three traditionally used nuclear regions (phyA, ITS, PEFC) and two chloroplast regions (trnl, trnH-psbA).

2.3. DNA extraction, amplification, and sequencing

We extracted DNA from silica-dried plant material and herbarium specimens (Appendix 1) using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA). PCR primers for ITS, PEFC and phyA have been described previously (White et al., 1990; Olson, 2002; Beilstein et al., 2008). Primers “CP-C” and “CP-D” (Taberlet et al., 1991) were used to amplify the ~0.5 kb intron in the plastid trnl gene. The plastid trnH-psbA intergenic spacer region was amplified using primers described elsewhere (Kress et al., 2005). We optimized PCR conditions for the different genomic regions as specified in Table 1, and sent PCR amplicons for Sanger bidirectional sequencing to Beckman Coulter Genomics (http://www.beckmangenomics.com). When more than one band amplified, we isolated bands, reamplified, and sequenced directly. In a few cases in which cloning was necessary, we gel-purified PCR products (QIAquick Gel Extraction Kit, QIAGEN), ligated them into pGEM T-Vector (Promega Corporation), cloned into E. coli DHB-5x-competent cells (Invitrogen, Carlsbad, California, USA), reamplified (eight colonies per PCR product), and sent to be sequenced, as above.

2.4. Phylogenetic analyses

We edited and assembled sequences in Sequencher v. 4.7 (Gene Codes, Ann Arbor, Michigan, USA). For those cases in which we cloned, we determined alleles by visual inspection of sequences. Consistent sequence variation among multiple clones was considered true polymorphism, and variation that was rare and non-repeated was treated as PCR error. We aligned manually in MacClade 4.08 (Maddison and Maddison, 2005), giving preference to transitions over transversions and treating gaps as missing data; we did not code indels.

We analyzed each dataset separately and in combination, following a total evidence approach (Kluge, 1989). To minimize artifacts that could result from high proportions of missing data
in our combined analyses, we decided to include only those accessions for which data was available for a minimum of four partitions. We employed a conservative approach by selecting the set of alleles that yielded the shortest trees in individual analyses. When conflicting allele relationships were recovered, we kept both alleles in subsequent analyses. For our final allele selection for combined analyses we gave preference to those that yielded the shortest, least resolved tree. To accommodate sequence rate heterogeneity across partitions, we implemented a partition scheme by region in our analyses.

We selected a model of molecular evolution based on the Akaike information criterion (AIC; (Akaike, 1974)) as implemented in the program ModelTest v3.7 (Posada and Crandall, 1998). We only considered models that account for site-to-site rate heterogeneity using a discrete approximation to a gamma distribution of rates (Gamma; (Stamatakis, 2006; Stamatakis et al., 2008)). When implementing the selected model was not possible, we used the closest more general model that was implementable. We conducted all of our analyses on the Cyberinfrastructure for Phylogenetic Research cluster (CIPRES; (Miller, 2009)) at the San Diego Supercomputer Center (http://www.phylo.org/).

We performed our Maximum Likelihood (ML) analyses in RAxML under the GTR+G model of evolution in a partitioned dataset. Analyses in a Bayesian framework included Metropolis coupled Markov chain Monte Carlo (MCMCMC) tree sampling (Larget and Simon, 1999; Mau et al., 1999) as implemented in the programs MrBayes v. 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) and BEAST v.1.7.4 (Drummond et al., 2012).

In MrBayes, we analyzed each of the partitions under its best fitting model, linking only the topology and branch lengths across partitions (unlinked parameters: tratio, revmat, statefreq, and shape). Two independent analyses of two runs each were performed, each with 4 chains, 20 million generations with a sampling frequency of 2000 generations, and a heating parameter of 0.05. We adjusted heat and parameters in an iterative process based on evaluation of mixing guided by split frequencies among runs, the moves in the “cold” chain for both runs within a single analysis (until we achieved acceptance rates of 10–60%), and chain swap information (until rates of 0.1–0.7 were achieved) (Huelsenbeck and Ronquist, 2001). Based on generation-by-likelihood plots, we discarded 10–25% of the samples as burn-in.

For both MrBayes and BEAST analyses, we examined the behavior of the MCMC chains using Tracer v.1.5 (Drummond and Rambaut, 2007) to ensure an ESS > 200 and an even sampling of the posterior and different parameters, including the tree likelihoods of different partitions. We tested the performance of the MCMC analyses across a wide variety of conditions and parameters, including sampling (taxa and regions included in the analyses), partition scheme (including codon positions for coding regions), model of evolution, model of speciation, and chain exchange behavior (narrow vs. wide). Eliminating “rogue” taxa has been shown to improve performance of phylogenetic analyses (Thomson and Shaffer, 2010; Aberer et al., 2013), and we observed that eliminating between three (MrBayes) and five (BEAST) accessions out of 143 produced the most reliable behavior of the analyses (as examined in Tracer v.1.5). Thus our final combined dataset consisted of a total of 138 accessions. The five accessions eliminated correspond to one *C. flavescens* (*e130*), one *C. coulteri* (*e080.a1*), one *S. hyacinthoides* (*e134*), one *Sisymbrium officinale* (*e033a*), and one *S. albidas* (*e081a*). Our complete sampling for individual genes, including rogues, is presented in the individual gene trees in Figs. S2–S9, and the five that were removed are highlighted in the Appendix 1 as well.

Our preliminary analyses in BEAST consisted of 20 million generation runs, based on which we optimized the operator settings iteratively to achieve a stable and reliable behavior of the MCMC chains and adequate ESS, examined in Tracer v.1.5 as above. We unlinked the substitution models across all partitions, and used an uncorrelated estimated lognormal relaxed clock. We used a Yule model as a prior of speciation and an exponential ucld mean of 10. We modeled the tree root height as a truncated and bounded (low = 0, upper = 35) normal, with a mean and initial value of 20 and a standard deviation of 4 (Beilstein et al., 2010). We used a UPGMA starting tree and did not enforce any topological constraints in our analyses. Finally, we also ran BEAST using an empty data set to assess the influence of our priors on the posterior distribution (Drummond et al., 2012).

Our final BEAST analyses consisted of three 50-million-generation independent runs with sampling every 5000 generations, each of which resulted in 10,000 samples of the posterior. We verified stable behavior of the MCMC chain as above, combined the post-burnin trees from stable runs with the program LogCombiner v1.7.4 (Drummond et al., 2012), and created a maximum clade credibility summary tree (MCCT; using median node heights) using TreeAnnotator v1.7.4 (Drummond et al., 2012). We used FigTree v1.4 (2006–2012, Andrew Rambaut, Institute of Evolutionary Biology, University of Edinburgh, UK) to visualize our posterior and summary trees (MCCT and Bayesian consensus).

### 2.5. Evolution of association with serpentine

We explored the evolution of serpentine tolerance and serpentine endemism using the phylogenetic hypothesis for the group generated from our BEAST analyses. A species was classified as tolerant of serpentine if at least one population has been found growing on serpentine substrates. Likewise, a species was considered a serpentine endemic if all of its recorded populations occurred only on serpentine substrates. Tolerant and endemic status was based on herbarium records and floristic treatments in the Flora of North America (Al-Shehbaz, 2010) and the Jepson Manual (Baldwin et al., 2012), along with personal observations.

### Table 1

<table>
<thead>
<tr>
<th>Region</th>
<th>Primers</th>
<th>Cycling conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-56 (AT1G56590)</td>
<td>F: AARGAYATTCTCATCATCTGTCTGAGR: GCATCCATCYTCTTCACAAGTGTCG</td>
<td>(94°C, 0:30; 55°C, 1:10; 72°C, 2:00) 35X</td>
</tr>
<tr>
<td>G-61 (AT1G61620)</td>
<td>F: GCAAGACACTGCAAGAACAA: AGGTGTGTCACACCAAGACTT</td>
<td>(94°C, 0:30; 60°C, 1:10; 72°C, 2:00) 35X</td>
</tr>
<tr>
<td>G-34 (AT4G34700)</td>
<td>F: GAATCTCAATGTCAACCAAGATG: ACCATGAGCAAATGTTGT</td>
<td>(94°C, 0:30; 63°C, 4:00) 30–34X</td>
</tr>
<tr>
<td>phylA, and PEPC</td>
<td>Beilstein et al. (2008) and Olson (2002)</td>
<td></td>
</tr>
<tr>
<td>ITS</td>
<td>White et al. (1990)</td>
<td></td>
</tr>
<tr>
<td>trnL, trnH-psbA</td>
<td>Taberlet et al. (1991) and Kress et al. (2005)</td>
<td></td>
</tr>
</tbody>
</table>
To reconstruct the evolution of serpentine endemism and tolerance in *Streptanthus* and its relatives we selected a single accession per species. We selected accessions that were homoygous at the marker loci, had the fewest missing data, and yielded the shortest topology. In the two instances where our data revealed some lack of cohesiveness, *S. morrisonii* and *C. lasiophyllus*, we kept the accessions that yielded the shortest tree. Because neither of these two cases spanned a transition across soil types, our selection was unlikely to have altered the patterns we observed. Our final dataset for this analyses consisted of 51 tips. For the main purposes of our study, we considered the Glandulosus complex (SG) as a single evolutionary lineage that exhibits tolerance but not endemism to serpentine. However, because this complex is composed of several closely related species that lack reciprocal monophyly with respect to *S. glandulosus*, and because there is considerable variation of soil use within this clade, we also explored the evolution of serpentine associations under a scenario in which we expanded the SG complex from one lineage to nine (keeping one accession for every lineage within the SG that is reciprocally monophyletic with previously recognized species within the SG clade). This dataset was composed of 59 tips.

We used Bayesian stochastic character mapping ((Huelsenbeck et al., 2003) and references therein) to infer the number of transitions separately for serpentine tolerance and for serpentine endemism, while accounting for uncertainty in tree topology, branch lengths, and the character mapping. Analyses were implemented in the program SIMMAP v1.5 (Bollback, 2006) using a random sample of 10,000 trees from the posterior distribution generated from our BEAST analyses. We simulated 100 possible character histories for each of the 10,000 topologies, generating a total of 1,000,000 character-mapping realizations for each, serpentine endemism and serpentine tolerance. We used the branch lengths as direct estimate of the rate, and thus did not use a rate parameter. SIMMAP uses a beta prior on the symmetry of the transition rate matrix i.e., the extent to which transitions favor one state (0) over the other (1). The shape of the beta distribution is described by the alpha parameter and discretized into k categories. We implemented the maximum number of categories possible to discretize the beta distribution \((k=101)\), and also verified that our results did not change under alternative values for \(k\). Because we had no strong a priori reason to suspect that serpentine use was more easily gained than lost, we focused primarily on an uninformative prior (\(alpha=1.0\)). We also explored whether highly biased transition models affected the number of estimated transitions to serpentine tolerance and serpentine endemism using additional values for alpha (\(alpha=0.05, 0.5, \text{and } 10\); after (Wainwright et al., 2012)).

### 3. Results

We present the diagnostics of our individual and combined data matrices, including their dimensions, parsimony-informative characters, and models of evolution selected and implemented, in Table 2. Individual gene trees vary in their degree of resolution, ranging from quite poor (e.g., chloroplast regions and *phyA* to moderate (e.g., *G34, G61, and G56*; Figs. S2–S9). Combined analyses with multiple accessions per species (selecting alleles that minimized tree length) resulted in good support for reciprocal monophyly in the vast majority of the species across analyses while resolution of relationships among major clades varied. RAxML and MrBayes combined analyses yielded strong support for main clades but failed to resolve relationships among the main clades; BEAST analyses provided the most resolution among main clades (Figs. 2, S1). For the remainder of the paper we focus on individual gene trees from MrBayes analyses (Figs. S2–S9) and on combined analyses from BEAST (Figs. 2, S1).

#### 3.1. Species monophyly

Both the individual-gene and combined analyses showed overall support for species monophyly, exceptions to which are discussed below. Some of our markers were variable enough to reveal subspecific differentiation where enough sampling was available (e.g., *S. tortuosus*, *C. couleri*, *S. hyacinthoides*; Figs. S2–S9). Conflicting relationships below the species level and between markers were also observed in a few instances (Figs. S2–S9) but overall, we observed limited discordance at the species level: when two (or more) alleles were recovered for a single individual, they tended to be sister (or cluster) to one another, or to others in the same species. This greatly facilitated the process of combining our datasets from different markers. In general, for combined analyses we kept individuals that were homozygotes, and when this was not possible our general approach was to be conservative and keep the alleles that would yield the shortest (least resolved) trees. If species discordance was present, we kept both alleles in subsequent analyses.

There were several instances of non-monophyly. *Caulanthus lemmonii* and *C. couleri* are not reciprocally monophyletic, and our data suggest that the former might be a variety of the latter. Our combined and individual gene analyses also indicate a lack of reciprocal monophyly between *S. brachiatans* and *S. morrisonii* from the BM clade, the different species traditionally recognized within the Glandulosus complex, and within *C. lasiophyllus* (Figs. S1 and S2–S9). All of these entities are highly variable (Al-Shehbaz, 2010) and more detailed study would be needed to carefully assess patterns of variation and discordance.

#### 3.2. Phylogenetics of Streptanthus and its allies

Our combined analyses suggest that *Streptanthus, Caulanthus,* and *Thelypodium* are not reciprocally monophyletic as currently circumscribed (Figs. 2 and S1). Instead, the following patterns emerge from our BEAST analyses results (Figs. 2 and S1). The species of *Streptanthus* segregate in two strongly supported clades. One, which we refer to as the Streptanthoid Complex (SDC), roughly corresponds to the majority of the species of *Streptanthus* and *Caulanthus*. The other, which we refer to as Streptanthus Clade II (SC-II), encompasses the rest of the species assigned to *Streptanthus* and a few to *Caulanthus*, in addition to *Stanleya* and *Streptanthes*. There is an overall pattern of correspondence between phylogeny and geography such that, roughly speaking, the species whose ranges are centered in the California Floristic Province (CFP) are part of the SDC clade, and species with more southeastern ranges that extend beyond California belong to the SC-II clade. The relationships of these two main clades with *Sibaropsis* are not clear. One marker did not amplify in *Sibaropsis* (G34) and four (*phyA, ITS, trnL, trnH-psbA*) lacked resolution on their own (see Figs. S6–S9). Among the nuclear markers that informed the relationships with *Sibaropsis*, there was variation in signal: One marker (G61) suggests *Sibaropsis* as part of the SDC clade; another (PEPC) suggests it is closely allied to some of the taxa of the SC-II clade, and a third (G56) has undergone a gene duplication in this species, with one copy suggesting *Sibaropsis* as part of the SDC clade, and the other as part of a larger clade that also contains the SDC and SC-II clades (but with unresolved relationship to them). The SDC clade is composed of three main clades with strong support, here referred to as Streptanthus Clade I (SC-I clade; \(pp=98\)), Caulanthus Clade (CC clade; \(pp=1\)), and Guillenia Clade...
The distribution of the times that serpentine tolerance evolved is bimodal, with one scenario involving 8–10 gains and few subsequent losses, and a second less likely scenario of only 1–2 gains with 7–9 subsequent losses (Figs. 3, 4B, S12–13). Alternative resolutions of the CC, GC and SC-I clades of the Streptanthoid complex and the stochastic component of our character mapping approach likely contribute to this bimodal pattern in the evolution of serpentine tolerance. To investigate the extent to which these two scenarios merely represented alternative topological resolutions of the Streptanthoid complex, we compared the distribution of alternative topologies across the two serpentine tolerance scenarios. We found no evidence of alternative topologies segregating with the two scenarios (Fig. S15), indicating that both scenarios are possible, and that a ‘many-gains’ scenario is more likely than a ‘few-gains’ one.

In summary, taking into account uncertainty in tree topology, branch length estimation and character history, our data supports that serpentine endemism evolved a minimum of four times in the Streptanthoid Complex (Figs. 3A and 4A), and that serpentine tolerance evolved between eight and ten times being subsequently lost about twice (SDC; Figs. 3C and 4B). We found no association with serpentine in the SC-II clade (Fig. 4), a fact that may reflect the rarity of serpentine substrates in the ranges of the species in this clade.

We show that serpentine endemism has likely evolved between four and five times in the Streptanthoid complex, being lost only once (Figs. 3 and S10–13). Visual inspection of the character mapping realizations suggests that serpentine endemism most likely evolved on the branches leading to *S. polygaloides*, the BBM clade, *S. howellii*, and *S. barbatus* (Fig. 4A).

### 4. Discussion

Integrating algorithmic approaches at identifying SCNGs, available ESTs, and low-coverage genomic scans we have developed the most resolved phylogenetic hypothesis to date of *Streptanthus* and allied genera, which together represent a significant component of the endemic flora of California. Moreover, we take advantage of *Streptanthus*’ uniquely high proportion of taxa associated with serpentine in California (Safford et al., 2005), and our phylogeny to explore the origin and evolution of serpentine use in this group. We also provide important new information on the phylogenetic relationships within the tribe Thelypodieae, a group long recognized as challenging in crucifer systematics (Warwick et al., 2010), and hope that our markers will also help elucidating relationships among other members of this tribe.

---

**Table 2**

| Matrix dimensions, partition scheme and models of evolution implemented. |
|------------------|------------------|------------------|------------------|------------------|------------------|
|                  | **ITS**          | **G-34** (AT4G34700) | **G-56** (AT1G56590) | **G-61** (AT1G61620) | **PEPC**          |
| **Individual**   |                  |                  |                  |                  |                  |
| **datasets**     |                  |                  |                  |                  |                  |
| # taxa           | 146              | 136              | 124              | 119              | 131              |
| # chars          | 684              | 808              | 937              | 1632             | 319              |
| cte              | 471              | 236              | 380              | 903              | 205              |
| NPIC             | 52               | 201              | 150              | 313              | 79               |
| % PIC            | 23.5             | 45.9             | 43.4             | 416              | 79               |
| best model (AIC) | SYM+G            | TPM2uf+G         | TIM3+G           | HKY+G            | TPM2uf+G         |
| best model (LRT) | SYM+G            | SYM+G            | Tn+G             | HKY+G            | HKY+G            |
| model MrBayes²   | GTR+G            | HKY+G            | GTR+G            | HKY+G            | HKY+G            |
| model BEAST³    | SYM+G            | TPM2uf+G         | TIM3+G           | HKY+G            | TPM2uf+G         |
|                  |                  |                  |                  |                  |                  |
| **Partitioned**  |                  |                  |                  |                  |                  |
| **datasets**     |                  |                  |                  |                  |                  |
| # taxa           | 137              | 86               | 105              | 113              | 126              |
| # chars          | 683              | 766              | 922              | 1633             | 318              |
| cte              | 476              | 326              | 394              | 919              | 213              |
| NPIC             | 49               | 192              | 150              | 314              | 29               |
| % PIC            | 23.1             | 32.4             | 41.0             | 24.5             | 23.9             |
| model BEAST³    | SYM+G            | TPM2uf+G         | TIM3+G           | HKY+G            | TPM2uf+G         |

---

² Forward and backwards criteria produced same model.
³ Final model implemented that guarantees stable chain behavior.
4.1. Systematics of the Streptanthoid complex

Our analyses reveal that *Streptanthus*, *Caulanthus*, *Streptanthella*, *Stanleya* and *Thelypodium* are not reciprocally monophyletic as currently circumscribed, and that major taxonomic revisions are needed in these genera. While formalizing taxonomic changes is beyond the scope of this paper, it is the subject of ongoing collaborations.

The lack of resolution among sub-clades of the SDC and SC-I clades, patterns in individual gene trees, and subtle correlations in the tree likelihood values among gene regions observed in the BEAST analyses suggest discordance among the genomic regions analyzed, in support of the hypothesis that the Thelypodieae underwent a rapid diversification potentially spurred by instances of past hybridization (Warwick et al., 2009, 2010). One example that can be interpreted as a mixed signal suggesting hybridization is the relationship between *C. simulans* and the *C. heterophyllus* clade. Two gene regions support their close relationship (G34, pp = 0.97, Fig. S2; G56, pp = 0.85, Fig. S3) and none contradicted it. However, the predominant signal in the combined analyses suggests that *C. simulans* is part of the larger SC-I clade. Morphology supports this relationship but only partially, as *C. simulans* has only slightly bilateral flowers with clearly urn-shaped calyces. Analyses relying on data at genome-wide scales might be the only way to resolve this and similar instances.

Within the Streptanthoid clade (SDC), our analyses support phylogenetic relationships that are consistent with those recovered in earlier studies using chloroplast restriction sites (Mayer and Soltis, 1994). Examples are the sister relationships between *S. polygalooides* and the Tortuosus Alliance, and the close association of *S. tortuosus* complex with *S. farnsworthianus*.
relationship between *S. barbiger* and *S. brevii*, and among *C. californicus*, *C. inflatus* and *C. coulteri*.

The combination of markedly urn-shaped calyces and strongly zygomorphic flowers is a synapomorphy of the species of *Streptanthus* that belong to the SC-I clade, and provides morphological evidence of their cohesiveness. However, floral zygomorphy and urn-shaped calyces have, separately, undergone some convergent evolution. Species in the Caulanthus clade (CC) have

![Fig. 3. Instances in which serpentine endemism evolved (A) and was lost (B), and in which tolerance to serpentine was acquired (C) and lost (D) in *Streptanthus* and its allies. Serpentine endemism was gained 4–5 times (A) in the Streptanthoid Complex, and lost once (B). The most likely scenario regarding the evolution of serpentine tolerance in this group involves 8–10 gains, and about two losses (D). Histograms represent a total of 1,000,000 character mapping realizations for each, endemism and tolerance (100 character realizations over 10,000 trees of the posterior distribution generated by phylogenetic analyses using BEAST v. 1.7.4).](image)

![Fig. 4. Examples of stochastic character mapping realizations (SIMMAP v. 1.5.2.) of serpentine endemism (A) and tolerance (B) in *Streptanthus* and its allies. This scenario fits the most probable scenario, as inferred from 100 character mapping of serpentine endemism and tolerance across 10,000 trees (total of 1,000,00 realizations) from the posterior distribution of BEAST analyses (see Fig. 3).](image)
somewhat inflated calyces but their flowers are not markedly zygomorphic. Some *Streptanthus* species in the SC-II clade can present strongly zygomorphic flowers (e.g., *S. cutleri*) but not in combination with inflated calyces.

There is a tight correspondence of life history and morphology with the phylogenetic relationships of *Caulanthus*. All the annual species, which also bear a terminal cluster of densely pigmented flower buds or bracts that give its conspicuousness has been termed a ‘flag’ (Al-Shehbaz, 2010), form a well-supported clade (CC clade). Together with the *Streptanthus* Clade I (SC-I) they are part of the Streptanthoid Complex (SDC). The perennial and bivernual species of *Caulanthus* are associated with the eastern taxa (SC-II clade) and do not form a clade: *C. major* and *C. crassicaulis* are sister species and close relatives of *Stanleya*; and, *C. glaucus* and *C. pilosus* are allied to the species of *Thelypodium* included in our study with strong support (pp = 0.95–1.0).

In support of some author’s views (Buck, 1995), there is a third well-supported clade in the SDC, which we refer to as the Guillema clade (GC). Members of the GC clade lack urn-shaped calyces, and are morphologically and ecologically diverse: *C. aniceps* is associated with alkaline soils (Al-Shehbaz, 2010; Baldwin et al., 2012), *C. flavescens* with serpentine soils (Baldwin et al., 2012), and *C. lasiophyllus* is a variable and widespread taxon associated with a variety of soils, serpentine among them (Al-Shehbaz, 2010; Anacker et al., 2011; Baldwin et al., 2012).

In summary, our results reveal that zygomorphy and urn-shaped calyces have evolved in a convergent fashion in the Streptanthoid Complex, and suggest that formal revisions to the genera *Streptanthus*, *Caulanthus*, and possibly *Thelypodium* and *Stanleya*, are needed. Ongoing collaborations will formalize the taxonomy of these alliances. Further studies with wider gene sampling would be required to resolve conclusively the relationships of some taxa, including *C. pilosus*, *C. hallii*, and *C. simulans*, as well as among the major clades identified here (e.g., SC-I, SC-II, CC, GC).

4.2. Origin and evolution of associations with serpentine in the Streptanthoid complex

We find support for the evolution of serpentine use in the Streptanthoid Complex from a non-serpentine ancestor. Also, our data supports a southern origin of the group, with subsequent northward expansion. Previous authors have hypothesized that the onset of a mediterranean climate in the CFP facilitated the northward expansion of warm-adapted taxa coming from the southeast (Raven and Axelrod, 1978). Our phylogenetic hypothesis supports this scenario, with the clade SC-II conforming by species with ranges with substantial portions occurring outside of the CFP and of which no species is a serpentine user. The lack of associations with serpentine in this clade might reflect the geographic distribution of serpentine outcrops, which are more common and extensive in northern than in southern California (see Fig. 1 in Harrison et al. (2004)). Our analyses support at least four instances of evolution of endemism related to serpentine, all in the northern Californian taxa. It is possible that adaptation to dry and highly exposed environments might be an extaptation for serpentine use in this group (Pepper and Norwood, 2001). We also show that transitions out of the serpentine endemic state in the Streptanthoid complex are rather uncommon, a finding that is consistent with patterns in other taxonomic groups (Anacker et al., 2011).

Subspecific variation within widespread taxa is an important contributor to diversity, and our results suggest that the interplay of intraspecific variation with serpentine tolerance has had an important role in promoting diversity in the SDC clade. Two of the largest and most diverse clades of the SDC clade (in terms of species, subspecies and varieties described, range size, and total amount of branch length in the trees) are primarily associated with serpentine: the GH clade (*Glandulosus complex* + *Howellii alliance*) and the Batrachopus-Breweri-Morrisonii (BBM) clades. Also, the widespread species *S. tortuosus*, while primarily a non-serpentine taxon, has populations that grow on serpentine. The isolated and discrete nature of serpentine outcrops may promote speciation through drift and founder effects. Moreover, it has been suggested that rare serpentine endemics could evolve through self-fertilization, specifically in *S. vernalis* (O’Donnell and Dolan, 2005). Our results support these authors’ earlier finding that *S. vernalis*, known from a single locality, is sister to the serpentine endemic *S. morrisoni*, with which it co-occurs.

Species with extremely narrow distributions limited to single geographic features like mountains or peninsulas are common in the SDC clade (*S. fenestratus*, *S. callistus*, *S. hispidus*, *S. niger*, *S. insignis*, *S. albidus*, *S. vernalis*). While these rare endemics are themselves not all limited to serpentine, serpentine tolerance might have played a role in their evolution. We show that *S. niger*, *S. callistus*, *S. albidus* and other lineages from the Glandulosus complex have evolved from an ancestor that likely resembled *S. glandulosus* ssp. *glandulosus* (Fig. S1), in agreement with previous hypotheses (Mayer and Soltis, 1994), and which our analyses indicate was associated with serpentine (Fig. S14). In fact, when taking into account the sub-clade diversity in the Glandulosus complex, the predominant signal for the evolution of serpentine tolerance consists of only 1–2 gains with many subsequent losses (10–11 losses; Fig. S14). Within this complex, *S. callistus* and *S. hispidus*, two mountaintop endemics not associated with serpentine, might represent cases of colonization of non-serpentine suitable habitat by a more widespread taxon most likely associated with serpentine (both occur near extensive areas of serpentine geology). In this way, our results suggest that in Streptanthus, where highly variable species are associated with serpentine (e.g., *S. brevii*, *S. morrisoni*) or with diversity in soil use (e.g., *S. tortuosus*, Glandulosus complex), associations with serpentine in evolutionary time are more complex than a mere ‘dead-end’.

New approaches to marker development, especially those that can take advantage of new sequencing technologies, make it possible to resolve previously murky relationships, and thus to address fundamental questions in plant evolutionary biology (Hughes et al., 2006). By integrating previous studies and ESTs with low coverage genomic scans we were able to sequence single copy nuclear regions with enough phylogenetically informative variation to improve the phylogenetic resolution in *Streptanthus* and its allies. We hope that both our improvement in phylogenetic resolution and our inferences of associations to serpentine will lay a framework to address the evolutionary ecology of edaphic specialization using the Streptanthoid complex as a model. The genetic basis of serpentine tolerance is being studied with diverse strategies to cope with challenges posed by different soil environments makes this an ideal group in which to address questions about the ecology and evolution of edaphic specialists. More generally, knowledge about the evolutionary relationships of this group opens the door to study the relative contributions of biotic and abiotic factors to ecological specialization. Through such studies, the Streptanthoid complex can emerge as a model to study the pathways and factors promoting and maintaining edaphic specialists, which constitute a large component of biodiversity.
Author contributions

NIC and SYS envisioned the study and collected most of the specimens. NIC visualized gene region identification, designed primers, sequenced nuclear genes (G34, G61, G56, PEPC, phyA, ITS), performed phylogenetic and comparative analyses, and drafted the manuscript. AMB and AEP contributed chloroplast (trnL-psHA, trnH) and some ITS sequences. NIC, SYS, AMB and AEP revised the manuscript.

Acknowledgments

We are deeply grateful to Luca Comai for laboratory space and resources for primer design. Bryan Drew and Patrick McIntyre provided helpful discussion for this study. Ryan O’Dell, Dick O’Donnell, Elinor Gates, and Roger Raiche assisted locating some populations. Jon Rebman from the San Diego Natural History Museum Herbarium (SDNHM), Ellen Dean from the UCD Herbarium (DAV), Ryan O’Dell, Walter Holmes, Norma Fowler, Mark Fishbein, and Phyllis Scherich, provided material for this study. Paul J. Greer and Ryan J. Williams assisted with specimen collection and generation of chloroplast sequences. This study was supported by the NSF through awards DEB 0919559 (S.Y. Strauss), DEB 0416169 (A.E. O’Dell), and DBI 0733857 “Functional genomics of plant polyplody” (L. Comai), and by a Conacyt Fellowship to N. I. Cacho (EPSCI #187083).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.11.018.

References

Miller, M.A., 2009. The CIPRES Portals. CIPRES. Arch. by WebCite(r) at http://www.webcitation.org/6m3tI6OQo.
Olson, M.E., 2002. Combining data from DNA sequences and morphology for a phylogeny of Moringaceae (Brassicaceae), Syst. Bot. 27, 55–73.


