Fossil-calibrated phylogeny of Onagraceae: patterns of floral and chromosome evolution

William A. Freyman[a,*]

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

Abstract

blah blah

1. Introduction

blah blah

2. Methods

Supermatrix assembly. I downloaded all available DNA sequences from Gen-Bank release 200 PLN division and performed an exhaustive all-by-all BLASTn (Camacho et al., 2009) comparison for sequences in Onagraceae and Lythraceae. Using a BLASTn e-value of $1.0e^{-10}$ 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. Code written to assemble the supermatrix is available as the Python module SUMAC (Freyman,

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

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.

2014, `https://github.com/wf8/sumac`), and can be used to assemble super-matrices for other taxonomic groups recognized in GenBank.

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

3. Results

Supermatrix assembly. blah blah

Divergence time estimates. Here is a table with the mean divergence time estimates including %95 confidence intervals....

Character evolution. blah blah

4. Conclusion

blah blah

References

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

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.

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 2: Bayesian divergence time estimates of major clades.
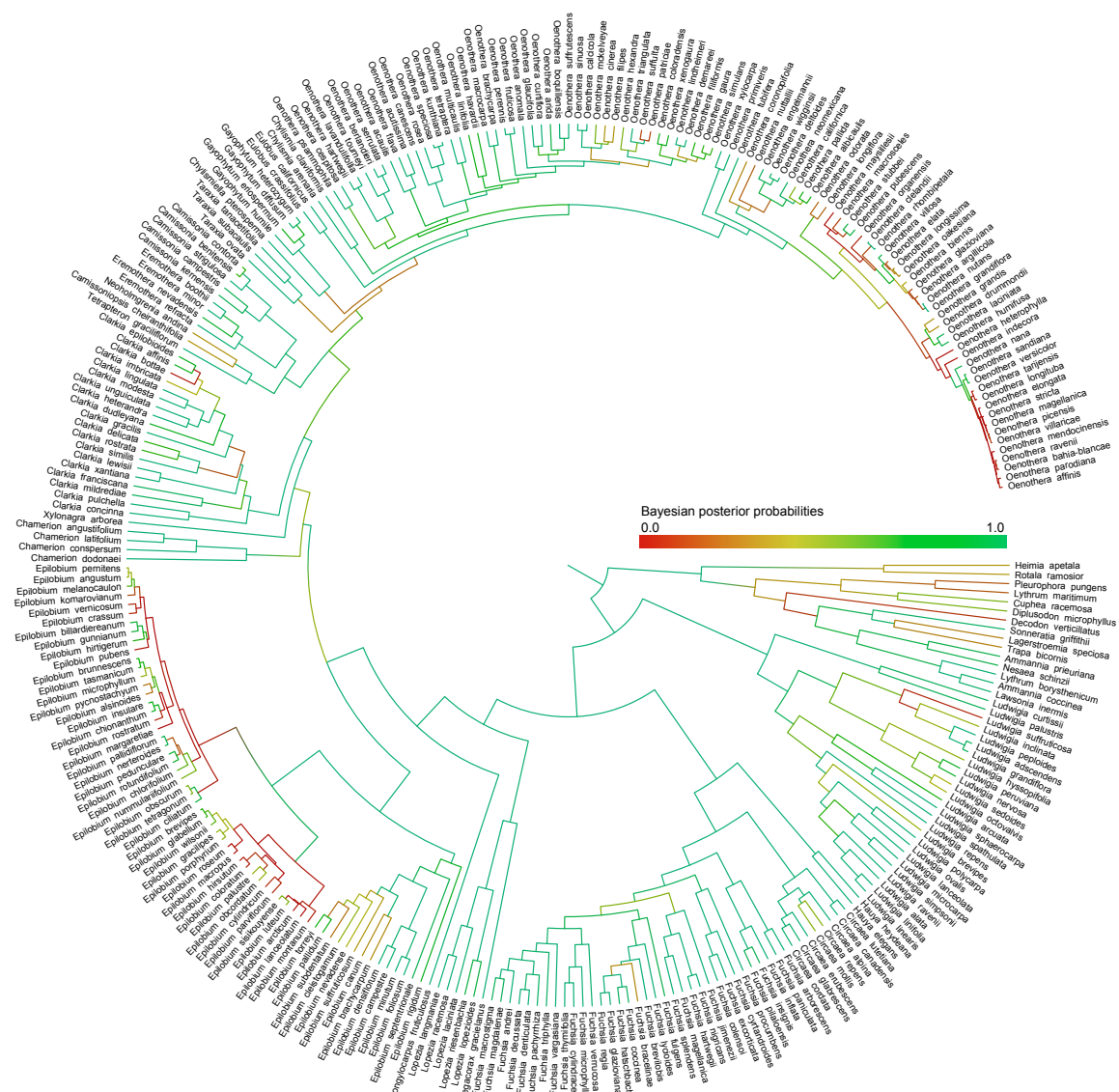
Figure 1: Bayesian maximum clade credibility phylogeny of 280 Onagraceae taxa and 15 Lythraceae taxa. Estimated posterior probabilities close to 1.0 are shown in green. All genera described in Wagner et al. (2007) are monophyletic with posterior probabilities of > 0.95 except for sister genera *Neoholmgrenia* and *Camissoniopsis* (posterior = 0.31).
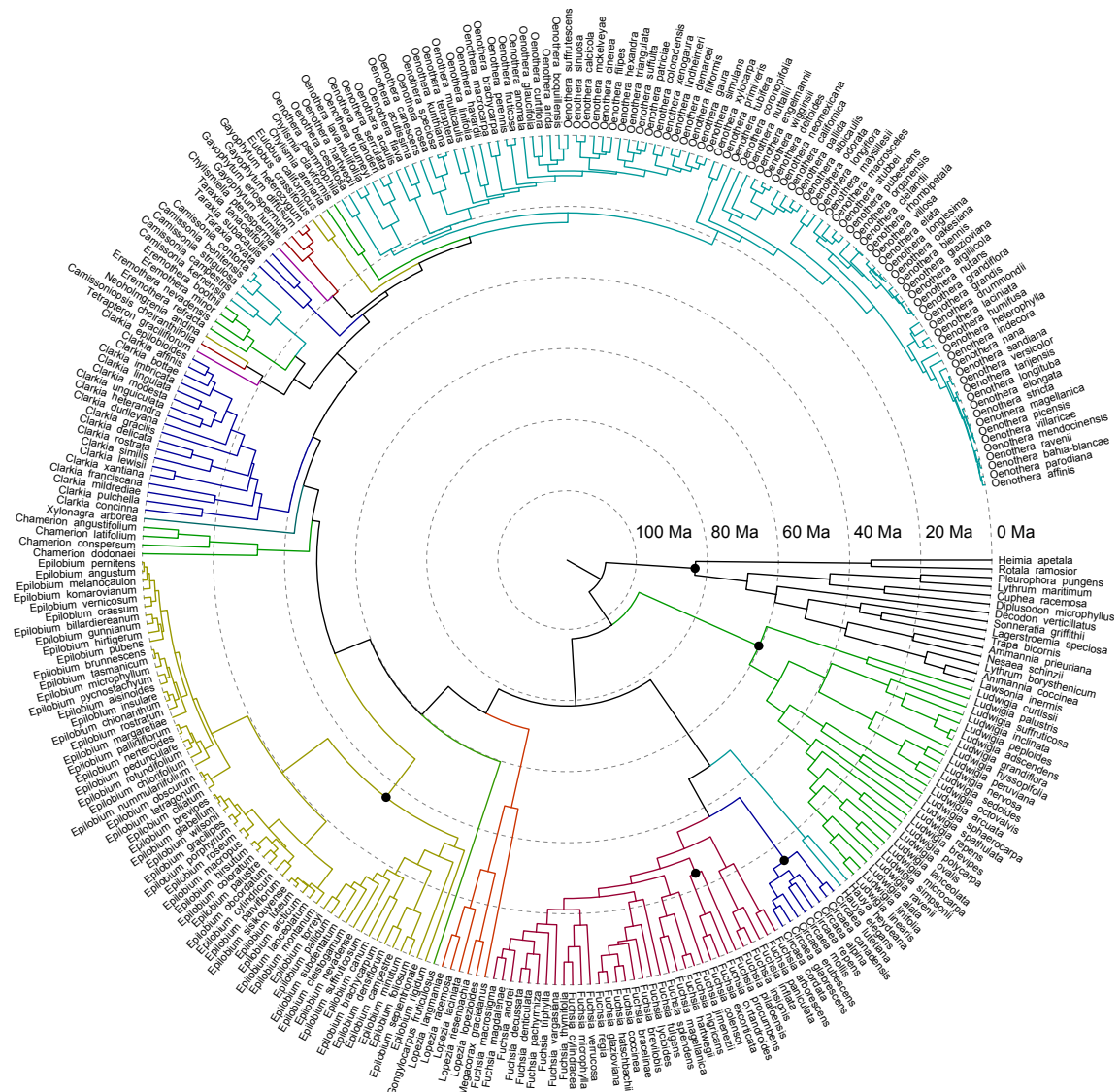
5

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 2.

Freyman, W.A., 2014. SUMAC: Supermatrix Constructor. `https://github.com/wf8/sumac`.

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.

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.

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.

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

Stamatakis, A., 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics , btu033.

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

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

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