Patterns of floral and chromosome evolution in Onagraceae: a fossil-calibrated supermatrix analysis

William A. Freyman[a,*]

[a]*Jepson Herbarium and Department of Integrative Biology, University of California, Berkeley*

Abstract

Onagraceae is a cosmopolitan plant family in the order Myrtales comprised of 22 genera and 657 species. I used a GenBank data-mining approach to construct an 11 gene supermatrix of 521 taxa in Onagraceae and Lythraceae. Maximum likelihood and Bayesian inference phylogenetic analyses were performed, and divergence time estimates were calibrated using 5 fossils. Chromosome number evolution was inferred using probabilistic models, and ancestral states of petal color and petal number were reconstructed using maximum likelihood. Correlated evolution between petal color and petal number was tested using Bayesian stochastic character mapping. The monophyly of all genera in Onagraceae was supported, and Onagraceae was estimated to have diverged from Lythraceae 109 Mya. The base chromosome number of Onagraceae was inferred to be $x = 5$. Petal color and petal number were found to follow a pattern of correlated evolution ($p = 0.00$), suggesting concerted shifts in floral traits may play an important role in shaping diversification of Onagraceae. Further analyses should be performed to test whether shifts in floral traits are correlated with pollinator shifts, and to determine how these shifts affect diversification rates. Furthermore, this work should be extended to test historical biogeographical hypotheses that explain the intercontinental disjunct distribution of Onagraceae.

[*]Corresponding author
Email address: `freyman@berkeley.edu` (William A. Freyman)

1

1. Introduction

In pursuit of large-scale evolutionary questions, biologists are increasingly using massive phylogenetic datasets to reconstruct ever-growing portions of the tree of life. A common approach to building large phylogenetic trees has been the supermatrix approach, in which multiple datasets are combined and analyzed simultaneously (de Queiroz and Gatesy, 2007). However, assembling and utilizing supermatrices is challenging due to difficulties such as determining homology of molecular sequences, assembling chimeric operational taxonomic units, limiting the amount of missing data, and managing data integrity and computational feasibility (Roure et al., 2007; Yan et al., 2005). Despite these challenges, considerable bioinformatic advances have made supermatrix construction more common, particularly in large-scale evolutionary plant studies. For example, Smith et al. (2011) inferred a phylogeny of 55473 seed plants, Hinchliff and Roalson (2013) estimated the phylogeny of 1500 members of Cyperaceae, and Särkinen et al. (2013) built a time calibrated tree of Solanaceae with 1076 tips.

The plant family Onagraceae is an ideal group to apply a data-mining supermatrix approach due to the over 5000 DNA sequences available in GenBank from previous studies. Onagraceae, the evening primrose family, is a well-studied monophyletic clade of flowering plants with 22 described genera and around 660 species (Wagner et al., 2007). The family is found on all continents except Antartica, and pollen fossils date Onagraceae to the Upper Cretaceous (Grímsson et al., 2012). Onagraceae falls within the order Myrtales, and previous molecular phylogenetic studies find it most closely related to Lythraceae (Conti et al., 1997; Sytsma et al., 2004). Though there have been many genus-level molecular phylogenetic studies in Onagraceae (Berry et al., 2004; Evans et al., 2009; Hoggard and Kores, 2004; Xie et al., 2009; Baum et al., 1994; Wagner and Lammers, 2005), the most recent family-level analyses (Levin et al., 2003, 2004) used only one or two exemplars from some genera. Furthermore, the historical biogeography, divergence times of the major clades, and family-level patterns and processes of diversification remain poorly understood and untested over a

molecular phylogeny.

Here, I use a well-resolved and densely-sampled supermatrix phylogeny of Onagraceae to test the monophyly of the taxonomic groups described in Wagner et al. (2007) and estimate when they diverged. I then inferred the pattern of chromosome number evolution using probabilistic models and examined the patterns of petal color and petal number evolution, including a correlation test of the evolution of these two floral traits. Lastly, I consider ways in which these analyses can be extended to further examine the evolution of Onagraceae, and discuss the novel method used here to construct the phylogenetic supermatrix.

2. Methods

Supermatrix assembly. Lythraceae was selected as an outgroup since previous molecular phylogenetic analyses place it sister to Onagraceae (Conti et al., 1997; Sytsma et al., 2004). I downloaded all DNA sequences from GenBank release 200 PLN division and performed an exhaustive all-by-all BLASTn (Camacho et al., 2009) comparison of sequences in Onagraceae and Lythraceae. Using a BLASTn e-value of 1.0×10^{-10} threshold and a sequence length percent similarity cutoff of 0.5, I constructed clusters of putative homologs using a single-linkage hierarchical clustering algorithm. Subspecies names were removed from all sequences, and all but one sequence of each species was pruned from each cluster. Clusters that were not phylogenetically informative (< 4 taxa) were discarded, and each cluster was aligned using MUSCLE (Edgar, 2004). The alignments were concatenated by species, and any species that was not present in at least two clusters was removed from the supermatrix. The program written to data-mine GenBank and assemble the supermatrix is available as the open source Python module SUMAC (Freyman, 2014). SUMAC will assemble supermatrices for any taxonomic group recognized in GenBank, and is optimized to run on multicore processors and clusters by utilizing multiple parallel processes.

Phylogenetic analyses. Maximum likelihood (ML) analyses were performed with RAxML-HPC (Stamatakis, 2014) on the CIPRES Scientific Gateway (Miller

3

Group	Age (Mya)	Prior Distribution	Mean	SD	Offset	Reference
Circaea (Onagraceae)	12	lognormal	0.0	2.0	12	(Grímsson et al., 2012)
Epilobium (Onagraceae)	12	lognormal	0.0	2.0	12	(Grímsson et al., 2012)
S. Pacific *Fuschia* (Onagraceae)	23	lognormal	0.0	1.0	23	(Lee et al., 2013)
Ludwigia (Onagraceae)	Paleocene	normal	60.0	3.0	-	(Zhi-Chen et al., 2004)
Lythraceae	82	lognormal	0.0	2.0	82	(Graham, 2013)

Table 1: Fossils used as priors in the Bayesian divergence time analysis.

et al., 2010) using the rapid bootstrap heuristic and the GTRCAT nucleotide substitution model. I used the ML tree to select 15 taxa phylogenetically widely distributed in Lythraceae to act as outgroup for the divergence time analysis; all other members of Lythraceae were subsequently removed from the supermatrix. Bayesian estimates of divergence times were inferred using BEAST v1.8 (Drummond and Rambaut, 2007; Suchard and Rambaut, 2009) on CIPRES and calibrated with five fossils identified with morphological synapomorphies (Table 1). The *Ludwigia* fossil pollen was dated broadly to the Paleocene (Grímsson et al., 2012), so I set the prior to a normal distribution with a wide standard deviation to cover the entire time period. For all other calibration points I used a lognormal prior distribution with the offset (the minimum age of the node) corresponding to the fossil age. The BEAST analysis utilized the GTR+Γ nucleotide substitution model with a relaxed molecular clock (uncorrelated lognormal model) and a Yule process tree prior. The Markov Chain Monte Carlo (MCMC) was run for 100 million generations, sampling every 10 thousand generations. Tracer v1.6 (Rambaut et al., 2013) was used to assess the MCMC output for parameter convergence and ensure that the effective sample size for all parameters was above 200. The first 1000 trees were discarded as burn-in, and the remaining 9000 trees were summarized as a maximum clade credibility (MCC) tree with mean divergence times.

Character state reconstruction. I scored six characters, including chromosome number, floral merosity, petal color, and self-compatibility/incompatibility. Character data was assembled from the comprehensive Wagner et al. (2007) Ona-

graceae monograph. Ancestral chromosome numbers were inferred using maximum likelihood and Bayesian methods as implemented in ChromEvol 2.0 (Glick and Mayrose, 2014). Eight different models of chromosome evolution were fit to the Bayesian MCC phylogeny using ChromEvol, and the best fit model was selected using Akaike's information criterion (Akaike, 1981). Ancestral character state reconstructions of petal number and petal color were performed using Mesquite v2.75 (Maddison and Maddison, 2011) over the Bayesian MCC tree. Characters were treated as unordered categorical data, and optimized using maximum likelihood with the Markov k-state 1 parameter (Mk1) model (Lewis, 2001). Additionally, Bayesian stochastic character mapping (Huelsenbeck et al., 2003) was used to test whether petal color and petal number covaried over the phylogeny. Models of petal color and petal number evolution were configured in the program SIMMAP v1.5 (Bollback, 2006) using unordered states, gamma rate priors, and equal state bias priors. The correlation analysis used SIMMAP's default predictive sampling configuration and calculated the following correlation statistics: D the overall association between the two characters, and d_{ij} the association between the individual states of each character (Huelsenbeck et al., 2003).

3. Results

Supermatrix assembly. SUMAC evaluated 5571 Onagraceae and 2832 Lythraceae nucleotide sequences to construct the supermatrix. The completed supermatrix consisted of 11 clusters of homologous sequences (Table 2). As used in the maximum likelihood analyses (before pruning the number of outgroup taxa), the supermatrix contained 521 taxa, was 31862 nucleotides long, and contained 93.0% missing data.

Phylogeny and divergence time estimates. The topologies of the ML and Bayesian phylogenies were nearly identical for all major clades within Onagraceae, so only the Bayesian MCC tree (Figures 1 and 2) is shown here. The only exception is the tribe Epilobieae *sensu* Wagner et al. (2007); the ML analysis recovered

DNA Region	# of Taxa	Aligned Length	Missing data (%)	Taxon Coverage Density
ITS	453	1746	13.2	0.87
trnL	234	1429	55.2	0.45
rpl16	91	1414	82.6	0.17
rbcL	77	1474	85.2	0.15
rps16	74	1016	85.8	0.14
rbcL2	64	1310	87.7	0.12
PgiC2	47	4028	91.0	0.09
matK	37	921	92.9	0.07
ndhF	37	2063	92.9	0.07
pgiC	26	14709	95.0	0.05
R5	18	3129	96.6	0.03

Table 2: Clusters of homologous sequences used to assemble the supermatrix.

tribe Epilobieae as monophyletic, whereas according to the Bayesian estimate it is paraphyletic. All Onagraceae genera described in Wagner et al. (2007) were recovered as monophyletic clades in the Bayesian MCC with posterior probabilities > 0.95, except for sister genera *Neoholmgrenia* and *Camissoniopsis* (posterior $= 0.31$) (Figure 1). Onagraceae was found to diverge from Lythraceae at 109 Mya (Figure 2). Divergence time estimates of other major clades and 95% highest posterior density (HPD) intervals can be seen in Table 3.

Chromosome number evolution. Out of the eight models of chromosome evolution tested by ChromEvol, the four parameter $CONST\_RATE\_DEMI\_EST$ model was the best fit according to the AIC scores. This model has separate rate parameters for chromosome gains (λ), losses (δ), demi-polyploid events (μ), and polyploid events (ρ). Demi-polyploid events occur when reduced and unreduced gametes unite resulting in a $3x$ cytotype. The base chromosome number of Onagraceae was estimated to be $x = 5$, and the ancestral number of the large genus *Epilobium* was estimated to be $n = 9$, though the extant taxa are split in two large monophyletic clades of $n = 10$ and $n = 18$, respectively (Figure 3).

6

Clade	Mean Age (Mya)	95% HPD Min	95% HPD Max
Onagraceae / Lythraceae	109	88	131
Ludwigia	97	76	118
Hauya	49	35	64
Circaea / Fuchshia	37	28	47
Lopezia	71	55	68
Gongylocarpus	60	45	77
Epilobium	49	38	60
Chamerion	47	36	57
Xylonagra	43	33	52
Clarkia	40	32	48
Terapteron	19	10	29
Camissoniopsis / Neoholmgrenia	14	5	23
Eremothera / Camissonia	24	16	33
Taraxia	30	22	38
Chylismiella / Gayophytum	20	10	30
Eulobus	26	19	34
Chylismia / Oenothera	25	18	31

Table 3: Bayesian divergence time estimates of major clades.

	Number of Petals		
	2	4	5
Petal Color			
Pink	-0.005	.011	ns
Yellow	-0.008	.021	ns
White	0.008	0.013	-0.006
Green	ns	-0.011	ns
Red	ns	ns	ns

Table 4: Test statistics for the correlation between petal color evolution and petal number evolution. d_{ij} values are shown for the pairwise comparison of states with $p < 0.01$. ns indicates no significant association. Negative d_{ij} values express negative correlation. The overall D value was 0.263 ($p = 0.00$).

Floral evolution. The maximum likelihood estimate of petal color for the last common ancestor of Onagraceae was pink with a proportional likelihood (PL) of 0.56 (Figure 4). White was the second most likely estimate ($PL = 0.19$). The maximum likelihood estimate of petal number for the last common ancestor of Onagraceae was 4 ($PL = 1.0$), with 8 independent transitions to 0, 2, or 5 petals (Figure 5). Petal color and petal number were found to significantly covary over the phylogeny ($D = 0.263, p = 0.00$). Multiple significant correlations (d_{ij}) were found between individual states of each character and are summarized in Table 4.

4. Discussion

The paraphyly of tribe Epilobieae in the Bayesian MCC tree (Figure 2) conflicts with both the maximum likelihood analysis (not shown) and the previous studies of Levin et al. (2003, 2004), all of which found tribe Epilobieae to be monophyletic. This incongruence is likely due to the placement of the fossil constraint on the genus *Epilobium* used in the Bayesian analysis. In this study I followed the taxonomy of Wagner et al. (2007), which splits tribe Epilobieae
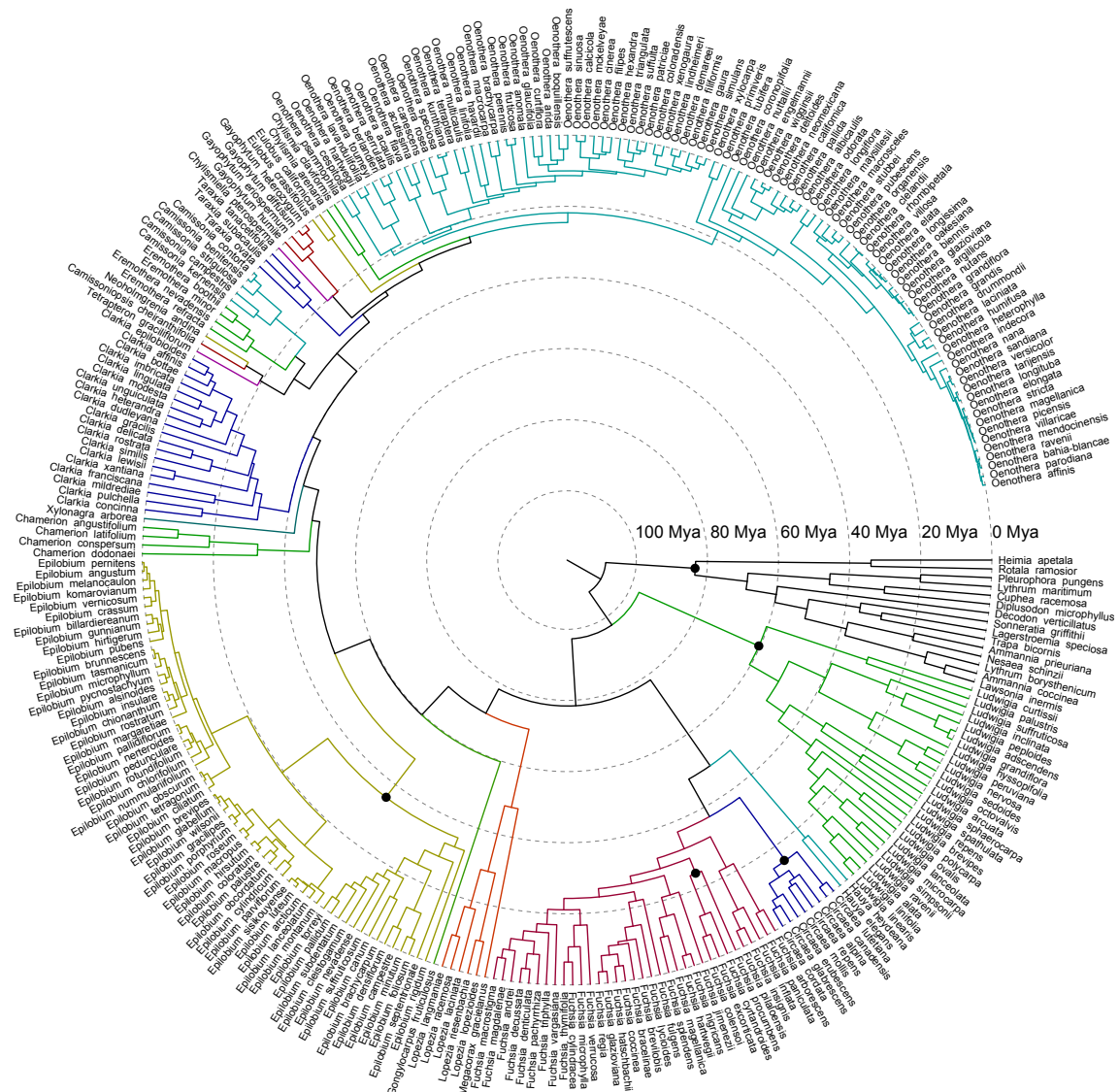
8

Figure 1: Bayesian maximum clade credibility phylogeny of 280 Onagraceae taxa and 15 Lythraceae taxa. Estimated posterior probabilities close to 1.0 are shown in green. All genera described in Wagner et al. (2007) were found to be monophyletic with posterior probabilities of > 0.95 except for sister genera *Neoholmgrenia* and *Camissoniopsis* (posterior = 0.31).

9

Figure 2: Bayesian chronogram of 280 Onagraceae taxa and 15 Lythraceae taxa. Approximate positions of fossil calibration points are shown as black circles. All genera described in Wagner et al. (2007) are colored, and their divergence time estimates and %95 HPD intervals can be seen in Table 3.
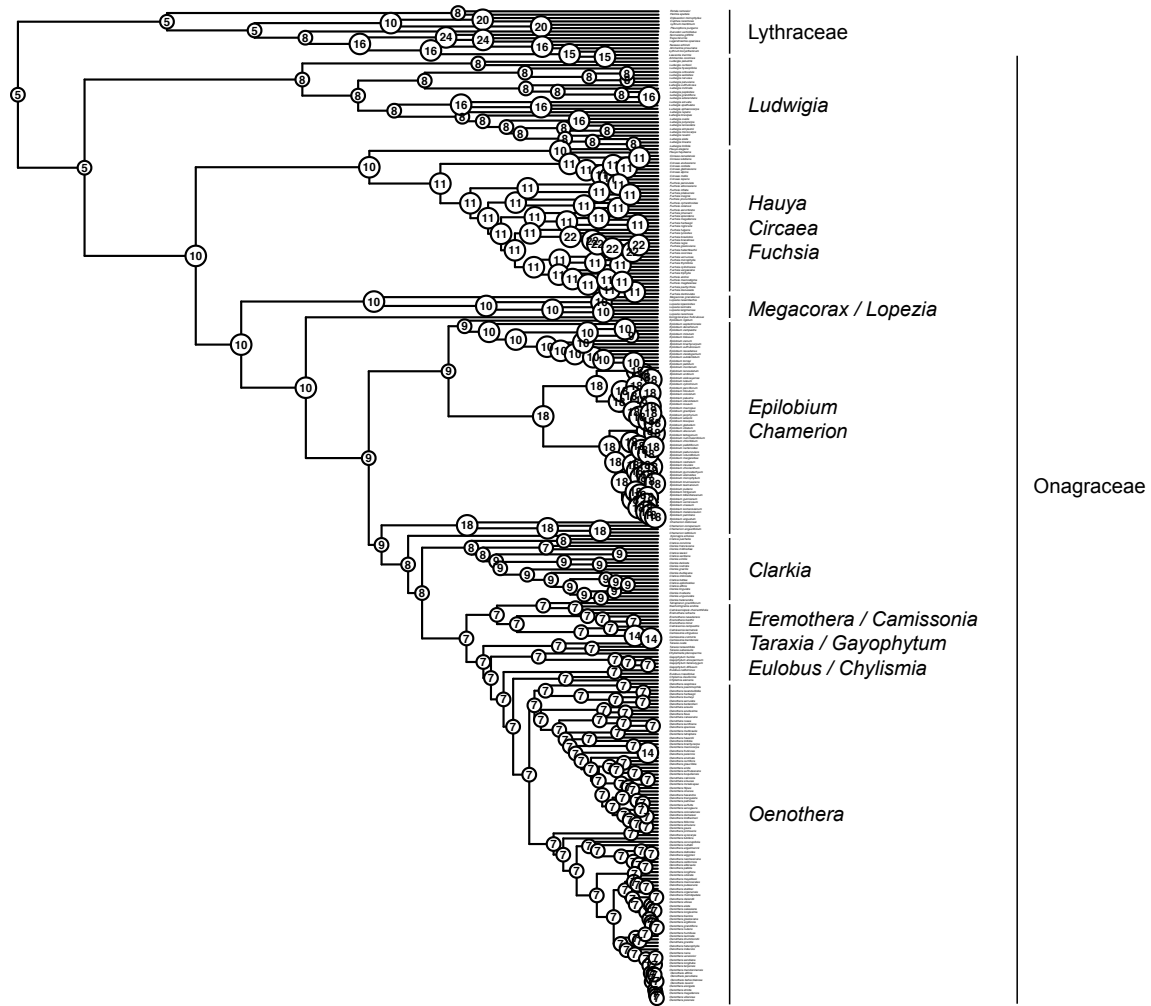
Figure 3: Maximum likelihood estimates of ancestral chromosome numbers over the Bayesian MCC phylogeny using the 4 parameter $(\lambda, \delta, \mu, \rho)$ model of chromosome evolution selected by Akaike's information criterion. The base number of Onagraceae is inferred to be $x = 5$.
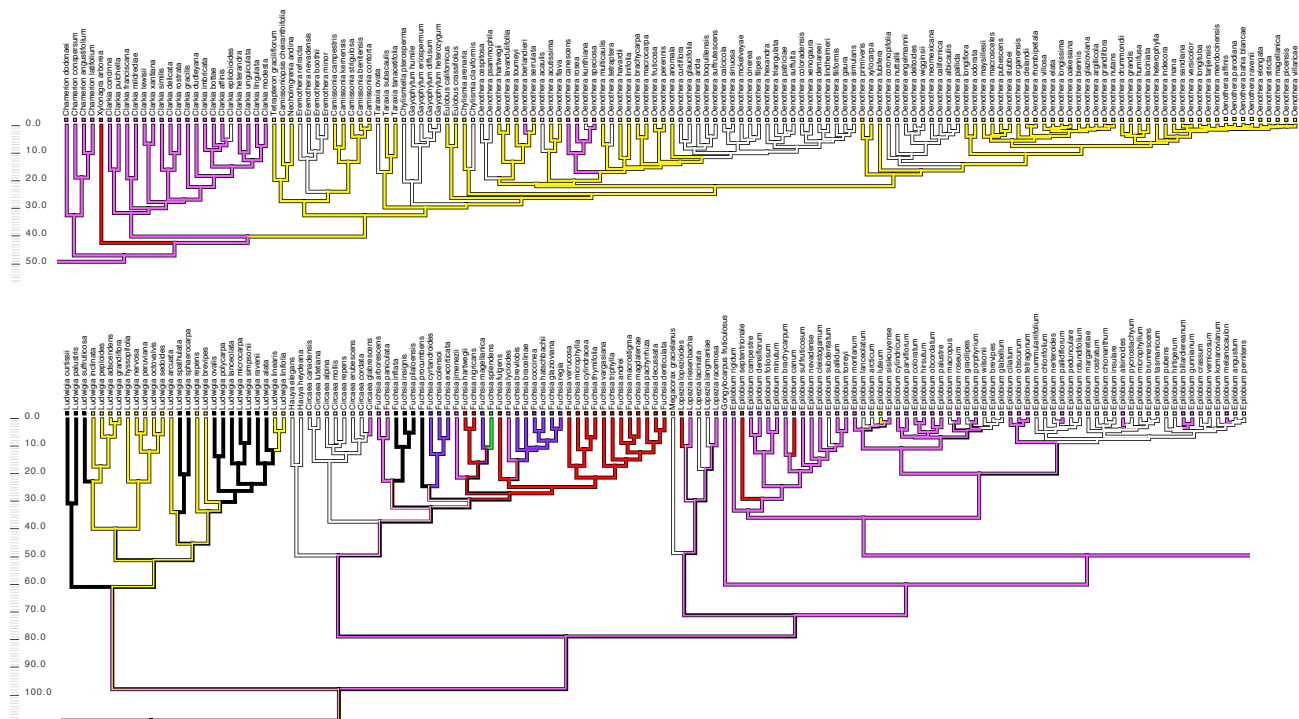
Figure 4: Maximum likelihood ancestral state reconstruction of petal color. Black represents absence of petals. Vertical time scale is in millions of years before present.
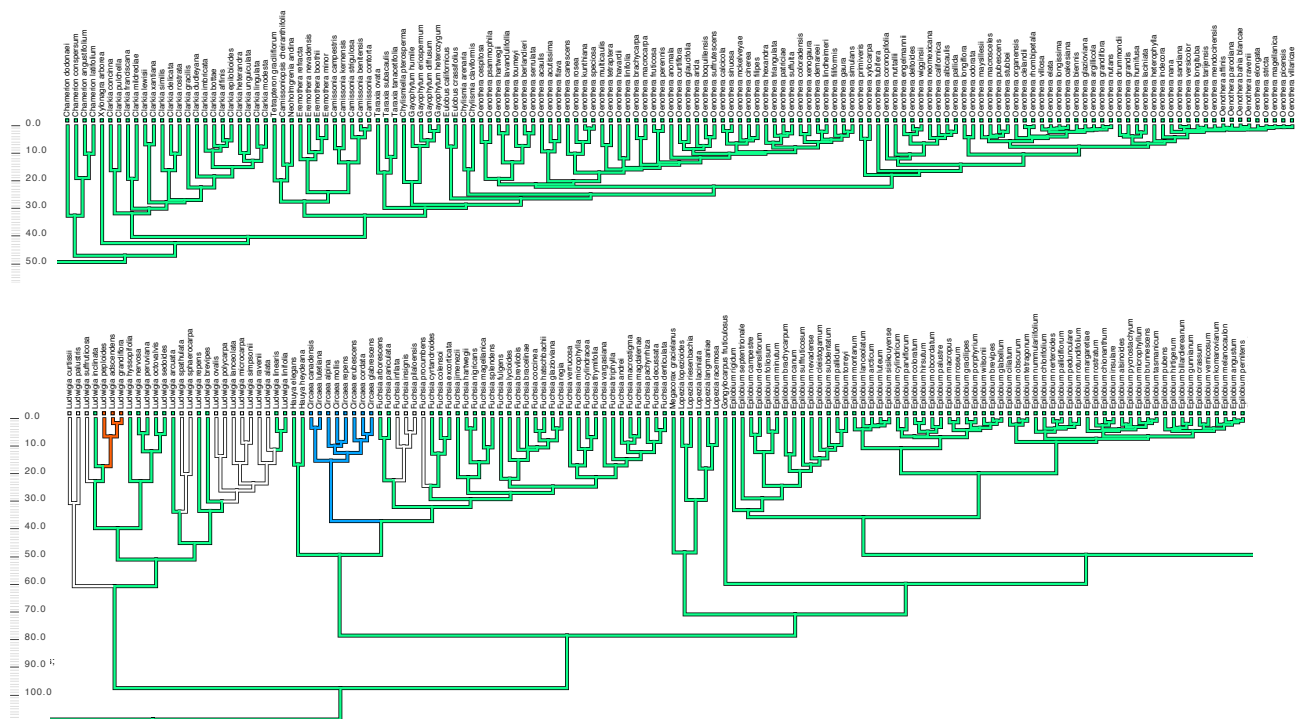
Figure 5: Maximum likelihood ancestral state reconstructions of petal number. White = 0, Blue = 2, Green = 4, Orange = 5. Vertical time scale is in millions of years before present.

13

into the sister genera *Epilobium* and *Chamerion*. However, the pollen morphology descriptions used by Grímsson et al. (2012) to identify the *Epilobium* fossil did not recognize *Chamerion* as separate from *Epilobium* (Punt et al., 2003; Patel et al., 1984). The divergence times will be reestimated using the *Epilobium* fossil to constrain both genera. This will likely result in topological changes to the Bayesian MCC, which means that the trait evolution estimates presented here should be considered preliminary.

The reconstruction of ancestral chromosome numbers validates many estimates made by Raven (1979). Raven's estimates of the base chromosome number for the genera *Ludwigia*, *Fuchsia*, *Circaea*, *Hauya*, *Epilobium*, *Camissonia*, and *Oenothera* all agree with the reconstruction presented here (Figure 3). However, for two genera I report estimates that differ from Raven's: for *Clarkia* $x = 9$ and for *Lopezia* $x = 10$. Furthermore, Raven (1979) proposed that the ancestral chromosome number of all Onagraceae was $x = 11$, whereas my estimate suggests $x = 5$. Further testing with the ChromEvol models should be carried out, particularly to see if the inclusion of the outgroup (Lythraceae) affected the estimate of the base chromosome number of Onagraceae.

The analysis of petal color and petal number evolution using Bayesian stochastic character mapping revealed that the two floral traits followed a pattern of correlated evolution (p = 0.00), suggesting concerted shifts in floral traits may play an important role in shaping diversification of Onagraceae. To test this, further analyses will be performed to determine if these shifts in floral morphology affect diversification rates. Other floral traits, such as corolla width and diurnal or nocturnal flower opening, could be included in future analyses. Tests will also be performed to evaluate whether shifts in floral traits are correlated with pollinator shifts. These analyses could provide evidence that pollinators are driving the evolution of floral morphology in Onagraceae (DeWitt Smith, 2010).

This work will also be extended to test historical biogeographical hypotheses that explain the intercontinental disjunct distribution of Onagraceae. The family-level supermatrix phylogeny will be used to infer the location of the last

common ancestor of Onagraceae, which has been hypothesized to have been in South America (Raven and Axelrod, 1974; Raven, 1988). The biogeographical analysis will also test the suggestion made by Berry et al. (2004) that *Fuchsia* originated in Central America. To my knowledge this will be the first reconstruction of the historical biogeography of Onagraceae over a molecular phylogeny.

The strength of the supermatrix approach to phylogenetic estimation is that it combines data from diverse sources into one large analysis. This strength is amplified by data-mining GenBank in an exploratory fashion, so that sequence data not necessarily used in previous systematic studies can also be incorporated into the supermatrix. A number of software tools for building supermatrices are already available that use methods similar to those performed here. For example, the PhyLoTA Browser provides a web interface to view all GenBank sequences within taxonomic groups clustered into homologs (Sanderson et al., 2008). A different approach is implemented in the program PHLAWD, which constructs clusters of homologs around guide sequences provided by the user (Smith et al., 2009). However, the method used here (implemented in SUMAC) combines elements of both approaches; the user can perform an exploratory clustering of all GenBank sequences within a taxonomic group, and also provide guide sequences to build homologous clusters in a more targeted manner. SUMAC is being extended to calculate metrics that assess the informativeness of phylogenetic supermatrices that are unavailable in other software tools. For example, the next version of SUMAC will calculate an index of substitution saturation based on information entropy (Xia et al., 2003) for each alignment, and also calculate the partial phylogenetic decisiveness of the supermatrix, which measures the percentage of trees (out of all possible trees) that can be inferred given the data missing from the supermatrix (Sanderson et al., 2010; Fischer, 2012). By calculating these metrics SUMAC will provide a unique framework with which GenBank can be repeatedly mined using different settings and the resulting data matrices can be compared. With further methodological refinement, supermatrix approaches will be indispensable to future large-scale phylogenetic studies.

References

Akaike, H., 1981. Likelihood of a model and information criteria. Journal of econometrics 16, 3–14.

Baum, D., Sytsma, K., Hoch, P., 1994. A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA sequences. Systematic Botany 19, 363–388.

Berry, P.E., Hahn, W.J., Sytsma, K.J., Hall, J.C., Mast, A., 2004. Phylogenetic relationships and biogeography of *Fuchsia* (Onagraceae) based on noncoding nuclear and chloroplast DNA data. American Journal of Botany 91, 601–14.

Bollback, J.P., 2006. Simmap: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7, 88.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Conti, E., Litt, A., Wilson, P.G., Graham, S.A., Briggs, B.G., Johnson, L., Sytsma, K.J., 1997. Interfamilial relationships in Myrtales: molecular phylogeny and patterns of morphological evolution. Systematic Botany 22, 629–647.

DeWitt Smith, S., 2010. Using phylogenetics to detect pollinator-mediated floral evolution. The New Phytologist 188, 354–63.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797.

Evans, M.E.K., Smith, S.a., Flynn, R.S., Donoghue, M.J., 2009. Climate, niche evolution, and diversification of the "bird-cage" evening primroses

16

(*Oenothera*, sections Anogra and Kleinia). The American Naturalist 173, 225–40.

Fischer, M., 2012. Perfect taxon sampling and phylogenetically decisive taxon coverage. arXiv preprint arXiv:1206.3472 .

Freyman, W.A., 2014. Supermatrix Constructor (SUMAC): a Python package for data mining GenBank and building phylogenetic supermatrices. `https://github.com/wf8/sumac`.

Glick, L., Mayrose, I., 2014. Chromevol: Assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Molecular biology and evolution , msu122.

Graham, S.A., 2013. Fossil records in the Lythraceae. The Botanical Review 79, 48–145.

Grímsson, F., Zetter, R., Leng, Q., 2012. Diverse fossil Onagraceae pollen from a Miocene palynoflora of north-east China: early steps in resolving the phytogeographic history of the family. Plant Systematics and Evolution 298, 671–687.

Hinchliff, C.E., Roalson, E.H., 2013. Using supermatrices for phylogenetic inquiry: an example using the sedges. Systematic Biology 62, 205–19.

Hoggard, G., Kores, P., 2004. The phylogeny of *G*aura (Onagraceae) based on ITS, ETS, and trnL-F sequence data. American Journal of Botany 91, 139–148.

Huelsenbeck, J.P., Nielsen, R., Bollback, J.P., 2003. Stochastic Mapping of Morphological Characters. Systematic Biology 52, 131–158.

Lee, D.E., Conran, J.G., Bannister, J.M., Kaulfuss, U., Mildenhall, D.C., 2013. A fossil *Fuchsia* (Onagraceae) flower and an anther mass with in situ pollen from the early Miocene of New Zealand. American Journal of Botany 100, 2052–2065.

Levin, R., Wagner, W., Hoch, P., 2004. Paraphyly in tribe Onagreae: insights into phylogenetic relationships of Onagraceae based on nuclear and chloroplast sequence data. Systematic Botany 29, 147–164.

Levin, R.a., Wagner, W.L., Hoch, P.C., Nepokroeff, M., Pires, J.C., Zimmer, E.a., Sytsma, K.J., 2003. Family-level relationships of Onagraceae based on chloroplast rbcL and ndhF data. American Journal of Botany 90, 107–15.

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

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

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees, in: Gateway Computing Environments Workshop (GCE), 2010, IEEE. pp. 1–8.

Patel, V., Skvarla, J., Raven, P., 1984. Pollen Characters in Relation to the Delimitation of Myrtales. Annals of the Missouri Botanical Garden 71, 858–969.

Punt, W., Rovers, J., Hoen, P., 2003. Onagraceae. Review of Palaeobotany and Palynology 123, 107–161.

de Queiroz, A., Gatesy, J., 2007. The supermatrix approach to systematics. Trends in Ecology & Evolution 22, 34–41.

Rambaut, A., Suchard, M., Drummond, A.J., 2013. Tracer v1.6. http://tree.bio.ed.ac.uk/software/tracer/.

Raven, P.H., 1979. A survey of reproductive biology in Onagraceae. New Zealand Journal of Botany 17, 575–593.

Raven, P.H., 1988. Onagraceae as a model of plant evolution, in: Plant evolutionary biology. Springer, pp. 85–107.

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

Roure, B., Rodriguez-Ezpeleta, N., Philippe, H., 2007. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evolutionary Biology 7 Suppl 1, S2.

Sanderson, M.J., Boss, D., Chen, D., Cranston, K.a., Wehe, A., 2008. The PhyLoTA Browser: processing GenBank for molecular phylogenetics research. Systematic Biology 57, 335–46.

Sanderson, M.J., McMahon, M.M., Steel, M., 2010. Phylogenomics with incomplete taxon coverage: the limits to inference. BMC evolutionary biology 10, 155.

Särkinen, T., Bohs, L., Olmstead, R.G., Knapp, S., 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evolutionary Biology 13, 214.

Smith, S.a., Beaulieu, J.M., Donoghue, M.J., 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evolutionary Biology 9, 37.

Smith, S.a., Beaulieu, J.M., Stamatakis, A., Donoghue, M.J., 2011. Understanding angiosperm diversification using small and large phylogenetic trees. American Journal of Botany 98, 404–14.

Stamatakis, A., 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics .

Suchard, M.A., Rambaut, A., 2009. Many-core algorithms for statistical phylogenetics. Bioinformatics 25, 1370–1376.

Sytsma, K., Litt, A., Zjhra, M., 2004. Clades, clocks, and continents: Historical and biogeographical analysis of Myrtaceae, Vochysiaceae, and relatives in the

southern Hemisphere source. International Journal of Plant Sciences 165, S85–S105.

Wagner, W., Lammers, T., 2005. Systematics of *Oenothera* Sections Contortae, Eremia, and Ravenia (Onagraceae). Systematic Botany 30, 332–355.

Wagner, W.L., Hoch, P.C., Raven, P.H., 2007. Revised classification of the Onagraceae. Systematic Botany Monographs 83.

Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitution saturation and its application. Molecular phylogenetics and evolution 26, 1–7.

Xie, L., Wagner, W.L., Ree, R.H., Berry, P.E., Wen, J., 2009. Molecular phylogeny, divergence time estimates, and historical biogeography of *Circaea* (Onagraceae) in the Northern Hemisphere. Molecular Phylogenetics and Evolution 53, 995–1009.

Yan, C., Burleigh, J.G., Eulenstein, O., 2005. Identifying optimal incomplete phylogenetic data sets from sequence databases. Molecular phylogenetics and evolution 35, 528–35.

Zhi-Chen, S., Wei-Ming, W., Fei, H., 2004. Fossil pollen records of extant angiosperms in China. The Botanical Review 70, 425–458.