

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
5 (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 sim-
ilarity cutoff of 0.5, I constructed clusters of putative homologs using a single-
linkage hierarchical clustering algorithm. Subspecies names were removed from
10 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
15 assemble the supermatrix is available as the Python module Supermatrix Con-
structor (SUMAC) (Freyman, 2014, <https://github.com/wf8/sumac>).

*Corresponding author

Email address: freyman@berkeley.edu (William A. Freyman)

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

Table 1: Fossil priors used in BEAST analyses.

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 species phylogenetically widely dis-
tributed in Lythraceae to act as outgroup; all other members of Lythraceae were
subsequently removed from the supermatrix before estimating divergence times.
Bayesian estimates of divergence times were inferred using BEAST (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 devi-
ation 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+I
nucleotide substitution model with a relaxed molecular clock (uncorrelated log-
normal model) and a Yule process tree prior. The Markov Chain Monte Carlo
(MCMC) was run for 100 million generations, sampling every 10 thousand gen-
erations. 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
tree with mean divergence times.

Character evolution. blah blah

40 3. Results

Supermatrix assembly. blah blah

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

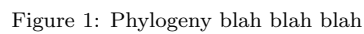
Character evolution. blah blah

45 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
50 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.
- 55 Freyman, W.A., 2014. Supermatrix Constructor (SUMAC). <https://github.com/wf8/sumac>.
- Graham, S.A., 2013. Fossil records in the lythraceae. The Botanical Review 79, 48–145.
- 60 Grímsson, F., Zetter, R., Leng, Q., 2012. Diverse fossil onagraceae pollen from a miocene palynoflora of north-east china: early steps in resolving the phyto-geographic history of the family. Plant Systematics and Evolution 298, 671–687.



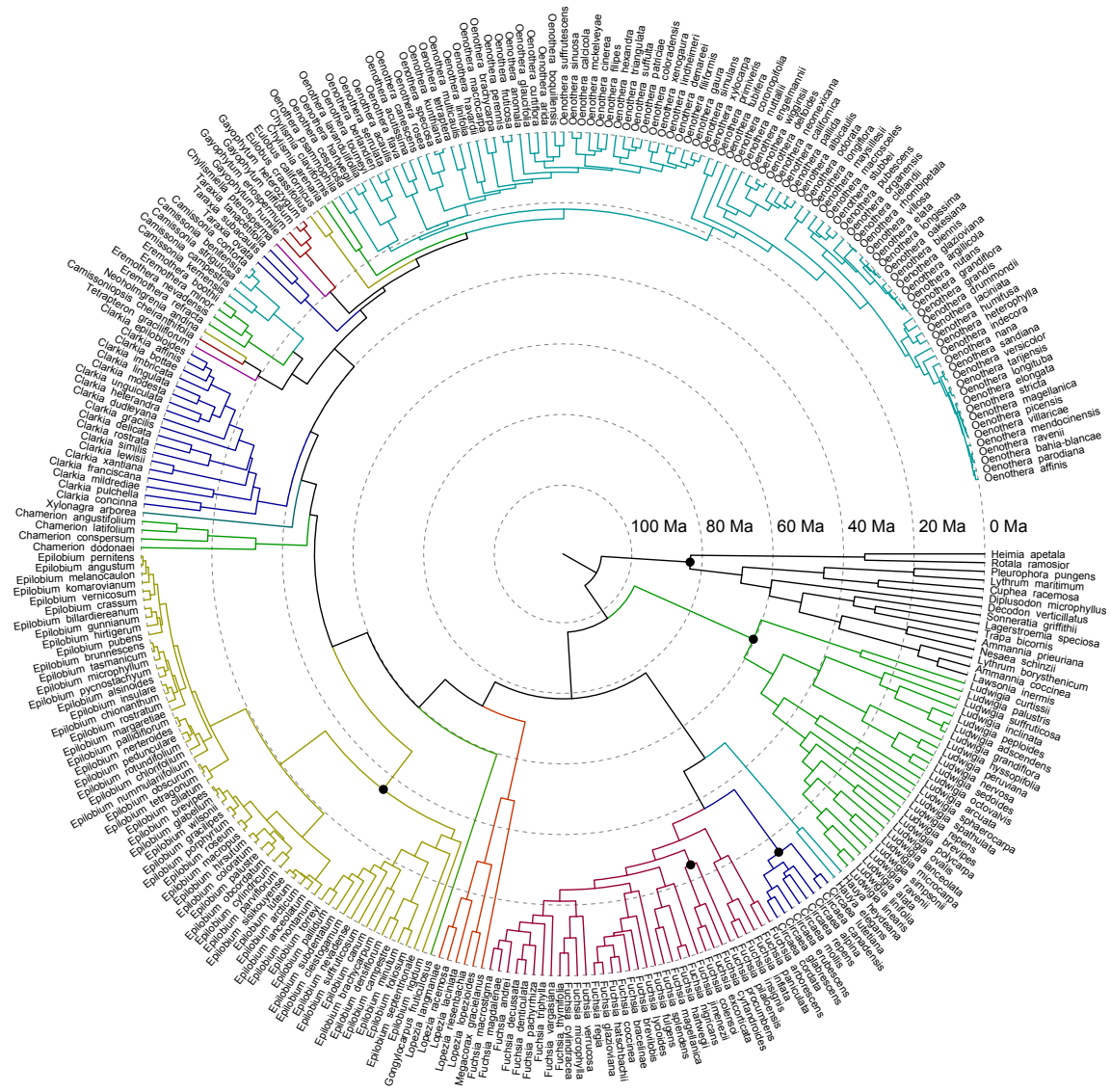


Figure 2: Sequence diagram

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