

**When your favorite trait is not enough to explain diversification: An example of diversification
linked to polyploidy and breeding system in Solanaceae**

Rosana Zenil-Ferguson,¹ J. Gordon Burleigh,² William A. Freyman,³ Boris Igić,⁴ Itay Mayrose,⁵ and Emma
E. Goldberg⁶

¹University of Minnesota Twin Cities

²University of Florida

³23 and me

⁴University of Illinois

⁵Tel Aviv University

⁶University of Minnesota Twin Cities

Author for correspondence: Rosana Zenil-Ferguson

Running head: Polyploidy and breeding systems Solanaceae

Keywords: Polyploidy, Breeding System, Diversification, SSE models

Abstract

Abstract text...

Introduction

Linking changes in diversification to polyploidy and breeding system has been one of the key botany questions for the past n number of years. The prevalence of polyploidy, and the multiple evolutionary changes of breeding systems in important angiosperm clades, leads to the inevitable question of the evolutionary changes.

- Diversification and polyploidy

An important debate regarding the diversification of angiosperms has been ongoing since the publication of Mayrose et al. (2011). The authors discovered that using the latest diversification model linked to two states diploid and polyploid, the net diversification of polyploids was much slower than the net diversification rate than polyploids. This result was surprising and an sparked a discussion about the long-term evolutionary consequences of polyploidy. Soltis et al. (2014) questioned if polyploidy should be regarded as an evolutionary dead-end since the net diversification of polyploids was negative, despite overwhelming evidence about the incidence of polyploidy especially at the root of highly diverse angiosperm claims (ref here). A year later, diversification models were re-tested and corrected, and still found the same pattern (Mayrose et al. 2015), to the disappointment of plenty of botanists there was no denying of the weak trend of diversification that polyploids left behind.

There are two key questions that at the time of the polyploidy debate were difficult to ask. The first is if the models used to measure the diversification of polyploids were correct, and the second question is if the models have potentially included more evidence and potential traits that are not polyploidy or other lines of evidence that could be driving the patterns. At the time, in a different context Beaulieu and O'Meara (2016) were finding an alternative solution to the first question, coming up with a new model that could represent the broad heterogeneity of the diversification process and parse out the signal between the trait of interest and the noise in diversification. Their model, the hidden state speciation and extinction model is a key component to detect whether polyploidy our something else unknown but related to it is driving the speciation and extinction patterns that we see in angiosperms.

The second question (talk here about breeding system and polyploidy, and how this could be different across clades and this is one of the reasons why we focused on Solanaceae). -Diversification and breeding systems

Goldberg and Igic 2012, different perspectives

-What we did

35 We investigate the relative importance of both b

Debate of Mayrose

Methods

Data

Chromosome number data were obtained for all Solanaceae taxa in the the Chromosome Counts Database
40 (CCDB; Rice et al. 2015), and the ca. 14,000 records were cleaned semi-automatically using the CCDBcu-
rator R package (Zenil-Ferguson et al. 2017). This large dataset includes the compilation of Solanaceae
ploidy states from (Robertson et al. 2011). Species were coded as either diploid (D) or polyploid (P). For
the majority of species, ploidy was assigned according to information from the original publications and the
Kew Royal Botanic Gardens C-value DNA resource (Bennett and Leitch 2005). For taxa without ploidy in-
45 formation but with information about chromosome number, we assigned ploidy based on the multiplicity of
chromosomes within the genus. For example, *Solanum betaceum* did not include information about ploidy
level but it has 24 chromosomes, so because $x = 12$ is the base chromosome number of the *Solanum* genus
(Olmstead and Bohs 2007), we assigned *S. betaceum* as diploid. Species with more than one ploidy level
were assigned the smallest most frequent ploidy level recorded.

50 Breeding system was scored as self-incompatible (SI) or self-compatible (SC) based on results hand-
curated from the literature (as in Igić et al. 2006; Goldberg et al. 2010; Robertson et al. 2011; Goldberg and
Igić 2012). Most species could unambiguously be coded as either SI or SC (Raduski et al. 2012). Following
previous work, we coded as SI any species with functional SI systems, even if SC or dioecy was also
reported. Dioecious species without functional SI were coded as SC.

55 To those existing data sets, we added some additional records for chromosome number and breeding
system. The Supplementary Information contains citations for the numerous original sources for all of the
data. Synonymy followed Solanaceae Source (PBI *Solanum* Project 2012). Hybrids and cultivars were
excluded because ploidy and breeding system are likely to be altered by domestication. As in Robertson
et al. (2011), we examined closely the few species for which the merged ploidy and breeding system data
60 indicated the presence of SI polyploids. Although SI populations frequently contain some SC individuals
and diploid populations frequently contain some polyploid individuals, in no case did we find individuals
that were both SI and polyploid. Because of this empirical observation and the functional explanation for
whole genome duplication disabling gametophytic self-incompatibility (reviewed in Ramsey and Schemske
1998; Stone 2002), we consider only three observed character states: SI-D, SC-D, and SC-P.

Matching our character state data to the largest time-calibrated phylogeny of Solanaceae (Särkinen et al. 2013) yielded 595 species with ploidy and/or breeding system information on the tree. Binary or three-state classification of ploidy and breeding system for the 595 taxa is summarized in Table 1. We retained all of these species in each of the analyses below, because pruning away tips lacking breeding system in the ploidy-only analyses (and vice versa) would discard data that could inform the diversification models.

Models

To investigate the associations of these traits with lineage diversification, we fit six models from the BiSSE (Maddison et al. 2007) family. The simplest models consider ploidy only or breeding system only, as previous work has done. We also consider the two traits simultaneously, to investigate whether the effect of one might override the effect of the other. Additionally, to allow for the possibility that something other than these traits might affect diversification, we add models that include a hidden trait.

We first considered only the ploidy trait, scoring each species in either the diploid state (D) or the polyploid state (P). To allow for the influence of this binary trait on lineage diversification, we used the BiSSE model (Maddison et al. 2007). In a Bayesian framework, we obtained posterior probability distributions of speciation rates (λ_D , λ_P), extinction rates (μ_D , μ_P), net diversification rates ($r_D = \lambda_D - \mu_D$, $r_P = \lambda_P - \mu_P$), and relative extinction rates ($v_D = \mu_D/\lambda_D$, $v_P = \mu_P/\lambda_P$) for each state. We also simultaneously estimated the polyploidization rate (the transition rate from D to P, ρ), and we allowed for diploidization (the transition rate from P to D, δ). We call this the “ploidy only” model.

Second, we considered ploidy and also a unobserved, hidden trait. As discussed by Beaulieu and O’Meara (2016), the previous model might attribute diversification differences to the ploidy trait merely because there are no other explanatory factors in the model. By including a hidden trait in a state speciation and extinction model, HiSSE-like models allow for heterogeneous background diversification while also parsing out the possible signal of diversification due to the trait of interest. Our second model thus has four states, with D and P subdivided by an unobserved binary trait with states A and B. We therefore estimated the posterior probability distributions of four speciation rates (λ_{DA} , λ_{DB} , λ_{PA} , λ_{PB}), four extinction rates (μ_{DA} , μ_{DB} , μ_{PA} , μ_{PB}), and the corresponding four net diversification rates and relative extinction rates. We again include transition due to polyploidization (rate ρ) and diploidization (rate δ), and we assume that the changes between hidden states are symmetrical with rate α . We call this the “ploidy and hidden trait” model.

In our third, “breeding system only” model, each species is scored as either self-incompatible (SI) or self-compatible (SC). In this BiSSE analysis, state-dependent speciation rates (λ_I , λ_C) and extinction rates

(μ_I , μ_C) are estimated, with the corresponding net diversification and relative extinction rates. The rate for transitions from SI to SC is q_{IC} , and regain of SI is prohibited, as explained above. This is the same as the analysis of Goldberg et al. (2010).

The fourth model considers both breeding system and a hidden trait, again using a HiSSE strategy to test whether diversification shifts unrelated to the focal trait might be attributed to it by the previous BiSSE analysis. In this “breeding system and hidden trait” model, there are four speciation rates (λ_{I_A} , λ_{I_B} , λ_{C_A} , λ_{C_B}), four extinction rates (μ_{I_A} , μ_{I_B} , μ_{C_A} , μ_{C_B}), and correspondingly four rates of net diversification and relative extinction. Transition rates are again q_{IC} for loss of SI and α for changes in the hidden state.

Our fifth model investigates the link among diversification, breeding system, and ploidy simultaneously. There are three possible states: self-incompatible diploids (ID), self-compatible diploids (CD), and self-compatible polyploids (CP); self-incompatible polyploids are not present in this family, as explained earlier. This “ploidy and breeding system” MuSSE model (FitzJohn 2012) has six rates for speciation and extinction (λ_{ID} , μ_{ID} , λ_{CD} , μ_{CD} , λ_{CP} , μ_{CP}). It also has four rates for transitions: polyploidization of SI diploids (transition from ID to CP at rate ρ_I), polyploidization of SC diploids (transition from CD to CP at rate ρ_C), diploidization, which does not restore SI (transition from CP to CD at rate δ_C), and loss of SI without a change in ploidy (transition from ID to CD at rate q_{IC}).

Finally, the sixth model considers a hidden trait along with ploidy and breeding system. This again provides a means to account simultaneously for diversification rates linked to ploidy and breeding system while also allowing for the presence of heterogeneity from unknown sources. (A similar approach was used for a different variant of BiSSE by Huang et al. 2018; Caetano et al. in press.) The full model would have 26 parameters, but our goal is to look for diversification rate differences so we fit a simplified version with 17 parameters by fixing the transition rates among hidden states to be equal with parameter α , and the transition rates between breeding system and ploidy states as defined in the previous model (ρ_I , ρ_C , δ_C , q_{IC}) independent of the hidden state. We estimated twelve speciation and extinction rates (λ_{ID_A} , μ_{ID_A} , λ_{CD_A} , μ_{CD_A} , λ_{CP_A} , μ_{CP_A} , λ_{ID_B} , μ_{ID_B} , λ_{CD_B} , μ_{CD_B} , λ_{CP_B} , μ_{CP_B}), and the net diversification and relative extinction rates associated with them.

All the models were performed using RevBayes (Höhna et al. 2016) software that performs Bayesian inference via MCMC using Metropolis-Hastings algorithm. A correction for sampling bias was done in all models by assuming that Solanaceae family has approximately 3,000 species ($s = 595/3000$) as the Solanaceae Source project indicates (PBI *Solanum* Project 2012). In all models, speciation and extinction parameters used log-normal prior distributions that with mean the expected net diversification

rate ($\frac{(\text{number of taxa})/2}{\text{root age}}$) and standard deviation 0.5. Prior distributions for parameters defining cladogenetic changes were gamma distributed with parameters $k = 0.5$ and $\theta = 1$. MCMC was performed for 96 hours in the cluster at Minnesota Supercomputing Institute which allowed for 5,000 generation of burn-in and a minimum of 200,000 generations of MCMC for each of the 6 models. Convergence of the MCMC was tested using R package coda (ref) and software Tracer (ref) to assess convergence and mixing (see supplementary information).

Results

Discussion

Acknowledgements

Literature Cited

- Beaulieu, J. M. and B. C. O'Meara, 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst Biol* 65:583–601.
- 140 Bennett, M. D. and I. J. Leitch, 2005. Plant DNA C-values database.
- Caetano, D. S., B. C. O'Meara, and J. M. Beaulieu, in press. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution*.
- FitzJohn, R. G., 2012. Diversitree : comparative phylogenetic analyses of diversification in r. *Methods Ecol Evol* 3:1084–1092.
- 145 Goldberg, E. E. and B. Igić, 2012. Tempo and mode in plant breeding system evolution. *Evolution* 66:3701–3709.
- Goldberg, E. E., J. R. Kohn, R. Lande, K. A. Robertson, S. A. Smith, and B. Igić, 2010. Species selection maintains self-incompatibility. *Science* 330:493–495.
- Huang, D., E. E. Goldberg, L. M. Chou, and K. Roy, 2018. The origin and evolution of coral species richness in a marine biodiversity hotspot. *Evolution* 72:288–302.
- 150 Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist, 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol* 65:726–736.
- Igić, B., L. Bohs, and J. R. Kohn, 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proc Natl Acad Sci USA* 103:1359–1363.
- 155 Maddison, W. P., P. E. Midford, and S. P. Otto, 2007. Estimating a binary character's effect on speciation and extinction. *Syst Biol* 56:701–710.
- Mayrose, I., S. H. Zhan, C. J. Rothfels, N. Arrigo, M. S. Barker, L. H. Rieseberg, and S. P. Otto, 2015. Methods for studying polyploid diversification and the dead end hypothesis: a reply to soltis et al. (2014). *New Phytol* 206:27–35.
- 160 Mayrose, I., S. H. Zhan, C. J. Rothfels, K. Magnuson-Ford, M. S. Barker, L. H. Rieseberg, and S. P. Otto, 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333:1257.
- Olmstead, R. G. and L. Bohs, 2007. A summary of molecular systematic research in Solanaceae: 1982–2006. *Acta Horticulturae* Pp. 255–268.
- 165 PBI *Solanum* Project, 2012. Solanaceae Source: a global taxonomic resource for the nightshade family.
- Raduski, A. R., E. B. Haney, and B. Igić, 2012. The expression of self-incompatibility in angiosperms is bimodal. *Evolution* 66:1275–1283.
- Ramsey, J. and D. W. Schemske, 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29:467–501.
- 170 Rice, A., L. Glick, S. Abadi, M. Einhorn, N. M. Kopelman, A. Salman-Minkov, J. Mayzel, O. Chay, and I. Mayrose, 2015. The chromosome counts database (CCDB) - a community resource of plant chromosome numbers. *New Phytol* 206:19–26.
- Robertson, K., E. E. Goldberg, and B. Igić, 2011. Comparative evidence for the correlated evolution of polyploidy and self-compatibility in solanaceae. *Evolution* 65:139–155.
- 175 Särkinen, T., L. Bohs, R. G. Olmstead, and S. Knapp, 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol Biol* 13:214.
- Soltis, D. E., M. C. Segovia-Salcedo, I. Jordon-Thaden, L. Majure, N. M. Miles, E. V. Mavrodiev, W. Mei, M. B. Cortez, P. S. Soltis, and M. A. Gitzendanner, 2014. Are polyploids really evolutionary dead-ends (again)? a critical reappraisal of mayrose et al. (2011). *New Phytol* 202:1105–1117.
- 180 Stone, J. L., 2002. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. *The Quarterly Review of Biology* 77:17–30.
- Zenil-Ferguson, R., J. M. Ponciano, and J. G. Burleigh, 2017. Testing the association of phenotypes with polyploidy: An example using herbaceous and woody eudicots. *Evolution* 71:1138–1148.

Type	Total of Taxa	Models		
		BiSSE/HiSSE Polyploidy	BiSSE/HiSSE Breeding System	MuSSE/MuHiSSE
Diploid Self Compatible	152	0	0	0
Diploid Self Incompatible	97	0	1	2
Diploid with unknown breeding system	219	0	(0,1)	(0,2)
Polyploid	81	1	0	1
Unknow ploidy and self compatible	34	(0,1)	0	(0,1)
Unknown ploidy and self incompatible	12	0	1	2

Table 1: Binary and three state classifications for 595 taxa with ploidy and/or breeding system data. The number of taxa in the t he sample was maximize by including tips with only ploidy or only breeding system and assigned them as uncertain in the unknown character.