

Accommodating hybrid/introgressed lineages in the multispecies coalescent:

The complex evolutionary history of *Antirrhinum* unraveled

‡ Isabel Liberal, ‡ Jose L. Blanco-Pastor, * Bernard E. Pfeil and ‡ Pablo Vargas

‡ Real Jardín Botánico de Madrid, CSIC, Madrid, Spain
* Gothenburg University, Gothenburg, Sweden

Abstract

Antirrhinum L. (snapdragons; c. 25 spp.) is an Old World plant genus primarily distributed in the western Mediterranean basin. Due to the interesting ecological and morphological features that exhibit this genus, active research has been undertaken on pollination biology and molecular floral development for many wild and cultivated *Antirrhinum* species.

In spite of this research attention, and regardless the extensive efforts made, the phylogenetic relationships within the genus are far from being well resolved. Previous phylogenetic analyses failed to infer species relationships. The difficulty in defining species boundaries and the low phylogenetic resolution suggested an early rapid diversification of the genus. Under this scenario, processes obscuring phylogenetic reconstruction such as hybridization/introgression and/or incomplete lineage sorting are likely to have occurred.

In order to obtain accurate phylogenetic relationships within *Antirrhinum* we obtained allelic data from two low-copy nuclear genes and three plastid loci. We carried out an exact test of population differentiation and pairwise MIGRATE-n analyses to detect gene flow among species. Additionally, we incorporated hybrids/introgressed species in the multispecies coalescent model by using a multilabelling approach in which a lineage of hybrid origin is represented with multiple sources of parental contribution. This methodology allowed the reconstruction of complex evolutionary patterns that have occurred in this group of species extensively used for evolutionary research.

Material and methods

Sequencing. We amplified three plastid regions (*trnK-matK*, *trnS-trnG* and *rpl32-trnL*) and two putative low-copy nuclear gene (LCNG) (*Ei3E* and *AroB*). PCR conditions follow Vargas et al. (2009) for *trnK-matK*, *trnS-trnG* and *rpl32-trnL* regions and Li et al. (2008) for the nuclear loci (two intron-spanning regions). All datasets were aligned with the MAFFT v. 6.814b alignment tool (Katoh et al. 2002). *Ei3E* exhibited properties of multi-copy gene (e.g. sites that were heterozygous in a high number of individuals). Haplotypes were determined as follows: (i) when only one polymorphic site is present, gametic phase was determined unambiguously; (ii) when two polymorphic sites are present we used PHASE (Version 2.1; Stephens et al. 2001) to infer the gametic phase statistically; (iii) if PHASE was unable to assign gametic phase with posterior probability greater than 0.95, then these sequences and those showing more than two ambiguities were cloned.

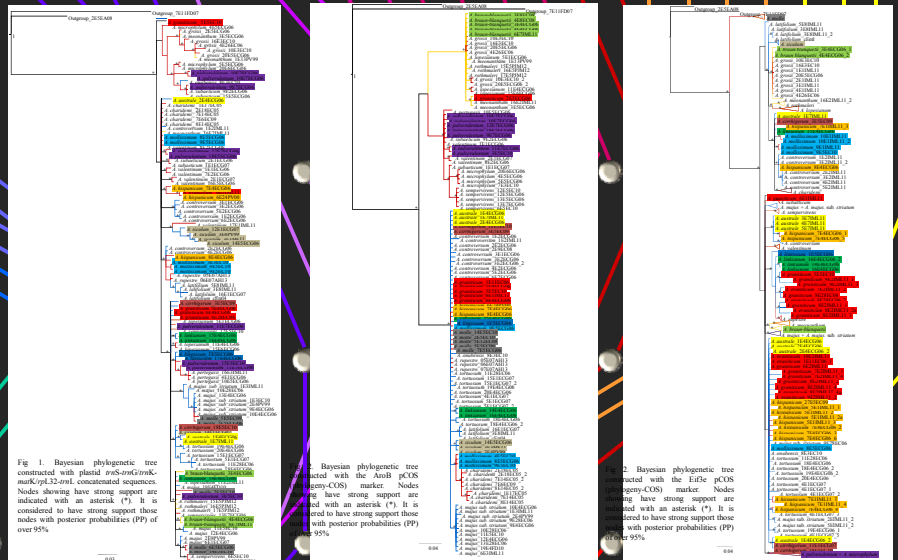
Gene trees. Different statistical test of recombination were performed for the two LCNG using the RDP (Recombination Detection Program) software. Phylogenetic analyses for *trnS-trnG*, *trnK-matK*, *rpl32-trnL* concatenated sequences, *AroB* and *Ei3E* were conducted using Bayesian inference (BI). In addition, Bayesian phylogenetic analyses were also performed on the separate plastid matrices to examine plastid gene tree congruence (not shown). *Gambelia speciosa* and *Mitspates orontium* were selected as the outgroup based on previous phylogenetic evidence (Vargas et al. 2004). Selection of model evolution → Akaike Information Criterion (AIC) in jModeltest 0.1.1 (Posada 2008). GTR+G, GTR and HKY+I+G models were selected for the plastid regions, *AroB* and *Ei3E* loci respectively. BI was performed in MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003). Two identical searches with 10 million generations each and a sample frequency of 1000 were performed. Chain convergence was assessed with Tracer 1.5 (Rambaut & Drummond 2009).

Genetic differentiation among species. We tested for genetic differentiation between each pair of *Antirrhinum* species by using the exact test of population differentiation (Raymond & Rousset 1995). The species that were not significantly differentiated for none of the three loci analyzed (pDNA, *AroB*, *Ei3E*) were considered to be connected by gene flow.

Directionality of gene flow. To assess for directionality of gene flow between pairs of non-differentiated *Antirrhinum* species, we compared models of gene flow by applying a Bayesian coalescent approach implemented in migrate-n (Beerli & Felsenstein 2001; Beerli & Pacerowski 2010). For each pair of non-differentiated species we completed two different analyses with either migration parameter (M_{12} and M_{21}) set to zero in the connection matrix of the migration model. Natural log Bayes factor (LBF) of each model was calculated via $LBF = 2 \ln(L(\text{model}_1)/L(\text{model}_2))$. Following Kass & Raftery (1995), LBF values smaller than -2 suggested preference for model 2 while values larger than 2 suggested preference for model 1.

Multilabelled species tree. We considered species receiving migrant alleles as species with multiple distinct origins (hereafter hybrid species) and therefore represented them with multiple labels in a *BEAST species tree analysis implemented in the BEAST package v1.7. (see Blanco-Pastor et al. 2012). The number of labels assigned to a hybrid species depended on the number of donor species. A species with n donors was represented with $n+1$ labels. With that approach we accommodated gene-tree incongruence caused by both incomplete lineage sorting and recent hybridization/introgression.

Gene trees

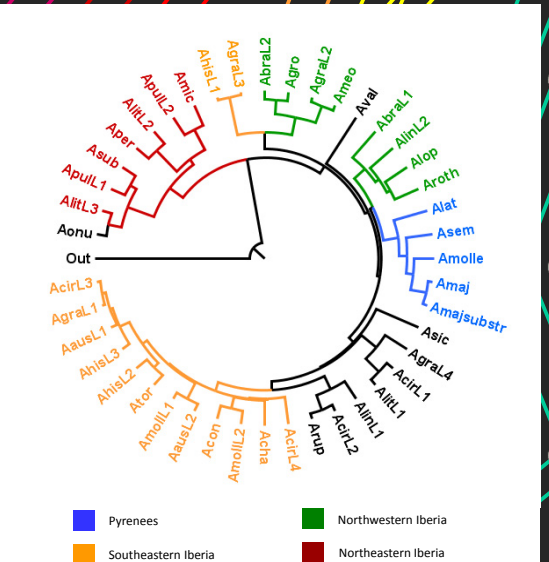


Differentiation test

	aus	bra	cha	cir	con	gra	gro	his	lat	lin	lit	lop	str	maj	moo	mic	moll	mollu	oss	per	pul	roth	rup	sem	sic	sub	tor	val
aus	n/a																											
bra	n/a	n/a																										
cha	n/a	n/a	n/a																									
cir	n/a	n/a	n/a	n/a																								
con	n/a	n/a	n/a	n/a	n/a																							
gra	n/a	n/a	n/a	n/a	n/a	n/a																						
gro	n/a	n/a	n/a	n/a	n/a	n/a	n/a																					
his	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a																				
lat	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a																			
lin	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a																		
lit	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a																	
lop	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a																
str	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a															
maj	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a														
moo	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a													
mic	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a												
moll	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a											
mollu	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a										
oss	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a									
per	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a								
pul	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a							
roth	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a						
rup	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a					
sem	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a				
sic	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a			
sub	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a		
tor	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	
val	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

AroB/Ei3E pDNA: "+" differentiation; "-" no differentiation; significance level=0.05

Multilabelled Species tree



References

Beerli, P. and J. Felsenstein (2001). "Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach." *Proceedings of the National Academy of Sciences* 98(8): 4563-4568.

Beerli, P. and M. Pacerowski (2010). "Unified framework to evaluate panmixia and migration direction among multiple sampling locations." *Genetics* 185(1): 313-326.

Blanco-Pastor, J. L., P. Vargas, et al. (2012). "Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean Linaria." *PLoS One* 7(6): e39089.

Huelsenbeck, J. and F. Ronquist (2006). "MrBayes: Bayesian inference of phylogeny v3.1.2." Distributed under the GNU general public license from the website: <http://mrbayes.cit.berkeley.edu>.

Kass, R. E. and A. E. Raftery (1995). "Bayes factors." *Journal of the American statistical association* 90(430): 773-795.

Katoh, K., K. Misawa, et al. (2002). "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform." *Nucleic acids research* 30(14): 3059-3066.

Li, M., J. Wunder, et al. (2008). "Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species." *Cladistics* 24(5): 727-745.

Posada, D. (2008). "jModelTest: phylogenetic model averaging." *Molecular biology and evolution* 25(7): 1253-1256.

Rambaut, A. and A. Drummond (2010). *Tracer v1.5*. <http://tree.bio.ed.ac.uk/software/tracer/>. Accessed.

Raymond, M. and F. Rousset (1995). "An exact test for population differentiation." *Evolution* 49(6): 1280-1283.

Stephens, M., N. J. Smith, et al. (2001). "A new statistical method for haplotype reconstruction from population data." *The American Journal of Human Genetics* 68(4): 978-989.

Vargas, P., E. Carró, et al. (2009). "A geographical pattern of *Antirrhinum* (Scrophulariaceae) speciation since the Pliocene based on plastid and nuclear DNA polymorphisms." *Journal of Biogeography* 36(7): 1297-1312.

Vargas, P., J. Rosello, et al. (2004). "Molecular evidence for naturalness of genera in the tribe Antirrhineae (Scrophulariaceae) and three independent evolutionary lineages from the New World and the Old." *Plant Systematics and Evolution* 249(3-4): 151-172.

Corresponding authors

iliber@rjb.csic.es; jblanco@rjb.csic.es; bernard.pfeil@bioenvy.gu.se; vargas@rjb.csic.es