Summer 8-13-2014

Phylogenetics and Mating System Evolution in the Southern South American Valeriana (Valerianaceae)

Lauren A. Gonzalez
University of New Orleans, lag117@gmail.com

Phylogenetics and Mating System Evolution in the Southern South American *Valeriana* (Valerianaceae)

A Thesis

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Master of Science
in
Biological Sciences

by

Lauren Gonzalez

B.S. University of New Orleans, 2012

August, 2014

Table of Contents

Abstract

Species of Valerianaceae in South America represent one of the best examples of rapid diversification on a continental scale. The phylogeny of Valerianaceae has received a lot of attention within the last 10 years, but relationships among the South American species are fairly unresolved. Results from previous studies have not been well resolved with traditional genetic markers, most likely due to its recent and rapid radiation. Species in this clade exhibit a variety mating systems and inflorescence types. For the first part of this research I used several traditional plastid markers, and 3 new low copy nuclear markers to better resolve the phylogeny and then explore mating system evolution within the clade. For the second part of this research I collected high-throughput "next-generation" genomic sequence data from reduced representation libraries obtained using genotyping-by-sequencing (GBS) protocols, along with several phylogenetic methods, to try to further resolve the phylogeny of this group.

Keywords: Valeriana; Valerianaceae; mixed mating systems; next-generation sequencing; gynodioecy; genotyping-by-sequencing

Chapter 1

Mating System Evolution in the Southern South American *Valeriana* (Valerianaceae)

Introduction

Botanists and evolutionary biologists alike have long been interested in the evolution of mating systems in flowering plants (Goodwillie et al., 2005 and references within). In particular, the evolution of dioecy (separate sexes) from a hermaphroditic ancestor has received a great deal of attention (Webb 1979, Bawa, 1984, Barrett 1992, Ashman 2000, Dorken et al. 2002) especially with respect to potential morphological and ecological correlates (see Renner and Ricklefs, 1995; Thomson and Brunet 1990). Based on population genetic theory, it has been suggested that gynodioecy (the presence of both hermaphrodites and female flowers) may be an important intermediate step in the evolution of dioecy from hermaphroditic ancestors (e.g., Charlesworth and Charlesworth, 1978). Theory predicts that this 'gynodioecy-dioecy' pathway, or G-D, (from Spigler and Ashman, 2011) could originate via a stepwise process. First, male sterility alleles could arise in a population creating 'females' within the population. If these females had some selective seed-fertility advantage over hermaphrodites, they could then successfully establish themselves in the population, creating a gynodioecious taxa (Charlesworth and Charlesworth, 1978). The mere presence of these new females leads to the hermaphrodites increasing their fitness in male function (e.g., pollen production). Next, males could spread throughout the population due to a gradual reduction in female fertility and an increase in male fertility, or via additional mutations that influence male and female fertility. Finally, exclusively male individuals eliminate any remaining hermaphrodites resulting in a dioecious population or species. In a recent review, Spigler and Ashman (2011) reviewed the literature and concluded there is evidence to female advantage in gynodioecious populations.

Valerianaceae comprises 350 species that occupy a variety of habitat types around the world, shows multiple shifts in mating systems. The basal lineages in the clade, *Patrinia* and *Nardostachys*, are exclusively hermaphroditic, but there was an early shift to dioecy within the clade (Bell, 2007; Bell at al. 2012). Preliminary phylogenies of the group (Bell, 2004; Bell and Donoghue, 2005a) would suggest that dioecy has evolved at least 2 times in Valerianaceae, and that gynodioecy has evolved independently more that 5 times (Bell and Donoghue, 2005a). In addition, other mating systems, like polygamodioecy (some plants with hermaphroditic and female flowers, some plants with hermaphroditic and male flowers), have been documented within Valerianaceae (Bell and Donoghue, 2005a; Bell, 2007; Bell at al. 2012). Nowhere is the presence of gynodieocious taxa more evident than in the South American radiation of Valerianaceae, especially in the species that occur in the southern Andes (i.e., Chile and Argentina). The southern South American clade is made up of ~40 species of *Valeriana* that occur over a wide ecological, as well as elevational gradient. Most species are found in mid to low elevation habitats with a few occurring at higher elevations.

Mating systems have been shown to have morphological correlates, including flower characteristics (Renner and Ricklefs, 1995; Thomson and Brunet 1990, Drew and Sytsma 2013,). For example, it has been demonstrated that presence of many, small and white flowers is strongly correlated with gynodioecy and dioecy in species of *Lepechinia* (Lamiaceae) (Drew and Sytsma 2013). Although the majority of species of *Valeriana* in South America have small, white flowers, there is some degree of variation in floral display. In general, species within the southern South American valerians show 4 distinct inflorescence types that could be evolving in some correlated fashion with mating systems, which is explored in this study.

Much work has been done on the phylogeny of Valerianaceae in recent years, but due to its recent, rapid radiation, the southern South American clade has been difficult to resolve. Based

on a previous study (Bell et al. 2012) it appears that traditional genetic markers are lacking enough variation to confidently resolve the relationships within this clade. In this study I use several new single-copy nuclear markers to further investigate the phylogeny of the southern South American valerians. I then use the resulting phylogeny to explore mating system evolution within the group, specifically to determine how many times these mating systems have arisen in the clade and whether they are correlated with the morphological character inflorescence type.

Materials & Methods

DNA extraction, PCR amplification, cloning, sequencing, and alignment

For this study, I sampled 31 individuals (Table 1) of southern South American *Valeriana* (Appendix 1). I extracted total DNAs using the standard CTAB methods or with Qiagen DNeasy plant mini extraction kits (Qiagen). I then amplified all regions using standard Polymerase Chain Reaction in 25 μL volume reactions. Reactions conditions were as follows: an initial denaturation at 94° C for 3 min; then 35 cycles consisting of 94° C for 1.5 min, 48°- 56° C for 2 min, and 72° C for 3 min. I then cleaned all amplified PCR products prior to sequencing using ExoSap-IT (USB-Affymetrix).

I amplified and sequenced 8 chloroplast regions that have previously been examined in Valerianaceae, including: *matk, accD, ndhJ, trnD, trnG, trnK, trn:, ycf5* (Bell et al., 2012). In addition to the chloroplast genome I amplified and sequenced 3 low copy conserved ortholog set (COS) markers (Fulton et al., 2002), including *Agt1, Chlp*, and *Hmgs*, using published primers (Li et al., 2008), as well as the nuclear ribosomal internal transcribed spacer region (*ITS*), using primers ITS2, ITS3, ITS4, and ITS5. For each of the nuclear markers, we cloned all PCR products using an Invitrogen Topo-TA cloning kit (Invitrogen, Inc., Carlsbad, California). We then screen 8-16 clones per sample to evaluate sequences heterogeneity.

I sequenced via dye terminator cycle sequencing using the protocol specified by the manufacturer and then visualized on an ABI 3100 capillary sequencer. Next, I visualized and edited sequence fragments using the computer package Sequencher (Gene Codes Corporation, Ann Arbor MI) to build contig sequences. Finally, I aligned all sequences visually with the help of MacClade version 4.0 (Maddison and Maddison, 2000).

Phylogenetic analysis and divergence time estimation

I performed a preliminary maximum likelihood analysis with a model of molecular evolution determined by using the Akaikie Information Criterion (AIC) using MrModelTest ver. 2 (Nylander, 2004). In both cases, the AIC favored a GTR+I+G model of molecular evolution for our set of aligned sequences. I performed 10 random-stepwise-addition searches for each data set. Maximum likelihood searches were conducted using heuristic search methods with tree bisection reconnection (TBR) branch swapping, collapse of zero-length branches, and all characters weighted equally. The analyses were repeated 100 times with the RANDOM ADDITION option. Sets of equally most parsimonious trees were summarized with a strict consensus tree. Bootstrap tests (Felsenstein, 1985) were performed using 300 replicates with heuristic search settings identical to those of the original search. All maximum likelihood analyses were performed using the computer software PAUP* vers. 4.0b10 for UNIX (Swofford, 2002).

In addition to maximum likelihood analyses, I estimated the group's phylogeny and divergence times simultaneously. For these analyses I used a Bayesian method (Drummond et al., 2006) with an uncorrelated lognormal (UCLN) relaxed clock implemented in the program BEAST ver. 1.7.2 to estimate divergence times within the southern South American valerians. I performed two analyses: 1) in the first, I assumed a single common model across the

concatenated dataset, and 2) in a second analysis, in which I partitioned the data set by gene, I estimated separate rates and rate-change parameters for each partition. Bayes factors, as calculated in Tracer, favored the uncorrelated lognormal (UCLN) model for rate change over the strict clock model (see Nylander et al., 2004, and references therein).

I set the underlying model of molecular evolution to be GTR + I + Γ, for each of the individual genes. I also used the UCLN model, which allows for rates of molecular evolution to be uncorrelated across the tree. BEAST also allows for uncertainty in the age of calibrations to be represented as prior distributions rather than as strict/fixed calibration points. For each analysis, I initiated four independent MCMC analyses from starting trees with branch lengths that satisfied the priors on divergence times. A starting tree with branch lengths satisfying the fossil prior constraint was created using r8s v.1.7 with nonparametric rate smoothing (NPRS). For each MCMC analysis, I ran six independent chains for 100 million generations and assessed convergence and stationarity of each chain to the posterior distribution using Tracer v.1.3 (Drummond and Rambaut, 2003) and by plotting time series of the log posterior probability of sampled parameter values. After stationarity was achieved, I sampled each chain every 1000 steps until an effective sample size (ESS) of greater than 200 samples was obtained. If convergence between the independent chains was evident, I combined the samples from each run using LogCombiner v.1.4.7 (part of the BEAST distribution).

Divergence times for Valerianaceae have been estimated in the broader context of Dipsacales evolution (Bell and Donoghue, 2005b). For the analyses here I set the age of the root node (i.e., the most recent common ancestor of the southern taxa and included outgroup taxa) to a uniform prior between 3.5 and 23 million years. These values represent a range in mean values obtained by Bell and Donoghue (2005b) across different dating estimation methodologies. Nevertheless, without a reliable fossil record divergence time estimation in Valerianaceae

remains tentative.

Mating system and inflorescence evolution

To investigate the evolution of the different mating systems and inflorescence types I reconstructed character states under the parsimony criterion using Mesquite ver. 2.75 (Maddison and Maddison 2011). I also inferred ancestral states of both characters in Mesquite under maximum likelihood using a one-parameter Mk1 model (Lewis, 2001) of character state change. I based ancestral state reconstructions on the Bayesian tree inferred with BEAST. For this analysis I coded mating systems as (0) hermaphroditic, (1) gynodioecious, (2) dioecious and (3) polygamodioecious, and inflorescence types as (0) capituliform, (1) paniculiform, (2) glomeruliform and (3) spiciform (Kutschker 2011).

To estimate rates of transitions among mating system character states I used a discrete model as implemented in BayesTraits (Pagel 1994, Pagel and Meade 2006) under the maximum likelihood criterion. For the first analysis, I estimated the rate of transition between (0) hermaphroditism and (1) mixed mating systems that included gynodioecy, dioecy and polygamodioecy. For the rest of the analyses, I estimated the transition rates of each individual mating systems under three different models: 1) an unconstrained model with 12 parameters, all transitions between each system are estimated with no restrictions; 2) a 1 parameter model, where all rates are equal; 3) a 9 parameter model where dioecy is restricted from transitioning to any other mating system (constraining the rate of change from dioecy to each other mating system to = 0).

Correlated evolution

I tested for correlated evolution between mating system and inflorescence type using Pagel's 1994 test of correlated (discrete) character evolution, implemented in Mesquite ver. 2.75,

which takes two binary characters and compares the likelihood ratios of two models. In the first model the rates of change of each character are independent of the other, and in the second model the rates of change are dependent on the state of the other character. Since in the null hypothesis (first model), each character has a separate rate of change both forwards and backwards (4 rates total), is nested within the more complicated second model that has 8 rates, as each rate from the null model is split, likelihood ratios must be compared instead of likelihoods. In this analysis, I coded mating systems as (0) gynodioecious, as this was the inferred ancestral state of the group, and (1) for all other mating systems (hermaphroditic, dioecious, and polygamodioecious). I coded inflorescence types as (0) paniculiform, as this was the inferred ancestral state of the group, and (1) for all other inflorescence types (capituliform, spiciform, and glomeruliform). I used 1000 simulations to generate likelihoods, from which a likelihood ratio is calculated. The distribution of likelihood ratios from the simulated data is then compared to the likelihood ratio of the actual data to calculate a p-value.

Results

Phylogenetic analysis and divergence time estimation

Maximum likelihood searches found a single tree with a $-lnL$ score of 20657.03. Bootstrap support for clades in the ML analysis can be found in Fig. 1.

The resulting tree from the simultaneous estimation of phylogeny and divergence times with BEAST is shown is Fig 1. Overall support values for most of the clades recovered here are fairly high, with 22 out of 30 clades > 0.95 posterior probability. There were a few differences in clades recovered with BEAST than maximum likelihood. In the ML analysis *V. nivalis* was most

closely related to *V. chilensis*, but in the Bayesian analysis it was most closely related to *V. lobata. V. laxiflora*'s placement is also different, being sister to a much larger clade, and different, clade in the BEAST analysis than the ML analysis. The incongruence can be better visualized in Fig. 1, where bootstrap values for clades that were recovered in both the Bayesian and ML analyses are mapped onto the phylogeny, alongside posterior probabilities for all clades.

The origin of the entire clade of southern South American valerians was estimated to be about ~12.8 million years ago (mya) here, with a 95% confidence interval ranging between ~8.7-16.8 million years. Ages for individual clades can be visualized in Figs. 1 and 2.

Mating system evolution

The distribution of character states across taxa can be seen in Fig. 2. Parsimony reconstruction of mating system character states inferred 12 changes (steps) across the tree, with 6 transitions from gynodioecious to hermaphroditic and 1 from gynodioecious to dioecious. Mesquite inferred 1 transition from hermaphroditic to dioecious (in *V. polystachya*) and 1 transition from hermaphroditic to gynodioecious near the base of the tree. The remaining 3 changes were equivocal and concerned the evolution to the polygamodioecious state (*V. macrorhiza*), and an additional change to dioecy (*V. polybotrya* and *V. stuckertii*) from either a polygamodioecious, gynodioecious, or hermaphroditic state. The results of the maximum likelihood ancestral state reconstruction are shown in Fig. 3, with pie charts at each node representing character state probabilities.

Maximum likelihood inference of transition rates of hermaphroditism and mixed mating systems showed that the rate of evolution towards a mixed mating system was 3 times the rate of going from a mixed mating system to an exclusively hermaphroditic system (Table 2, M3).

The other models of transition rates can also be found in Table 2. The constrained model assuming an equal rates one-parameter Mk model, estimated a rate of 0.0334 (-*ln*L = 38.201). The unconstrained model (12 parameters) showed that the rate of evolution going from a hermaphroditic system to a gynodioecious system is twice that of going from gynodioecious to hermaphroditic. The rate of going from gynodioecy to dioecy, and vice versa, is very low, <0.000. The rate of evolution for polygamodioecy to hermaphroditism is quite high, 0.483, but to gynodioecy or dioecy is very low, <0.000. The rates of going from dioecy to hermaphroditism and vice versa are also fairly low, 0.053 and 0.086 respectively. The 9 parameter model, where dioecy is restricted from evolving into any other mating system (equal to 0), the rate of evolution of polygamodioecy to gynodioecy is estimated the highest at 1.049, followed by the rate of hermaphroditism to polygamodioecy at 0.111. All other rates are fairly low, < 0.1, and can be seen in Table 2.

Correlated evolution

Pagel's 1994 test of correlated (discrete) character evolution, implemented in Mesquite, estimated a p-value of 0.005, indicating that I can reject the null hypothesis that the rates of change of mating system and inflorescence type are independent of each other. The correlation can be visualized in Fig. 3 where the phylogeny is mirrored against itself with one character mapped on each side. Gynodioecy is most often found in taxa that have a paniculiform inflorescence. More about possible correlations in mating systems can be found in the discussion section.

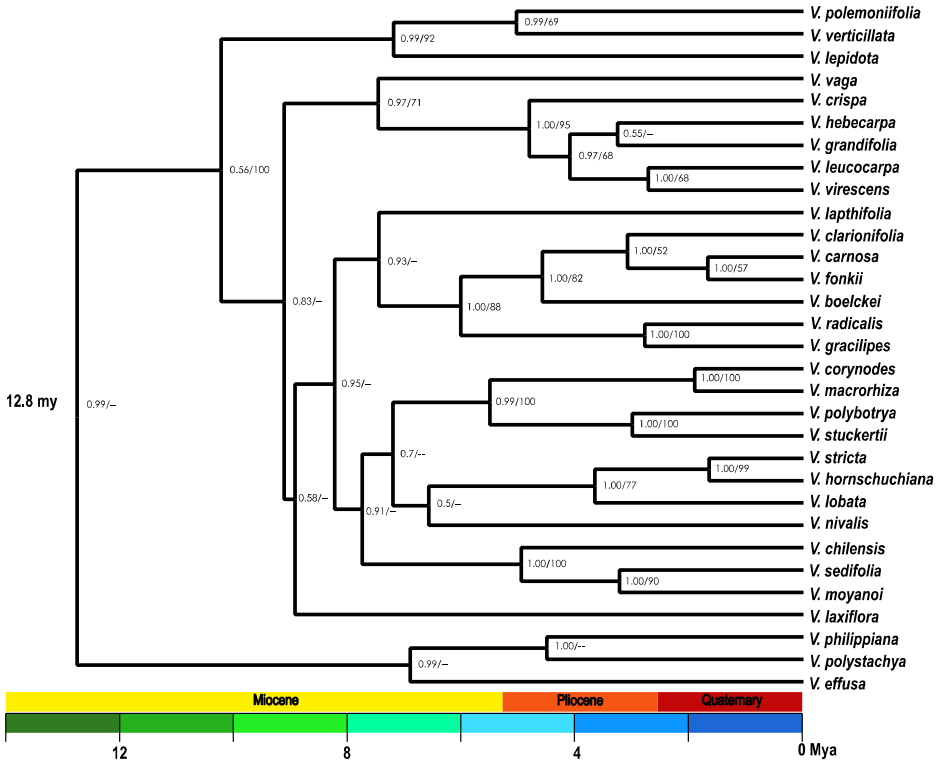


Figure 1. Phylogeny of the southern South American *Valeriana.* Inferred with BEAST, showing divergence times, with support values (posterior probabilities for all clades followed by bootstrap values for clades also recovered with ML).

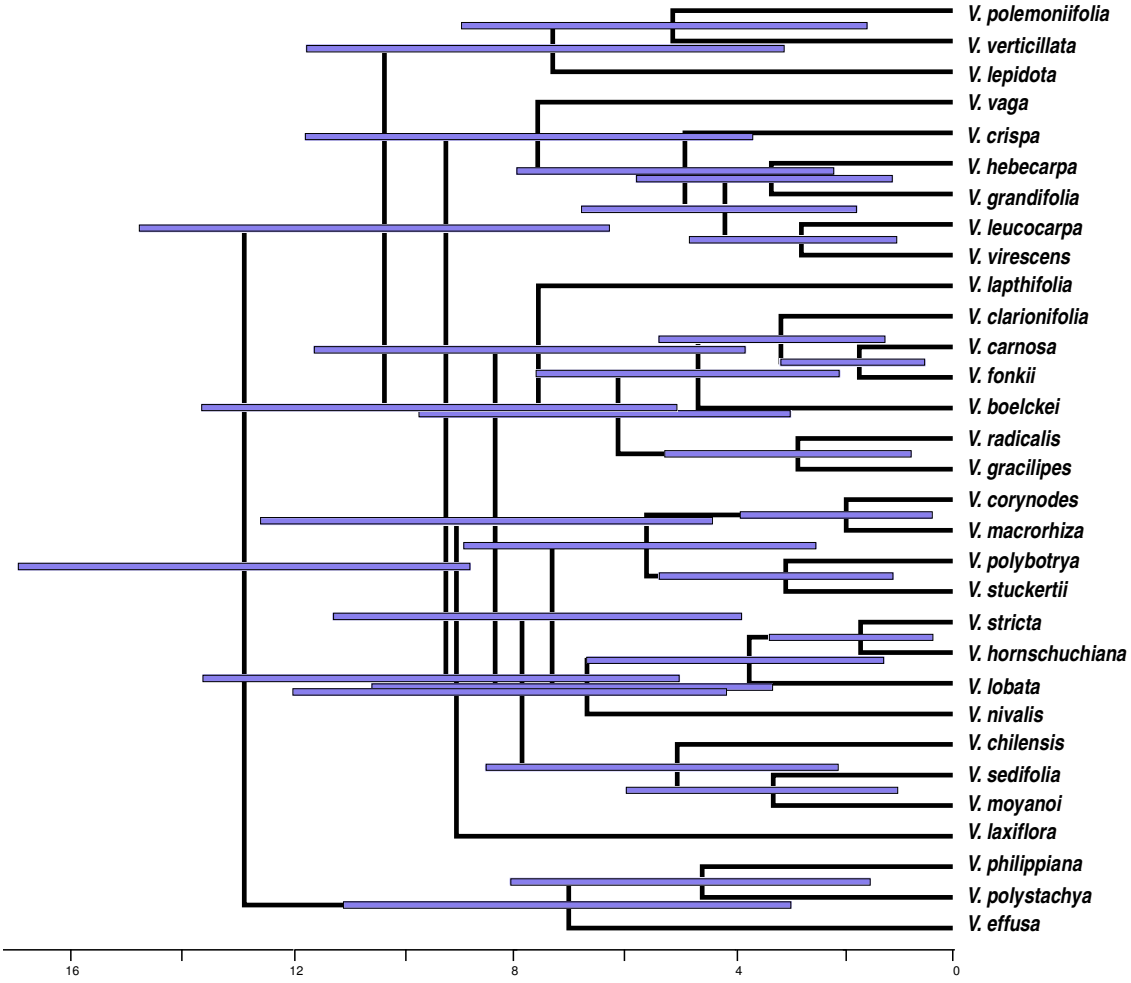


Figure 2. Chronogram showing 95% confidence intervals (blue bars).

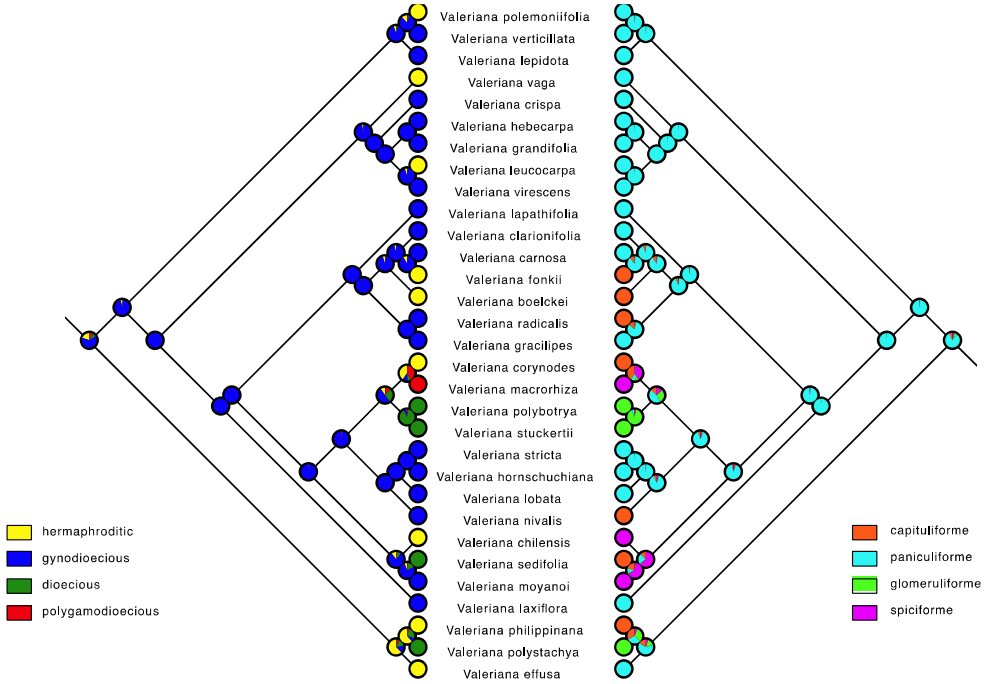


Figure 3. Mirror trees, using the topology inferred with BEAST, showing ML ancestral state reconstructions (mating systems left, inflorescence types right).

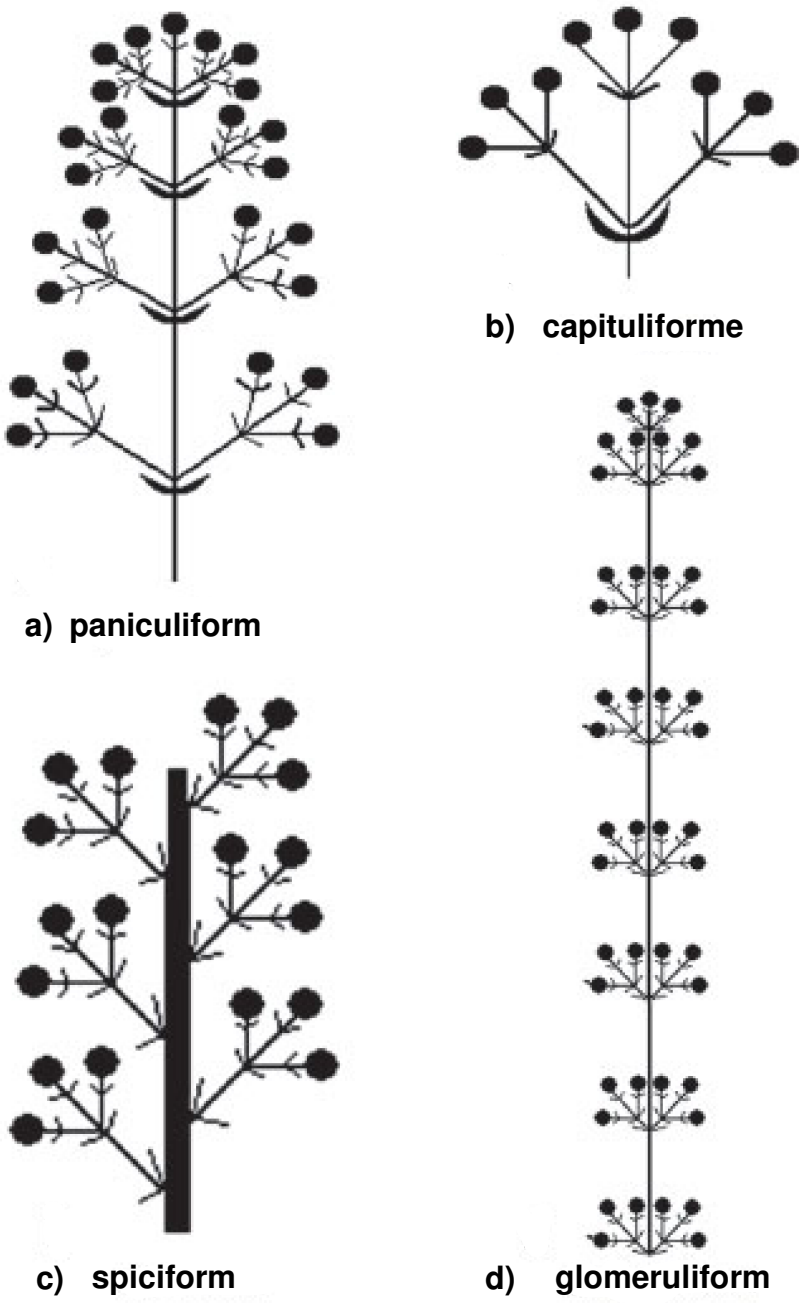


Figure 4. Inflorescence types in the southern South American *Valeriana.* Illustrations from Kutschker 2011.

Discussion

The diversity of mating systems present in the southern Andean valerians makes them an excellent system to study their origin and maintenance; however, it has been difficult to confidently resolve species relationships within this group. In this study I used new low copy markers, in addition to previously used sequence data for this group, to further resolve the phylogeny and then explore mating system evolution.

Phylogenetic analysis and divergence time estimation

Even with the addition of new low copy markers, relationships among the southern South American *Valeriana* remain uncertain. While the Bayesian analysis generally inferred well-supported clades (22 out of 30 clades > 0.95 posterior probability), maximum likelihood analyses generally showed low to no support for several clades, and only 11 out of 30 clades showed bootstrap support > 70 (Fig. 1). This study did however recover some different relationships than the most recent study from Bell et al. (2012). A few notable differences: here *V. virescens* was sister to *V. leucocarpa* rather than to *V. crispa*; *V. vaga* and *V. carnosa* are not closely related in our study as they were in Bell et al. 2012. More general conclusions are hard to make regarding differences in relationships recovered here and in Bell et al. (2012) as there are different taxa sampled in each study.

Divergence times estimated here are generally different for individual clades than those estimated in Bell at al. (2012), however the origin of the entire clade was estimated at ~12.8 mya here and ~13.7 mya in their study. While adding new low copy markers in this study gave different results (phylogeny and divergent times) than Bell et al. 2012 (where comparison is possible), bootstrap support was weak for several clades. Posterior probabilities were higher, but since these support values have been shown to be inappropriately high much of the time

(Simmons et al. 2004, Erixon et al. 2003, Doudy et al. 2003), it's difficult to be confident in them without high bootstrap values to corroborate.

Since the addition of new low copy markers did not confidently resolve the phylogeny of the southern South American valerians, it is clear that new phylogenetically informative data, in the form of more low copy nuclear markers or some genomic scale data that can be obtained through high-throughout sequencing, is needed. Other studies have had some success at resolving recent, rapid radiations with these types of data (Eaton & Ree 2013, Sanders et al. 2013). Until then, divergence times and species relationships in the southern South American *Valeriana* remain tentative.

Mating system evolution

This is the first study to explore mating system evolution in the southern South American valerians. Ancestral state reconstructions revealed multiple shifts in mating system, with hermaphroditism and dioecy arising independently multiple times from a gynodioecious state (Fig. 3), with the transition from gynodioecy to hermaphrodites most likely resulting from the loss of exclusively female flowers. There are also at least two polygamodioecious taxa in this clade, only one of which is included in this study (*V. macrorhiza*). While I didn't explicitly test for it, based on phylogeny and ancestral state reconstruction, there was no evidence to support gynodioecy as an intermediate step to dioecy. However, without incomplete taxon sampling and a better resolved phylogeny, this remains inconclusive.

Additionally, mating systems in this group have not been explored in any literature until now, and need further exploration to rule out any possibility of plasticity, which has been observed in other taxa, especially between mixed mating systems and hermaphroditic systems (Delph 2003, and references therein). These studies found that because gender in plants can often

be plastic, the environment can influence sex ratio and make mating system somewhat unstable. In particular, they suggest that female frequency is higher in low quality, harsh conditions. This could have implications for the southern South American valerians, as they inhabit a wide range of habitat-types.

Based on a correlation analyses, using Pagel's 1994 test of correlated character evolution, there is evidence that mating systems and inflorescence types are evolving in a correlated fashion, with a paniculiform inflorescence being most common in gynodioecious taxa. It should also be noted that 3 out of 4 dioecious taxa included in this study have a glomeruliform inflorescence, and there are no instances of glomeruliform with any other mating system. The only polygamodioecious species (*V. macrorhiza*) has a spiciform inflorescence, which is only present in 2 other species. Hermaphroditic species show an equal number with capituliform and paniculiform inflorescences, and one species with a spiciform inflorescence. Other studies have shown that mating systems are in some way correlated with certain morphological and ecological characters, such as woody habit, fleshy fruits, wind and unspecialized insect pollination, small flowers, and more (Renner and Ricklefs 1995); however, these correlations are not always consistent among different taxonomic groups. So while there are many possibilities for why certain mating systems (here, specifically the mixed mating systems) consistently display their flowers the same way (e.g., pollinators or environmental pressures), there is currently not enough data available to explore this relationship. Field observations of pollinators, georeference data for each species, and more morphological measures are needed to make any inferences.

Future directions

More phylogenetically informative sequence data is needed to further resolve the phylogeny of this group. Using a next-generation sequencing method to obtain large amounts of

genomic sequence data would be a logical next step. With more genomic sequence data, several different phylogenetic methods, such as multispecies coalescent and Bayesian Concordance analyses, can be employed for this group (see Chapter 2).

In addition, to further explore mating system evolution in this group, more geographic and ecological data will be needed. Ecological niche modeling could be used here to determine if environment variables are influencing mating system evolution and maintenance within this group. In line with this, more field observations of mating systems will be necessary to rule out plasticity.

Table 1. All samples used in this study. Mating systems and inflorescence types based on Kutschker 2011.

Species	**Collection details**	**Mating system**	**Inflorescence type**
Valeriana boelckei	Argentina: Neuquén. Dpto Huiliches.	Hermaphroditic	Capituliform

	A.L.V. 122 (BCRU)		
Valeriana carnosa	Argentina: Rio Negro, Bariloche, Cerro Otto. *Weberling 10715* (Herb. Weberling)	Gynodioecious	Paniculiform
Valeriana chilensis	Argentina: Chubut. Dpto. Futaleufú. Kutschker 808 (BCRU)	Hermaphroditic	Spiciform
Valeriana clarionifolia	Argentina: Chubut. Dpto. Futaleufú. Kutschker 802 (BCRU)	Gynodioecious	Paniculiform
Valeriana corynodes	Kiesling et al. 7712- SI (1991)	Hermaphroditic	Capituliform
Valeriana crispa	Chile, Metropolitan Region, Prov. Melipilla. Arroyo et al. # 207074 (CONC)	Gynodioecious	Paniculiform
Valeriana effusa	*Weberling s.n.* (Herb. Weberling), cultivated	Hermaphroditic	Paniculiform
Valeriana fonckii	Chile: Volcan Chillan. *Weberling 10686* (Herb. Weberling)	Hermaphroditic	Capituliform
Valeriana graciliceps	Chile, Metropolitan Region . Prov. Santiago. Arroyo & Humaña 991851 (CONC 167197)	Gynodioecious	Paniculiform
Valeriana grandifolia	*Weberling et al. 10979* (Herb. Weberling) Chile: Concepción	Gynodioecious	Paniculiform
Valeriana hebecarpa	Chile, Volcan Chillan. *Weberling & Grau 10675* (herb. Weberling)	Gynodioecious	Paniculiform
Valeriana hornschuchiana	Chile, Metropolitana Region. Prov. Santiago. Arroyo & Humaña 980630 (CONC 162930)	Gynodioecious	Paniculiform
Valeriana lapathifolia	Argentina, Chubut. Dpto. Futaleufú. Kutschker 812 (BCRU)	Gynodioecious	Paniculiform
Valeriana laxiflora	*Weberling & Grau 10663* (Herb. Weberling) Chile, Volcan Chillan	Gynodioecious	Paniculiform
Valeriana lepidota	Chile, VII Región. Prov. Talca, Arroyo & Becerra 209668 (CONC)	Gynodioecious	Paniculiform
Valeriana leucocarpa	Chile: Volcan Chillan. *Stutzel et Pfanzelt (Weberling) 10987* (Herb. Weberling)	Hermaphroditic	Paniculiform
Valeriana lobata	Chile, V Region. Prov. Petorca. Arroyo & Humaña 992267 (CONC)	Gynodioecious	Paniculiform
Valeriana macrorhiza	Argentina: Rio Negro, Bariloche, Cerro Catedral. *Weberling 10744* (Herb. Weberling)	Polygamodioecious	Spiciform

Valeriana moyanoi	Argentina: Rio Negro, Bariloche. *Puntieri 428* (Herb. *Weberling*)	Gynodioecious	Spiciform
Valeriana nivalis	Bolivia,La Paz. Eriksen & Molau, 4830 (YU)	Gynodioecious	Capituliform
Valeriana philippiana	Argentina: Rio Negro, Bariloche, Cerro Lopez. *Puntieri (& Weberling) 10746* (Herb. Weberling)	Hermaphroditic	Capituliform
Valeriana polemoniifolia	Chile: Refug. Asserradero, Chillan. *Weberling 10666* (Herb. Weberling)	Hermaphroditic	Paniculiform
Valeriana polybotrya	Argentina: Cordoba. *Bianco s.n.* (Herb. Weberling)	Dioecious	Glomeruliform
Valeriana polystachya	Argentina, Buenos Aires. Hurrell et al. 5336 (SI)	Dioecious	Glomeruliform
Valeriana radicalis	Chile, Metropolitan Region. Prov. Santiago. Arroyo et al. 201442 (CONC 162967)	Gynodioecious	Capituliform
Valeriana sedifolia	As "*V. magellanica" Weberling & Weberling 10998* (Herb. Weberling) Argentina: Tierra del Fuego	Dioecious	Capituliform
Valeriana stricta	Chile: Santiago, Los Farellones. *Weberling & Rosas 10927* (Herb. Weberling)	Gynodioecious	Paniculiform
Valeriana stuckertii	*Bianco s.n.* (Herb. Weberling), Argentina: Sierra de San Luis	Dioecious	Glomeruliform
Valeriana vaga	Chile, V Region, Prov. Quillota. Arroyo et al. 994006 (CONC)	Hermaphroditic	Paniculiform
Valeriana verticillata	Chile, VII Region. Prov. Talca. Arroyo et al. # 209817 (CONC)	Gynodioecious	Paniculiform
Valeriana virescens	*Puntieri 426* (Herb. Weberling) Argentina: Bariloche	Gynodioecious	Paniculiform

Table 2. Rates of character state transitions estimated with BayesTraits. Mating systems were coded as one of the following hermaphroditic (0), gynodioecious (1), dioecious (2), and polygamodioecious (3). Transitions among characters states are represented as q0→1 (transition from state 0 to state 1). M0 = unconstrained 12 rate model, M1 = constrained equal-rate model, M2 = constrained 9 state model. M3 = (0) hermaphroditism (1) mixed mating system

See text for more detail

Transition	M0 (*lnL* = -30.12) 12 parameters	M1 (-36.82) 1 parameter	M2 (*lnL* =-30.87) 9 parameters	M3 (*lnL* =-15.87)
q0→1	0.400	0.033	< 0.000	8.963
q0→2	0.086	0.033	0.067	
q0→3	< 0.000	0.033	0.111	
q1→0	0.203	0.033	0.096	2.614
q1→2	< 0.000	0.033	< 0.000	
q1→3	< 0.000	0.033	< 0.000	
q2→0	0.053	0.033	0	
q2→1	< 0.000	0.033	0	
q2→3	0.126	0.033	0	
q3→0	0.483	0.033	0.032	
q3→1	< 0.000	0.033	1.049	
q3→2	< 0.000	0.033	< 0.000	

Chapter 2

Exploring the utility of next-generation genomic sequence data on inferring relationships among the South American valerians

Introduction

Inferring a well-supported phylogeny of recently and rapidly diverged lineages has long been a struggle for biologists (Shaw 2002, Maddison and Knowles 2006, Weins et al. 2006, Lerner et al. 2011), particularly in plant taxa (Kelch & Baldwin 2003, Hughes & Eastwood 2006, Givnish et al 2009). Traditional markers often lack enough variation at the species level to be phylogenetically informative (Shaw 2002, Shaw et al. 2005).

Sequencing technologies have made incredible progress in the last decade, most recently with high-throughput sequencing (Mardis 2008, Kircher & Kelso 2010, Godden et al. 2013). These "next-generation" sequencing (NGS) methods produce large amounts of genomic sequence data quickly and in a more cost effective manor than traditional Sanger sequencing. Recently, phylogeneticists have begun taking advantage of reduced-representation genome methods, such as restriction-site associated DNA sequencing (RADseq; Baird et al. 2008) and genotyping-by-sequencing (GBS; Elshire et al. 2011), which produce datasets of many short sequences from all over the genome, at restriction enzyme cut-sites (Eaton & Ree 2013, Hipp et al. 2014, Jones et al. 2013, McCormack et al. 2012, Wagner et al. 2013). These "reduced-representation genome" methods are particularly useful for phylogenetic studies because they produce many loci that can be phylogenetically informative and used for organisms lacking a reference genome. Reduced-representation methods have shown promise for phylogenetic studies, especially among lineages that are <60 million years old (Rubin et al. 2012, Cariou et al. 2013, Emerson et al. 2010). This, along with recent progress in multi-locus species tree inference methods, presents a new way to overcome the longstanding problems associated with inferring the evolutionary history of recent, rapid radiations (Eaton & Ree 2013, McCormack & Faircloth 2013).

Traditionally, studies using reduced-representation methods have used RADseq; however, recently studies have begun using GBS (White et al. 2013, Lu et al 2013). GBS differs from RADseq in that the barcodes, unique short sequences used to identify samples after multiplexing, are included in one of the adaptor sequences instead of being added to each DNA sample by PCR. A second Illumina run for indexing is not needed, as it would be for RADseq, because the barcode is located just ahead of the restriction enzyme cut-site. Due to its simplicity (fewer purification steps and no fragment size selection), GBS is also more cost and labor efficient than RADseq, requiring much less prep than other methods, as it only uses one well on a sequencing plate for both DNA digestion and adaptor ligation (Elshire et al. 2011).

The advent of NGS and the ability to obtain large numbers of sequences, from multiple individuals per species across the entire genome, has led phylogeneticists to start using multilocus, and especially multispecies coalescent-based tree inference methods (eg. BEST, Liu 2008; STEM, Kubatko et al. 2009; *BEAST, Heled & Drummond 2010). It has been shown that using a concatenated approach with multiple genes can result in a well-supported, but incorrect, phylogeny (Kubatko & Degnan 2006), but multispecies coalescent-based approaches have had success in overcoming these challenges by taking into account the variation in gene histories (Delsuc et al. 2005, Rannala & Yang 2008, Kumar et al. 2012). This becomes exceedingly important for lineages that have diversified rapidly, as they are more likely to retain ancestral polymorphisms because they haven't had time to achieve reciprocal monophyly (Sanders et al. 2013, Eaton and Ree 2013). I chose to use the hierarchical Bayesian model implemented in *BEAST (Heled & Drummond 2010) for this study because it specifically models the discord between gene trees and species tree due to incomplete lineage sorting, and has shown to be superior to BEST in population size estimation (Heled and Drummond 2010).

Another multilocus tree inference method, Bayesian Concordance Analysis implemented in BUCKy (Ané et al., 2007; Larget et al., 2010), makes no assumptions about the reason for discordance among gene trees, and it doesn't assume a multispecies coalescent. BUCKy uses a non-parametric clustering of genes to reconstruct the primary concordance tree by estimating concordance factors (CFs) that measure the proportion of the genome for which each clade is true, and then builds a tree with the clades that have the highest concordance factors.

The phylogeny of Valerianaceae has received increased attention within the last 10 years with recent studies recovering strong support among the major lineages within the group (Chapter 1, Bell & Donoghue 2005a, Bell et al. 2012). These studies also found relatively strong support for a clade consisting of the bulk of the South American species. It is hypothesized that following a single introduction into South America, the group subsequently radiated and diversified, primarily in high Andean habitats. In addition, there is limited support for two South American clades, one consisting of species from the north (primarily paramo and puna habitats) and another southern clade (primarily Patagonian). However, the relationships of the taxa within each of these Andean clades have not been well resolved with traditional genetic markers (Bell et al. 2012). Because of this uncertainty, many questions about divergence times and phylogeography of this group have not been confidently resolved.

The southern South American valerians consists of about 40 described species that occur in a wide elevational as well as ecological gradient. They occur east and west of the Andes and at low and high elevations, encompassing many different habitat types. Because of this group's recent, rapid radiation and the fact that many of its species occur in one of the world's biodiversity hotspots (central Chile, Myers et al. 2000), it is a powerful model to study how

biogeography, ecology and genetics drive diversification and its implications for conservation. In order to conduct further studies, a well-supported, well-resolved phylogeny is essential.

In this study I use concatenated GBS data, along with several species tree methods, to infer the phylogeny of the southern South American radiation of *Valeriana* (Valerianaceae). Although I included only a subset of the species in this complex, this work will serve as a starting point to see if these methods will help confidently resolve these relationships and will help determine if further efforts will be valuable in understanding the evolutionary history of Valerianaceae.

Methods

Sampling & Sequencing

For this study, I originally sampled 31 species of southern South American valerians, with 48 total samples. I extracted genomic DNA from silica dried plant tissues using the CTAB method (Doyle & Doyle 1987, Cullings 1992). I prepared the GBS libraries using the protocol outlined in Elshire et al (2011). I used the restriction enzyme *Pst*I (CTGCAG) to digest the extracted genomic DNA from each individual, and then ligated the resulting fragments to a barcode adaptor and a common adaptor with the correct sticky ends. I put each individual into one well of a 96-well plate, with one well being a control containing no DNA. After digestion and ligation, I cleaned up the products using a Qiagen MinElute 96-well PCR purification kit. After PCR, I quantified the PCR products using PicoGreen and a qPCR machine, and then used the appropriate volume of each sample to end up with a 150 ng concentration. Once I obtained the correct concentration of DNA per sample, I pooled all samples into a single GBS library.

I sent the library to the Oregon State University Center for Genome Research and Biocomputing where it was run on one lane of an Illumina HiSeq sequencer to generate singe-end 100bp reads.

Clustering

I used the software pipeline pyRAD v.1.4 (Eaton 2014) to process the raw data from the Illumina FASTQ files. Unlike the pipelines that focus on preparing RADseq type data for population level analyses (e.g. Stacks; Catchen et al 2011), pyRAD aims to obtain variation across clades (species or higher) by using a global clustering and alignment method, allowing the detection of clusters with high levels of divergence. Our parameters in pyRAD were as follows: Nucleotides with Phred scores of <20 were coded as unknown bases, denoted by N's, and sequences with >5% N's were thrown out. Sequences were clustered within samples by 90% similarity via the uclust function in USEARCH (Edgar 2010). Clusters of less than 10 sequences were discarded and the minimum number of individuals per cluster was set to 5. Any locus that was heterozygous among more than 3 samples was discarded. The remaining clusters were treated as loci and assembled into a phylogenetic matrix.

I also used the R package RADami (Hipp 2014) to generate a figure showing the proportion of shared loci among individuals. This package takes as input the loci file that is output from *pyRAD* and uses pairwise comparisons of loci to calculate an average percentage of loci shared by each individual.

Phylogenetic Inference

To infer phylogenies with the GBS data I assembled 3 datasets: 1) a supermatrix that included all loci concatenated into a single alignment with N's present for loci with incomplete taxon sampling; 2) a concatenated dataset with only the loci that had full coverage among samples; and 3) a partitioned dataset of the loci that had full coverage among samples. I used RAxML 7.0.8 (Stamatakis 2006) to analyze each dataset. Models of substitution for both of the concatenated datasets and for each loci in the partitioned dataset were selected based on the Akaike Information Criterion (AIC) in MrModeltest (Nylander 2004) with likelihood calculation performed in PAUP* v.4.0a134 (Swofford, 2002). Both the supermatrix with missing data and the concatenated loci datasets were analyzed under the GTR+I+Γ nucleotide substitution model, with branch support estimated using 500 nonparametric bootstrap replicates. I analyzed the partitioned loci dataset under the GTRCAT nucleotide substitution model with branch support estimated using 500 nonparametric bootstrap replicates.

Multilocus species tree inference

I used the hierarchical Bayesian model implemented in *BEAST v1.7.5 (Heled and Drummond 2010) to estimate a species trees from the 140 loci (see results) that were present in all samples. *BEAST uses Markov chain Monte Carlo (MCMC) to estimate the posterior distribution of each of the 140 gene trees and the overall species tree. I used the previously determined substitution models for each locus, an uncorrelated lognormal (UCLN) relaxed clock and a Yule process tree prior. The MCMC analysis was run for 100 million generations, sampling every 1000 steps and discarding 10% as burnin. I used Tracer v.1.5 (Drummond and Rambaut 2009) to assess convergence and to be sure I achieved an ESS (effective sample size) of greater than 200.

Bayesian Concordance Analysis

I used the program BUCKy (Ané et al., 2007; Larget et al., 2010) to infer a species tree using the dataset consisting of the 140 loci that had full coverage among samples. For each locus I ran two independent runs in MrBayes 3.2 (Ronquist et al., 2012) for 1,000,000 generations using the previously determined nucleotide substitution models and checked for convergence in Tracer. Using the posterior sample of gene trees estimated by MrBayes, I ran BUCKy with 3 chains for 500,000 generations at various values of α (0, 5, 100, ∞), the *a priori* level of discordance among loci. Under these conditions, α=0 would indicate no expected discordance and all posterior distributions would have the same tree, while α= ∞ would indicate complete independence and each gene would have a different set of trees.

Consistency of inferred trees

To measure the consistency of the resulting trees, I used the software Compare2Trees (Nye et al. 2005) to perform pairwise comparisons of each of my resulting optimal tree topologies. This program allows you to compare two trees, obtained using different phylogenetic methods, to determine how similar or different the topologies are by calculating an overall topological score (%). I input into the program each tree, comparing two at a time, in newick format with only branch lengths labeled.

Table 1. Species names, with identifying collection details and total loci after processing with *pyRAD*, of the 18 samples used in phylogenetic analyses

Species	Collection details	Total # of loci after pyRAD
Valeriana clarionifolia	*Weberling 10707* (Herb. Weberling), Argentina: Chubut, El Condor	2109

Valeriana fonckii	Argentina, Chubut. Dpto. Futaleufú. Kutschker 803 (BCRU)	1955
Valeriana laxiflora	Argentina, Chubut. Dpto. Futaleufú. Kutschker 806 (BCRU)	1997
Valeriana leucocarpa	*Stützel et Pfanzelt (Weberling) 10987* (Herb. Weberling) Chile: Volcán Chillan	1639
Valeriana sedifolia	As "*V. magellanica" Weberling & Weberling 10998* (Herb. Weberling) Argentina: Tierra del Fuego	1838
Valeriana hornschuchiana	conc 162930	1823
Valeriana lapathifolia	Argentina, Chubut. Dpto. Futaleufú. Kutschker 812 (BCRU)	1337
Valeriana virescens	*Weberling 10828* (Herb. Weberling) Argentina: Río Negro	1999
Valeriana virescens	*Puntieri 426* (Herb. Weberling) Argentina: Bariloche	2209
Valeriana virescens	*Weberling 10714* (Herb. Weberling), Argentina: Río Negro, Bariloche	2258
Valeriana laxiflora	*Weberling & Grau 10663* (Herb. Weberling) Chile, Volcan Chillan	1418
Valeriana polemoniifolia	*Weberling 10692* (Herb. Weberling) Chile: Parque Nahuelbuta	2338
Valeriana stuckertii	*Bianco s.n.* (Herb. Weberling), Argentina: Sierra de San Luis	1376
Valeriana effusa	*Weberling s.n.* (Herb. Weberling), cultivated	1789
Valeriana lobata	*Weberling & Weberling 10938* (Herb. Weberling) Chile	1319
Valeriana grandifolia	*Weberling et al. 10979* (Herb. Weberling) Chile: Concepción	1579
Valeriana interrupta	Ruiz & Pavon *Denzinger s.n.* (Herb. Weberling) Bolivia: Copacabana	2059
Valeriana effusa	*Bianco s.n.* (Herb. Weberling) Argentina: Río Cuarto	1366

Results

Sequences

Illumina sequencing returned 283,325,239 total reads made up of 13,339 Mbases. I chose to leave out some of the samples due to poor coverage, possibly due to low quality of original

extracted DNA) and ended up with 14 species, for a total of 18 samples (Table 1). Clustering of consensus sequences with our previously mentioned parameters in pyRAD revealed 8,323 unique clusters, or loci, across all samples with 140 loci present in all 18 samples. Each of the 18 samples in the supermatrix dataset had 273,801 base pairs (those that were missing were coded as N's), resulting in a total of 4,928,418 base pairs (41% missing data). Each of the 140 loci that had full coverage was made up of 88-93 base pairs after barcodes were removed.

The output from R package RADami showing the proportion of shared loci among individuals, can be seen in Fig. 1. The average percentage of loci shared among individuals ranged from 0.33-0.51.

Phylogenetic inference

The maximum likelihood analyses recovered the same clades for each of the 3 datasets, with the supermatrix dataset having the highest bootstrap support (12 out of 15 clades with >95% and none <50%) (Fig 2.A). The partitioned loci dataset and the concatenated loci dataset returned the exact same trees with mostly high support (9 out of 15 clades with >100%, and 3 with <50%) (Fig 2.B and 2.C). In the supermatrix, *V. clarionifolia* was nested within the 3 samples of *V. virescens* (with 100% support to one sample and only 52% support to the other 2). This is the similar for the loci datasets, with 99% and 54% support respectively. The next lowest supported clade in the supermatrix dataset was the *V. effusa* clades relationship with the clade consisting of *V. fonckii, V. magellanica,* and *V. hornschuchiana*, which was 66%. In the loci datasets this relationship has an even lower bootstrap value of 35%. These analyses recovered different clades than the most recent study (Bell et al. 2012), though it should be noted that this study has less species, as well as some species that Bell et al. (2012) did not include.

Multilocus species tree inference

The *BEAST analysis recovered a slightly different topology than the ML analyses, with moderate support (only 2 out of 12 clades had a posterior probability (pp) of >0.95, and 7 out of 12 having a pp between 0.90-.94 (Fig. 2.D). Here, *V. effusa* and *V. fonckii* are sister to each other instead of *V. fonckii* being sister to *V. magellanica* and *V. hornschuchiana,* as in the ML analyses. However, the *V. effusa*/*V. fonckii* clade has a very low posterior probability (0.35).

Bayesian Concordance Analysis

BUCKy returned the same primary concordance trees, topology and concordance factors, for all runs with different values of α (0, 5, 100, ∞). The primary concordance tree recovered the same clades as the ML analyses, but with mostly low concordance factors (ranging from 0.5 as the highest, to 0.002 as the lowest), as seen in Fig. 2.E.

Consistency of inferred trees

The pairwise comparisons of each of our inferred trees (supermatrix ML, partitioned loci ML, non-partitioned loci ML, multilocus species tree, primary concordance tree) are presented as overall topological scores, the percent similarity between tree topolgies, in Table 2. The scores range from 74.5%-100% similar. These scores show how consistently the starting data infers the same tree using different phylogenetic inference methods. The ML trees from the loci datasets, both partitioned and not partitioned, returned the exact same tree. That topology was 93.1% the same as both the supermatrix ML tree and the primary concordance tree. The tree from the multilocus analysis was 75.9% similar to both the supermatrix ML tree and the primary concordance tree, and 74.5% similar to the loci ML trees.

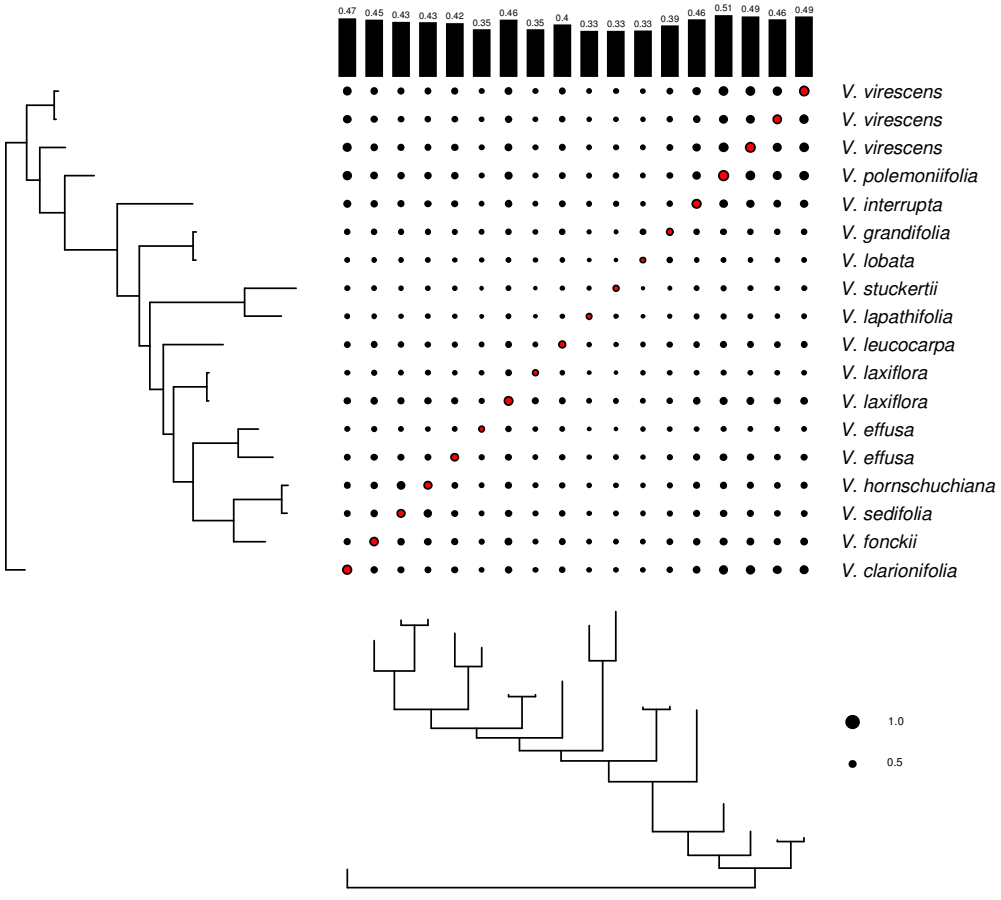
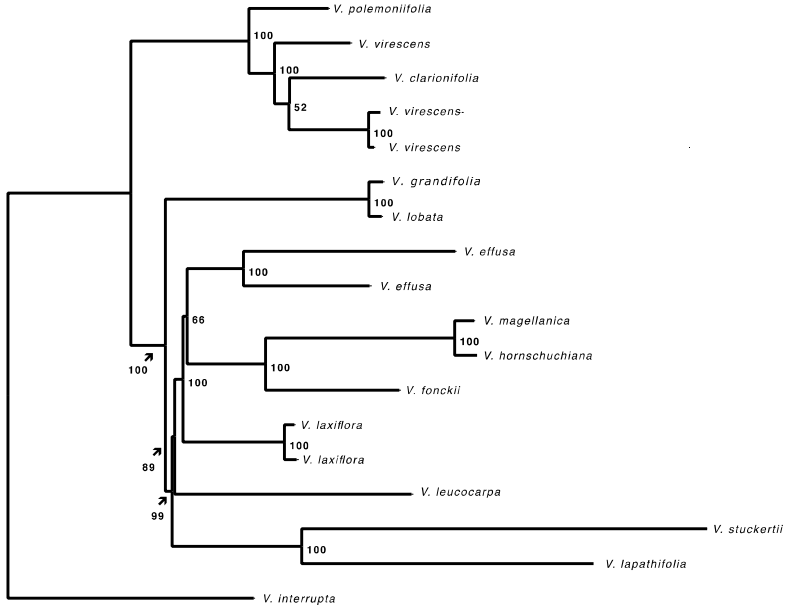


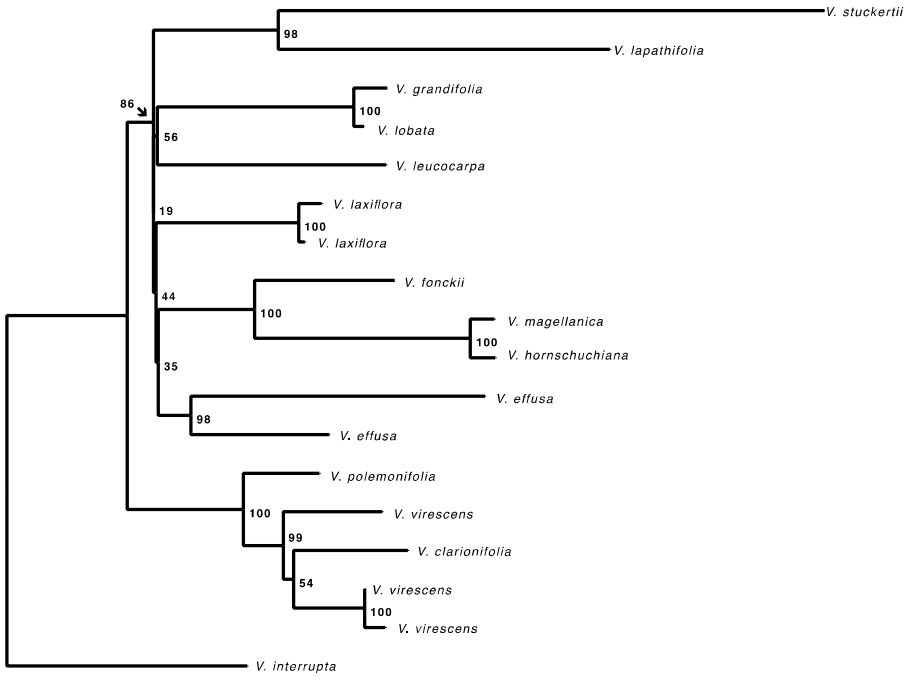
Figure 1. Proportion of loci shared among individuals. Red circles represent loci that were successfully sequenced within the individual, black circles represent loci that were shared

between individuals, expressed as a proportion of 0-1 (corresponding to the size of the circle) for all 8,323 loci returned from *pyRAD.* The bars above represent the average percentage of loci shared by each sample as an average of all the black circles for that individual.

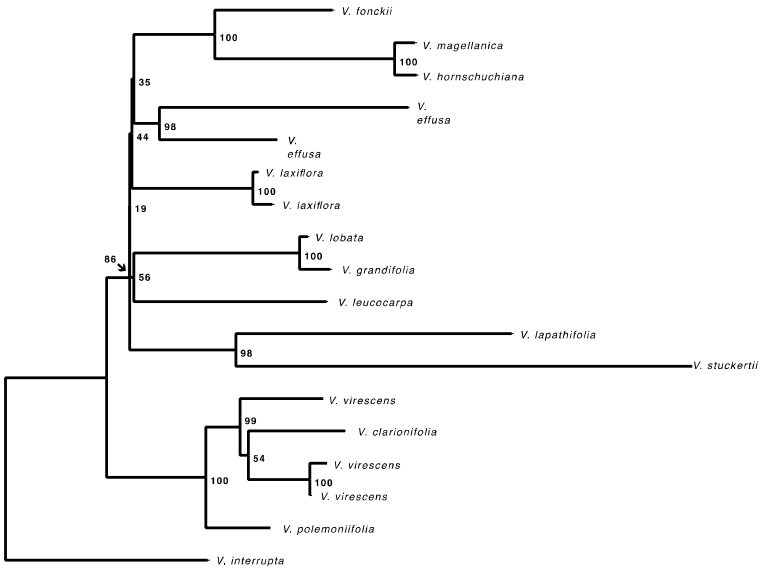
A)


V. polemoniifolia
100
V. virescens
100
V. clarionifolia
52
V. virescens-
100
V. virescens
V. grandifolia
100
V. lobata
V. effusa
100
V. effusa
66
V. magellanica
100
V. hornschuchiana
100
100
V. fonckii
V. laxiflora
100
V. laxiflora
89
V. leucocarpa
99
V. stuckertii
100
V. lapathifolia
V. interrupta

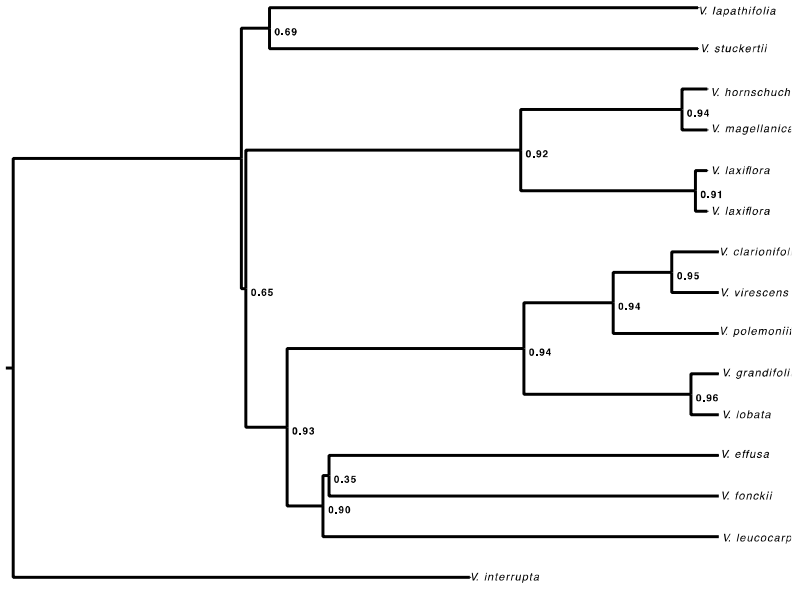
B)


V. stuckertii
98
V. lapathifolia
V. grandifolia
86
100
V. lobata
56
V. leucocarpa
19
V. laxiflora
100
V. laxiflora
V. fonckii
44
100
V. magellanica
100
V. hornschuchiana
35
V. effusa
98
V. effusa
V. polemonifolia
100
V. virescens
99
V. clarionifolia
54
V. virescens
100
V. virescens
V. interrupta

C)


V. fonckii
100
V. magellanica
100
V. hornschuchiana
35
V.
effusa
98
44
V.
effusa
V. laxiflora
100
V. laxiflora
19
V. lobata
100
86
V. grandifolia
56
V. leucocarpa
V. lapathifolia
98
V. stuckertii
V. virescens
99
V. clarionifolia
54
V. virescens
100
100
V. virescens
V. polemoniifolia
V. interrupta

D)


V. lapathifolia
0.69
V. stuckertii
V. hornschuch
0.94
V. magellanica
0.92
V. laxiflora
0.91
V. laxiflora
V. clarionifol
0.95
0.65
V. virescens
0.94
V. polemoniii
0.94
V. grandifoli
0.96
V. lobata
0.93
V. effusa
0.35
V. fonckii
0.90
V. leucocarp
V. interrupta

E)

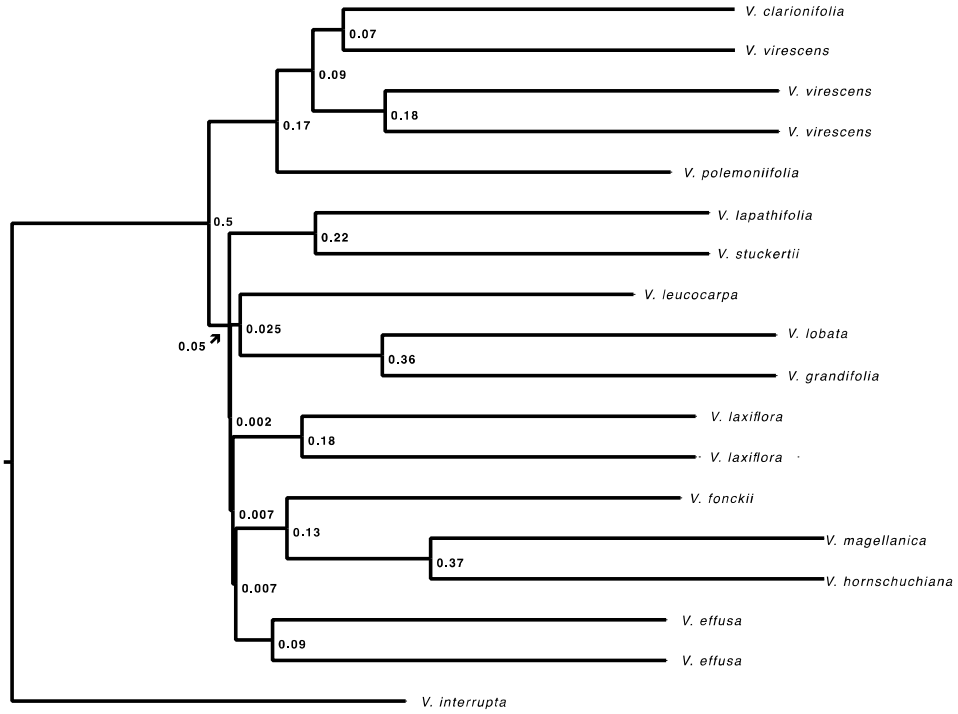


Figure 2. Phylogeny of the southern South American valerians. Using maximum likelihood, with bootstrap support values for each clade inferred with A) full-concatenated supermatrix dataset, B) partitioned 140 loci dataset, C) concatenated 140 loci dataset. Using multilocus tree inference methods D) maximum clade credibility tree inferred with 140 loci in *BEAST, posterior probabilities for each clade, E) primary concordance tree inferred with 140 loci in BUCKy, concordance factors for each clade.

Table 2. Overall topological scores (%), calculated with Compare2Trees, showing the similarity between tree topologies, with corresponding figure numbers

	Full concatenated supermatrix with ML (Fig2.A)	Partitioned loci with ML (Fig 2.B)	Concatenated loci with ML (Fig 2.C)	MCC with *BEAST (Fig 2. D)	Primary concordance tree with BUCKy (Fig 2.E)
Full concatenated supermatrix with ML (Fig2.A)		**93.1**	**93.1**	**75.9**	**93.1**
Partitioned loci with ML (Fig 2.B)	-		**100**	**74.5**	**93.1**
Concatenated loci with ML (Fig 2.C)	-	-		**74.5**	**93.1**
MCC with *BEAST (Fig 2. D)	-	-	-		**75.9**
Primary concordance tree with BUCKy (Fig 2.E)	-	-	-	-	

Discussion

Reconstructing phylogenies of recently diverged, closely related lineages is a problem that new sequencing technologies and tree inference methods are starting to overcome (Eaton & Ree 2013, Lerner et al. 2011). In this study I used a next-generation sequencing approach, GBS, to produce large amounts of genomic sequence data to infer the phylogeny of the recent radiation of the southern South American *Valeriana*. I obtained over 8,000 loci for 18 samples, consisting of 14 species, with 140 of the loci having full coverage among samples. This is a significant increase in data from the most recent study of this group, Bell et al. 2012, using only 10 gene regions (9 chloroplast, 1 nuclear).

Although most of the analyses here returned weak support (all but the concatenated supermatrix), the tree topology was fairly consistent, with different phylogenetic methods recovering mostly the same clades. Some studies that have compared several multispecies methods have recovered incongruent results and advise against using only a single species tree inference method (Lee et al. 2011, Mateos et al. 2012). Among the southern South American valerians there was some incongruence between the clades recovered with *BEAST and BUCKy, with the trees being ~75% similar. While *BEAST makes the assumption that discordance in gene trees is due to incomplete lineage sorting, a likely scenario in a recent, rapid radiation, BUCKy makes no such assumptions. Since I had no *a priori* support that only incomplete lineage sorting was responsible for discordance in this group, and *BEAST returned only low to moderate support, I chose to analyze the data with BUCKy. BUCKy returned very low concordance factors, independent of α, which indicates a lot of discordance among gene trees.

Similar studies to this one were able to recover more data (~45,00 loci per sample, Eaton & Ree 2012; ~21,000 per sample, Hipp et al. 2014), and also better supported phylogenies. In comparison, this study recovered an average of 1700 loci per sample (see Table 1), and only 140 loci with full coverage used in the multilocus analyses. However, this study had more consistency in proportion of shared loci among individuals. The average percentage of loci shared by each individual ranged from 0.33-0.51 here, but from 0.04-0.54 in Hipp et al. (2014)

However, it has also been noted that RAD loci may not be ideal for using multilocus phylogenetic methods (Eaton & Ree 2013). These loci are generally short sequences, in this study only 88-100 base pairs each, and contain very few variable sites. Because loci that lack variable sites are thrown out for phylogenetic analyses, it could be creating a bias if the variable regions are retaining ancestral polymorphisms and introgressed DNA (Eaton & Ree 2013, Ane et al. 2007). As sequencing technologies improve, these methods will become more reliable. Already, paired-end Illumina sequencing is yielding longer sequences from both RADseq and GBS methods, with loci consisting of several hundred base pairs (Etter 2011, Lemmon & Lemmon 2012).

In addition, there are a few reasons why more data, especially in the form of more accessions per species, would likely yield a better-supported phylogeny. There are some drawbacks of using only one individual to represent a species. Firstly, sequencing errors can appear to be polymorphisms, which can lead to inferring the incorrect relationships among taxa, especially if there is only one sample representing a species. Secondly, some of the multilocus coalescent species tree inference methods, including *BEAST, suggest multiple accessions per taxa in order to better estimate population size (Heled & Drummond 2010). Some of these problems might be overcome with greater sampling, both between and among species. Sampling

multiple individuals of a species will potentially drown out the effect of sequencing errors, as well as allow multilocus tree inference methods to more confidently estimate population size. This study used only one individual to represent 11 out of 14 of the species included, with only 3 species having multiple accessions.

Future directions

Based on the findings here, the next step with the southern South American valerians is to sequence more accessions per species, as well as use a method of next-generation sequencing that produces longer reads. While these steps will hopefully increase the reliability of phylogenetic tree inference methods, some of these methods should be explored using both collected and simulated data, to select for the best methods.

Additionally, in order to better understand the true evolutionary history of Valerianaceae and its closest relatives, including more taxa, such as the Northern Andean species will be crucial. Beyond phylogenetic data, more geographic (georeference points for each species) data is needed to further explore biogeography and trait evolution within this hyper-diverse clade.

Works Cited

Ané C., Larget B., Baum D.A., Smith S.D., Rokas, A. 2007. Bayesian estimation of concordance among gene trees. Molecular Biology & Evolution 24: 412–426.

Ashmann, T.L. 2000. Pollinator selectivity and its implication for the evolution of dioecy and sexual dimorphism. Ecology 81(9) 2577-91.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, e3376.

Barrett, S.C.H. 1992. Gender variation and the evolution of dioecy in *Wurmbea diocia* (Liliaceae). Journal of Evolutionary Biology 5: 423-444.

Bawa, S. S. 1984. The evolution of dioecy- concluding remarks. Annals of the Missouri Botanical Garden 71: 294-296.

Bell, C. D. 2004. Preliminary phylogeny of Valerianaceae (Dipsacales) inferred from nuclear and chloroplast DNA sequence data. Molecular Phylogenetics and Evolution 31:340-350.

Bell, C. D. 2007. Phylogenetic placement and biogeography of the North American species of Valerianella (Valerianaceae: Dipsacales) based on nuclear and chloroplast DNA. Molecular Phylogenetics & Evolution 44: 929-941.

Bell, C.D., Donoghue, M.J., 2005a. Phylogeny and biogeography of Valerianaceae (Dipsacales) with special reference to the South American valerians. Orgnaisms Diversity & Evolution. 5, 147–159.

Bell, C.D., Donoghue, M.J., 2005b. Dating the Dipsacales: comparing models, genes,and evolutionary implications. American Journal of Botany. 92, 284–296.

Bell, C.D., Kutschker, A., Arroyo, M.T.K., 2012. Phylogeny and diversification of Valerianaceae (Dipsacales) in the Southern Andes. Molecular Phylogenetics and Evolution. 63(3): 724-37.

Cariou, M., Duret, L., Charlat, S. 2013 Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and Evolution. 3: 846-852.

Catchen J.M., Amores A., Hohenlohe P., Cresko W., Postlethwait J.H. 2011. Stacks: Building and genotyping loci de novo from short-read sequences. G3 1:171–182.

Charlesworth, B., Charlesworth, D. 1978. A model for the evolution of dioecy and gynodioecy. American Naturalist. 112, 975-997.

Cullings, K.W. 1992. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. Molecular Ecology 1:233-240.

Delph LF. 2003. Sexual dimorphism in gender plasticity and its consequences for breeding system evolution. Evolution and Development. 5:31-39.

Delsuc, F., H. Brinkmann, and H. Philippe. 2005. Phylogenomics and the reconstruction of the tree of life. Nature Reviews: Genetics. 6: 361–375.

Dorken ME, Friedman J, Barrett SCH. 2002. The evolution and maintenance of monoecy and dioecy in Sagittaria latifolia (Alismataceae). Evolution 56:31-41

Douady, C. J., F. Delsuc, Y. Boucher, W. F. Doolittle, E. J. P. Douzery. 2003. Comparison of the Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Molecular Biology and Evolution 20: 248–254.

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

Drew, B.T., Sytsma, K.J. 2013. The South American radiation of Lepechinia (Lamiaceae): phylogenetics, divergence times, and evolution of dioecy. Botanical Journal of the Linnean Society. 170, 171–190.

Drummond, A.J., Ho, S.W.Y., Phillips, M.J., Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, 699–710.

Drummond, A.J., Rambaut, A. 2003. BEAST: Bayesian Evolutionary Analysis Sampling Trees. <http://evolve.zoo.ox.ac.uk/Beast/>.

Drummond A., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7: 214.

Eaton, D.A.R., 2014. *PyRAD:* assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics. doi: 10.1093/bioinformatics/btu121

Eaton D.A.R., Ree R.H., 2013. Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Systematic Biology 62: 689–706.

Edgar R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460 –2461.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E., Holzapfel, C.M. 2010. Resolving postglacial phylogeography using high- throughput sequencing. Proceedings of the National Academy of Sciences. USA 107, 16196–16200.

Erixson, P., Svennblad, B., Britton, T., Oxelman, B. 2003. Reliability of Bayesian Posterior Probabilities and Bootstrap Frequencies in Phylogenetics. Systematic Biology 52(5) 665-673.

Etter P., Preston J., Bassham S., Cresko W., Johnson E. 2011. Local de novo assembly of RAD

paired-end contigs using short sequencing reads. PLoS One 6:e18561.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Fulton, T. M., Van der Hoeven, R., Eannetta, N.T., Tanksley, S.D. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14: 1457–1467.

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 (Asteraceae: Campanulaceae). Proceedings of the Royal Society B 276: 406-417.

Godden, G.T., Jordan-Thaden, I.E., Chamala, S., Crowl, A.A., Garcia, N., Germain-Aubrey, C.C., Heaney, M., Latvis, M., Qi, X, Gitzendanner, M.A. 2013. Making next-generation sequencing work for you: approaches and practical considerations for marker development and phylogenetics. Plant Ecology & Diversity 5: 427-50.

Goodwillie, C., Kalisz, S., Ekert, C.G.. 2005. The evolutionary enigma of mixed mating systems in plants: occurance. Theoretical explanations, and empirical evidence. Annual Review of Ecology, Evolution, and Systematics 36: 47-79.

Heled, J., Drummond, A.J. 2010. Bayesian inference of species trees from multilocus data. Molecular Biology & Evolution. 27(3):570–580.

Hipp, A.L. 2014. RADami: R Package for Phylogenetic Analysis of RADseq Data. R package version 1.0-3. url: http://CRAN.R-project.org/ package = RADami.

Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S. 2014. A Framework Phylogeny of the American Oak Clade Based on Sequenced RAD Data. PLoS ONE 9(4): e93975.

Hughes, C., Eastwood, R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. Proceedings of the National Academy of Sciences USA 103, 10334–10339.

Jones, J.C., Fan, S., Franchini, P., Schartl, M., Meyer, A. 2013. The evolutionary history of Xiphophorus fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. Molecular Ecology, 22, 2986–3001.

Kelch, D.G., Baldwin, B.G. 2003. Phylogeny and ecological radiation of New World thistles (Cirsium, Cardueae-Compositae) based on ITS and ETS rDNA sequence data. Molecular Ecology 12: 141–151.

Kircher, M., Kelso, J. 2010. High-throughput DNA sequencing—concepts and limitations. Bioessays, 32, 524–536.

Kubatko, L.S., Degnan, J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56: 17–24.

Kubatko, L.S., Carstens, B.C., Knowles, L.L. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25:971–73.

Kumar, S., Filipski, A.J., Battistuzzi, F.U., Pond, S.L.K., Tamura, K. 2012. Statistics and truth in phylogenomics. Molecular Biology & Evolution 29: 457–472.

Kutschker A. 2011. Revisión del género Valeriana (Valerianaceae) en Sudamérica austral. Gayana Botanica 68 (2): 244—296.

Larget, B.R., Kotha, S.K., Dewey, C.N., Ané, C. 2010. BUCKy: gene Tree/Species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26:2910–2911.

Lee, J.Y., Joseph, L., Edwards, S.V. 2011. A species tree for the Australo-Papuan Fairy-wrens and allies (Aves: Maluridae). Systematic Biology 61(2): 253-271.

Lemmon, A.R., Lemmon, E.M. 2012. High throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. Systematic Biology 61:745–61

Lerner, H.R.L, Meyer, M., James, H.F., Hofreiter, M., Fleischer, R.C. 2011. Multilocus Resolution of Phylogeny and Timescale in the Extant Adaptive Radiation of Hawaiian Honeycreepers. Current Biology 21: 1-7.

Lewis, P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology. 50, 913–925.

Li, M., Wunder, J., Bissoli, G., Scarponi, E., Gazzani, S., Barbaro, E., Saedler, H., Varotto, C. 2008. Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. Cladistics 24:727–745.

Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542–2543.

Lu, F., Lipka, A.E., Glaubitz, J., Elshire, R., Cherney, J.H., Casler, M.D., Buckler, E.S., Costich, D.E. 2013. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. PLoS Genetics 9(1): e1003215.

Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55:21–30.

Maddison, D.R., Maddison, W.P., 2000. MacClade Version 4: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland Massachusetts.

Maddison, W.P., Maddison, D.R. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75. http://mesquiteproject.org

Mardis, E.R. 2008. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics, 9, 387–402.

Mateos, M., Hurtado, L.A., Santamaria, C.A., Leignel, V., Guinot, D. 2012. Molecular Systematics of the Deep-Sea Hydrothermal Vent Endemic Brachyuran Family Bythograeidae: A Comparison of Three Bayesian Species Tree Methods. PLoS ONE 7(3): e32066.

McCormack, J.E., Faircloth, B.C. 2013. Next-generation phylogenetics takes root. Molecular Ecology. 22: 19-21.

McCormack, J.E., Hird, S., Zellmer, A., Carstens, B., Brumfield, R. 2012. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics & Evolution. 66: 526-538.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G.A.B., Kent, J. 2000. Biodiversity hotspots for conservation properties. Nature 403, 853–858.

Nye, T. M. W., P. Lio, and W. R. Gilks. 2006. A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. Bioinformatics 22:117–119.

Nylander, J.A.A., 2004. MrModelTest v2. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Nylander, J. A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology. 53:47–67.

Pagel, M.D. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society 255: 37-45.

Pagel, M., Meade, A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump markov chain monte carlo. The American Naturalist. 167, 808–825.

Rannala, B., Yang, Z., 2008. Phylogenetic inference using whole genomes. Annual Review of Genomics and Human Genetics. 9, 217–231.

Renner, S.S., Ricklefs, R.E. 1995. Dioecy and its correlates in the flowering plants. American Journal of Botany 82: 596–606.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61 (3): 539-542.

Rubin, B.E.R., Ree, R., Moreau, C.S. 2012. Inferring phylogenies from RAD sequence data. PLoS ONE, 7, e33394.

Sanders, K.L., Lee, M.S.Y., Mumpuni, Bertozzi, T., Rasmussen, A.R. 2013. Multilocus phylogeny and recent rapid radiation of the viviparous sea snakes (Elapidae: Hydrophiinae). Molecular Phylogenetics and Evolution 66(3): 575-591.

Shaw, Kerry L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences. 99: 16122-16127.

Shaw, J., E. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Sirupin, C. T. Winder, E. E. Schilling, R. L. Small. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92: 142–166.

Simmons, M.P., Pickett, K.M., Miya, M. 2004. How Meaningful Are Bayesian Support Values? Molecular Biology & Evolution 21(1): 188-199.

Spigler, R.B., Ashman, T.-L., 2012. Gynodioecy to dioecy: are we there yet? Annals of Botany 109, 531–543.

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

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), Version b4. Sinauer Associates, Sunderland, Massachusetts.

Thomson, J.D., Brunet, J. 1990. Hypotheses for the evolution of dioecy in seed plants. Trends in Ecology and Evolution 5: 11-16.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular Ecology 22:787–798.

Webb, C.J. 1979. Breeding systems and the evolution of dioecy in New Zealand Apioid Umbelliferae. Evolution 33(2) 662-672.

White, T.A., Perkins, S.E., Heckel, G., Searle, J.B. 2013. Adaptive evolution during an ongoing range expansion: the invasive bank vole (Myodes glareolus) in Ireland. Molecular Ecology 22(11): 2971-2985.

Wiens, J.J., Engstrom, T.N., Chippindale, P.T. 2006. Rapid diversification, incomplete isolation, and the ‘‘speciation clock’’ in North American salamanders (Genus Plethodon): Testing the hybrid swarm hypothesis of rapid radiation. Evolution 60, 2585–2603.

Vita

The author was born in Metairie, Louisiana. She obtained her Bachelor's degree in Biological Sciences at the University of New Orleans 2012. She joined the graduate program at the University of New Orleans to pursue a Masters degree in Biological Sciences and became a member of Dr. Charles Bell's lab in 2012.