

# Coalescent Simulations Reveal Hybridization and Incomplete Lineage Sorting in Mediterranean *Linaria*

José Luis Blanco-Pastor<sup>1\*</sup>, Pablo Vargas<sup>1</sup>, Bernard E. Pfeil<sup>2</sup>

**1** Departamento de Biodiversidad y Conservación, Real Jardín Botánico (RJB-CSIC), Madrid, Spain, **2** Department of Biological and Environmental Sciences, Gothenburg University, Gothenburg, Sweden

## Abstract

We examined the phylogenetic history of *Linaria* with special emphasis on the Mediterranean sect. *Supinae* (44 species). We revealed extensive highly supported incongruence among two nuclear (ITS, AGT1) and two plastid regions (*rpl32-trnL*<sup>UAG</sup>, *trnS-trnG*). Coalescent simulations, a hybrid detection test and species tree inference in \*BEAST revealed that incomplete lineage sorting and hybridization may both be responsible for the incongruent pattern observed. Additionally, we present a multilabelled \*BEAST species tree as an alternative approach that allows the possibility of observing multiple placements in the species tree for the same taxa. That permitted the incorporation of processes such as hybridization within the tree while not violating the assumptions of the \*BEAST model. This methodology is presented as a functional tool to disclose the evolutionary history of species complexes that have experienced both hybridization and incomplete lineage sorting. The drastic climatic events that have occurred in the Mediterranean since the late Miocene, including the Quaternary-type climatic oscillations, may have made both processes highly recurrent in the Mediterranean flora.

**Citation:** Blanco-Pastor JL, Vargas P, Pfeil BE (2012) Coalescent Simulations Reveal Hybridization and Incomplete Lineage Sorting in Mediterranean *Linaria*. PLoS ONE 7(6): e39089. doi:10.1371/journal.pone.0039089

**Editor:** Thomas Mailund, Aarhus University, Denmark

**Received:** March 9, 2012; **Accepted:** May 18, 2012; **Published:** June 29, 2012

**Copyright:** © 2012 Blanco-Pastor et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** JLB-P received support from a Spanish National Research Council grant (CSIC: JAEpre). This work is framed within a project from the Spanish Ministry of Environment, reference 005/2008. BEP is supported by grants from VR (Swedish Research Council), KVA (Royal Swedish Academy of Sciences), Lars Hiertas Minne fund, The Royal Physiographic Society in Lund, Helge Ax:son Johnsons fund and the Lundgrenska fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: jlblanco@rjb.csic.es

## Introduction

Gene trees can differ from one another and do not always correspond to species trees [1–4]. Wendel and Doyle [5] listed three categories of processes that may cause incongruent patterns: technical causes, organism-level processes and gene- or genome-level processes. If technical causes, selection, paralogy and recombination can be ruled out, then (i) hybridization among fully differentiated species with subsequent fixation of nuclear and/or organellar loci and (ii) the incomplete random sorting of alleles at many loci independently due to short intervals between divergence events (hereafter incomplete lineage sorting) often remain as the main hypotheses that can explain gene tree incongruence [6–11]. Typically, phylogenetic analyses using single locus datasets (e.g. [12–14]) or concatenated datasets (e.g. [15–18]) have provided inferences of relationships in numerous plant groups. Nonetheless, a tree based on a single locus or concatenated genes may lead to a spurious representation of the history of the species [19,20]. Several methods that distinguish hybridization from incomplete lineage sorting have been recently described [21–23]. However, many independent loci are needed for their implementation and hybridization is difficult to uncover if multiple reticulation events have occurred. Ané *et al.* [24] implemented a method that can accommodate any source of incongruence even using a limited number of loci, but this method is unable to determine the process causing incongruence among phylogenies. Also, Maureira-Butler *et al.* [6] and Joly *et al.* [10] have proposed statistical frameworks, applicable to datasets with few independent

loci, where hybridization can be detected in the presence of incomplete lineage sorting. Alternatively, several models can estimate the correct species tree if incongruence is due to incomplete lineage sorting alone [20,25–29], but in such models hybridization signals need to be previously ruled out or excluded. If not, an incorrect species tree may be inferred by such methods [20,30].

Both polyploid and homoploid hybrid speciation might represent a large fraction of the source of plant biodiversity on Earth [31]. In the Mediterranean basin, several plant groups suffered secondary contacts in their postglacial colonization routes from their glacial maximum refugia located in southern peninsulas [32] or after altitudinal migrations in restricted areas within peninsulas (e.g. Iberian Peninsula, [33,34]). A considerable proportion of the present Mediterranean plant diversity may be the result of hybridization episodes, which *per se* represent a challenge for phylogenetic reconstruction. Besides this, species complexes that underwent rapid speciation also represent a major challenge for molecular systematics. In those groups species relationships could be obscured by the ancestral polymorphisms retained through speciation events as a consequence of incomplete lineage sorting [2,35]. In the Mediterranean region, rapid plant speciation has been recently detected [36–38] and associated with adaptation to the establishment of the Mediterranean climatic rhythm (summer drought) (3.2 Ma) or the Quaternary-type Mediterranean climatic fluctuations (2.3 Ma) [39].

Toadflaxes (*Linaria* Mill.) constitute the largest genus within the snapdragon lineage (tribe Antirrhineae). *Linaria* comprises c.150

species that are widely distributed in the Palearctic region, but the genus is most diverse in the Mediterranean basin. The origin of the genus has been placed in the Miocene [40] predating the Messinian Salinity Crisis [41]. The monophyly of *Linaria* has been suggested based on nrDNA (ITS) sequences of eight species representing all sections [42], however, whether the sections constitute natural groups remains uncertain. Numerous taxonomic treatments of *Linaria* have been proposed [43–51], but remarkable disagreement in the infrageneric classification suggests complex evolutionary processes. The latest classification of the genus recognizes seven sections (*Linaria*, *Speciosae*, *Diffusae*, *Supinae*, *Pelisserianae*, *Versicolores* and *Macrocentrum*) [45]. Section *Supinae* (Benth.) Wettst. (hereafter *Supinae*) is a clear example of the systematic complexity within *Linaria* because of the disagreement in taxonomic treatments (Table 1). *Supinae* comprises 44 diploid (2n = 12) [52] hermaphroditic annual and perennial species differentiated from other sections by their laterally-compressed winged seeds that have a horizontal arrangement in globose capsules [45]. *Supinae* species are distributed in the temperate regions of Europe, northern Africa and western Asia (circum-Mediterranean distribution), with the highest diversity found in the Iberian Peninsula (40 species) [44,45].

In *Linaria*, hybrid species have been historically described when intermediate characters of two species meet in a plant [53,54]. In section *Supinae* several natural hybrids have been previously reported [44,55–57]. Artificial experiments have also shown the potential of hybridization inasmuch as *Supinae* species that do not meet in nature can produce capsules after hand cross-pollination (Blanco-Pastor, unpublished), [53]. The highest fertilization success was found in crosses among *Supinae* species (13 successful crosses of 20 assayed), followed by clearly lower values in inter-sectional crosses (four successful crosses of 14) [53]. A lack of internal reproductive barriers among *Supinae* species is then suggested. Despite this, external barriers such as allopatry do exist at the present time within *Supinae* as few species have overlapping distributions. However, such geographical barriers may have not existed during glaciations.

The high chance for hybridization in *Linaria* may affect phylogenetic reconstruction in this genus. Nonetheless, incomplete lineage sorting cannot be discarded as a cause of phylogenetic incongruence.

Both processes can be difficult to distinguish, but may also occur simultaneously [58]. Within this framework, we investigate causes of incongruence between three presumably unlinked loci. Two nuclear (ITS and AGT1) and two linked plastid (*rpl32-trnL<sup>UAG</sup>* and *trnS-trnG*) regions are herein sequenced for *Linaria*, with special emphasis in *Supinae* species. Our aims are: (i) to test for the presence of reticulation signals by simulations under the coalescent model using the method of Maureira-Butler *et al.* [6], (ii) to detect individuals that may have been affected by historical hybridization (hereafter potential hybrids), (iii) to exclude potential hybrids and infer the species tree using a method that accounts for incomplete lineage sorting (\*BEAST) [20], (iv) to compare the \*BEAST species tree with our original gene trees to identify random sorting episodes, and (v) to recover the reticulation events by locating the parental lineages of the potential hybrids in a multilabelled species tree. The ultimate goal is to disclose the evolutionary history of *Supinae* by exploring the presence of incomplete lineage sorting and/or reticulation events that may have occurred during the course of the evolution of this plant group.

## Materials and Methods

### Sampling Strategy

Individuals were collected in the field and dried in silica gel or obtained from herbaria (MA, E, RNG) (Table S1). Total genomic DNA was extracted using the Dneasy Plant Mini Kit (QUIAGEN Inc., California). We amplified (using an Eppendorf Mastercycler Epgradient S, Westbury, NY) a low copy nuclear gene intron (AGT1) [59], the nuclear ribosomal internal transcribed spacer (ITS) [60] and two plastid regions (*rpl32-trnL<sup>UAG</sup>*, *trnS-trnG*) [61,62] in 52 individuals representing 46 *Linaria* species plus one individual of *Antirrhinum* and one individual of *Chaenorhinum*. In particular, we used one species of sect. *Macrocentrum* (*L. chalapensis*), three species of sect. *Versicolores* (*L. spartea*, *L. gharbensis*, *L. multicaulis*), five species of sect. *Linaria* (*L. meyeri*, *L. loeselii*, *L. odora*, *L. tibetica*, *L. vulgaris*), four species of sect. *Speciosae* (*L. ventricosa*, *L. dalmatica*, *L. peloponnesiaca*, *L. genistifolia*), seven species of sect. *Diffusae* (*L. albifrons*, *L. flava*, *L. triphylla*, *L. laxiflora*, *L. wariensis*, *L. haelava*, *L. joppensis*) and 24 of the 44 species of section *Supinae* [45]. We followed Sutton's species delimitation [45] for the non-Iberian species and Sáez & Bernal's delimitation [44] for the Iberian species but with minor changes

**Table 1.** Systematic classification of *Linaria* sect *Supinae* suggested in this study and its relation with previous classifications.

Bentham (1846)	Wettstein (1895)	Valdés (1970) and Viano (1978)	Sutton (1988), Sáez (2008)	Present study
<i>Linaria</i> sect. <i>Linariastrum</i> Chav.	<i>Linaria</i> Juss.	<i>Linaria</i> Miller	Sect. <i>Supinae</i> (Bentham) Wettst.	Sect. <i>Supinae</i> (Bentham) Wettst.
§ <i>Arvenses</i> Bentham p.p.max.	Sect. <i>Arvenses</i> (Bentham) Wettst. p.p.max.	Sect. <i>Arvenses</i> (Bentham) Wettst.	Subsect. <i>Supinae</i> p.p.	Subsect. <i>Arvenses</i>
§ <i>Supinae</i> Bentham p.p. § <i>Diffusae</i> Bentham p.p. § <i>Grandes</i> Bentham p.p.min.	Sect. <i>Supinae</i> (Bentham) Wettst. p.p.	Sect. <i>Supinae</i> (Bentham) Wettst. subsect. <i>Supinae</i>	Subsect. <i>Supinae</i> p.p.	Subsect. <i>Supinae</i>
§ <i>Supinae</i> Bentham p.p. § <i>Versicolores</i> Bentham p.p.min.	Sect. <i>Supinae</i> (Bentham) Wettst. p.p.	Sect. <i>Supinae</i> subsect. <i>Saxatile</i> Valdés p.p. max Sect. <i>Supinae</i> (Bentham) Wettst. subsect. <i>Supinae</i> p.p. Sect. <i>Supinae</i> subsect. <i>Amethystea</i> Valdés Sect. <i>Bipunctatae</i> Viano p.p.max.	Subsect. <i>Saxatile</i> Valdés Subsect. <i>Supinae</i> p.p.	Subsect. <i>Saxatile</i> Valdés
–	–	–	Subsect. <i>Trimerocalyx</i> (Murb.) – D.A. Sutton	–

p.p. = *pro parte*.

p.p.max = *pro parte maxima*.

p.p.min = *pro parte minima*.

doi:10.1371/journal.pone.0039089.t001

regarding the “*Linaria verticillata* group” and the “*Linaria alpina* group” [44,63] (see Methods S1). We also included one additional species neither considered by Sutton nor Sáez & Bernal: *L. almijarensis* Campo & Amo [64] (see Table S1). All necessary permits were obtained for the described field studies. In cases where plant locations were protected we obtained permissions from the “Consejería de Medio Ambiente” of Andalusian Government (Spain), references: GB-86/2010/EA/FL/FA/JMLV, ENSN/JSG/IHC/MCF. Amplification products were outsourced for sequencing to a contract sequencing facility (Macrogen, Seoul, South Korea) on an ABI Prism® 3730xi DNA sequencer, using the same primer set as for PCR. Sequence data were edited using Geneious software (Biomatters Ltd., Auckland, New Zealand). Sequences are available in GenBank (see Table S1).

### Deciphering of Haplotypes in Unphased Genotypes

More than one allele was found in both AGT1 and ITS in Sanger sequenced PCR amplicons. To decipher these, we first estimated the gametic phases of the sequences using Arlequin 3.5.1.2 [65]. This program performs a Gibbs sampling via the ELB algorithm [66] to obtain the posterior probability of phased haplotypes. The settings for the ELB algorithm were as follows: dirichlet alpha value: 0.01, epsilon value: 0.1, heterozygote site influence zone: 5, gamma value: 0.01, sampling interval: 500, no. of samples: 2000, burn-in steps: 100000 and 0% of recombination steps. AGT1 haplotypes retrieved with posterior probability under 0.95 were confirmed by cloning the purified PCR products using the Promega Corporation protocol (Madison, USA) with JM109 High Efficiency competent cells and pLysS plasmids. Four single recombinant colonies from each reaction were screened. Amplifications were performed using the T7-SP6 plasmid primers. All ITS haplotypes inferred with Arlequin were used to build allele trees. In only one case (*L. bubanii*) ITS haplotypes were not inferred as sister (or very closely related) sequences in the gene trees. As the phase posterior probability for this individual was low (0.41), we empirically confirmed the *L. bubanii* ITS haplotypes by sequencing the PCR product using allele-specific primers as described in Scheen *et al.* [67].

### Test for Recombination

Recombination was tested within ITS and AGT1 datasets using RDP 3.44 [68] with the following methods: RDP [69], Geneconv [70], MaxChi [71], Bootscan/Reccscan [72], SisScan [73], 3Seq [74] and Chimaera [75]. We selected 0.05 as the p-value cut-off in general settings and internal references only in the RDP method. A window size of 150 and step size of 20 was used in the Bootscan and SisScan methods and a variable window size was set in MaxChi and Chimaera methods. We considered that recombination was likely if it was accepted by more than two methods. For the remaining settings we used the default values.

### Gene Trees Estimation and Calculation of Dates

The haplotype sequences obtained from the three datasets (ITS, AGT1, cpDNA) were analyzed by Bayesian Inference in MrBayes 3.1.2 [76] after alignment with MAFFT v.6 [77] (with corrections by visual inspection) and optimal substitution model selection in jModeltest 0.1.1 [78,79].

For time calibration, we used the divergence time between *Antirrhinum* and *Linaria* (13.33–27.32 Ma) from a previous estimate obtained in a relaxed molecular-clock analysis of tribe Antirrhineae (Vargas *et al.*, unpublished). This analysis was in turn calibrated with five Lamiales fossils and a divergence time between Oleaceae and Antirrhineae modeled as a normal distribution with

mean = 74 Ma and Std = 2.5 Ma, on the basis of a relaxed molecular clock analysis of angiosperms [80], see [40] for details. We used the minimum age (13.33 Ma) as a fixed calibration point for the stem node of the *Linaria* clade to estimate the dates of the internal nodes with a penalized likelihood procedure implemented in r8s 1.71 [81]. Cross-validation to find the optimal smoothing parameter ( $10^k$ ) was done using increments of  $k$  of 0.1, from  $k = -3$  to 3, repeated for two trees from the stable posterior distribution of each gene; the smoothing values of both trees were very similar so we used the value with lower  $\chi^2$  error. After cross-validation we set the smoothing parameter to 1.5 for ITS, 3.2 for AGT1 and 0 for cpDNA and rate smoothed 20 trees drawn from the posterior distribution after burn-in to obtain the chronograms that were used in the coalescent simulations.

### Coalescent Simulations

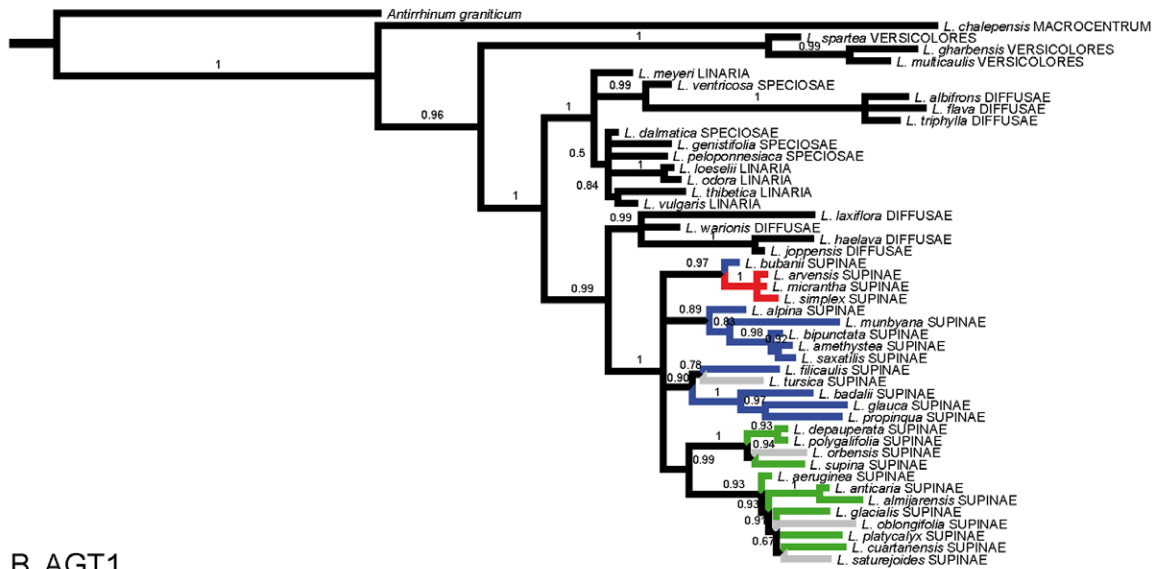
We used simulations under the coalescent model following Maureira-Butler *et al.* [6] to test whether incomplete lineage sorting alone could explain the observed incongruence among gene trees. As the test does not account for the uncertainty of tree topology and branch length estimation, here we used 20 trees from the stable posterior distribution of the Bayesian analysis for each gene, performed the simulations and calculated all tree-to-tree distances from this pool of trees (hereafter the base line distribution), rather than the consensus as was done previously [6]. The base line distribution was then compared to the distribution obtained by calculating pairwise tree-to-tree distances of the 20 chronograms for each gene—essentially a measure of how much the gene trees from each locus differ—hereafter the observed distribution (see Methods S1 for further details).

Effective population size estimates ( $N_e$ ) used in the coalescent simulations were derived from cpDNA haplotypes and obtained via  $\theta_w = 2\mu N_e$  with theta ( $\theta_w$ ) and mutation rate per generation ( $\mu$ ) taken from data of three *Linaria* species with contrasting range sizes (and potentially, contrasting  $N_e$ ) (table S2): *L. glacialis* (endangered, narrow endemic of Sierra Nevada, Spain), *L. elegans* (endemic to northern Iberia) and *L. simplex* (distributed across the Mediterranean basin). The effect of  $N_e$  estimates in the coalescent simulations was explored by repeating the set of simulations using the three  $N_e$  values separately (see Methods S1 for further details).

### Detection of Potential Hybrids

The detection of potential hybrids was addressed by examining the effect of taxon deletion on the observed and base line distributions. Theoretically, the potential hybrids detected by the test were the set of individuals that, after exclusion, retrieved overlapping observed distributions (pairwise tree-to-tree distances within their 95% HPD) and base line distributions (trees from coalescent simulations), thus the null hypothesis of incomplete lineage sorting alone was no longer rejected. Here, this approach was difficult to apply as the results were very dependent on the  $N_e$  values used (see Results). We identified that limitation, but we also recognized the significant challenge of getting exact estimates of population sizes through time in a phylogeny, especially with scarce genetic data [82,83]. We then made an exploration of the effect of the deletion of each terminal with an incongruent position, in order to identify the individuals causing the highest effect in the differences between the baseline and the observed distributions. This was done by excluding terminals with incongruent positions (one at the time) and calculating new base line and observed distributions for the three datasets under each  $N_e$ . The nine replications (three datasets x three  $N_e$ ) alleviated the non-reproducible effect of taxon exclusion due to the stochastic

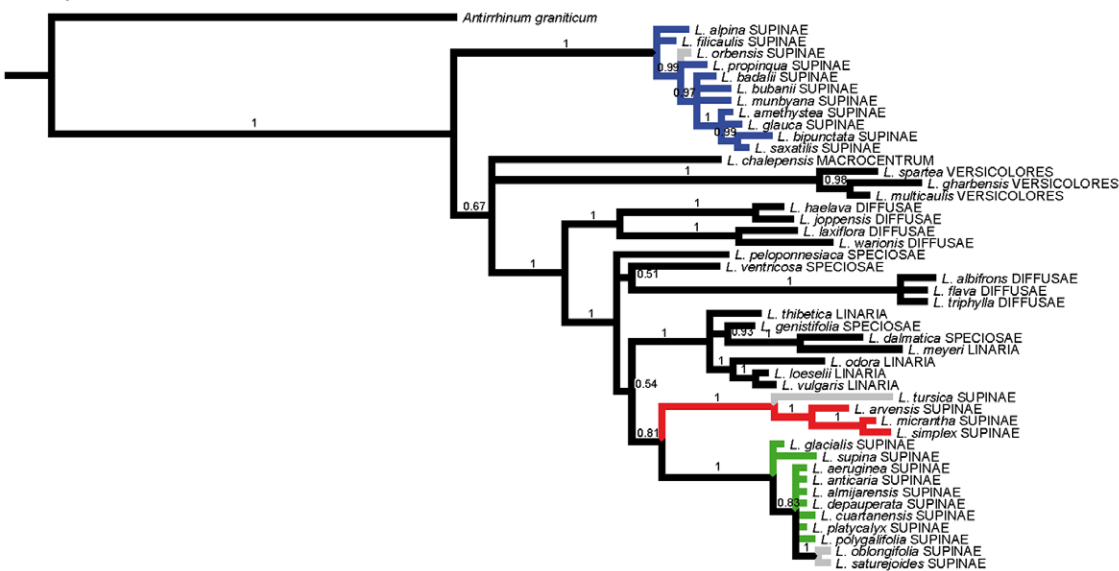
A. ITS



B. AGT1



C. cpDNA



**Figure 1. Gene trees.** Phylogenetic relationships of 47 samples representing 46 *Linaria* species and one individual of *Antirrhinum* as the outgroup. One species of sect. *Macrocentrum*, three species of sect. *Versicolores*, five species of sect. *Linaria*, four species of sect. *Speciosae* and 28 species of sect. *Supinae* are represented. 50% Majority-rule consensus tree obtained in the Bayesian analysis of ITS (A), AGT1 (B) and cpDNA (C) sequences are shown. Numbers above branches represent Bayesian posterior probabilities. Phylogenetic trees are based on one sample and one allele per species, when the two alleles were not sister we used the most incongruent one respecting the other two genes. *Linaria* sections following Sutton [45] are shown in capital letters. Colors represent the systematic nomenclature for *Supinae* clades as suggested in this paper (see Fig. 4). Species with key traits from two *Supinae* clades (Fig. 4) are represented in grey.  
doi:10.1371/journal.pone.0039089.g001

nature of simulations. The last step was to average the nine independent estimates obtained for each analyzed taxon.

### Testing Monophyly of *Supinae*

We used AGT1 and cpDNA datasets (one haplotype per sequence) with hybrids excluded to test support for the monophyly of *Supinae*. This was done to assess whether the incongruence (regarding *Supinae* naturalness) was exclusively explained by hybridization (as putative hybrids were excluded) and inference limitations, or whether additional processes generated real gene tree differences (in this case incomplete lineage sorting). In order to calculate support for the monophyly of *Supinae* we used two approaches: (i) the Shimodaira and Hasegawa [84] (S-H) test and the Bayes Factors [85,86] (BF) test. The S-H test was implemented by calculating the maximum likelihood tree with unconstrained and constrained topologies in RAxML (-f d function) to subsequently compare both ML trees using the -f g function, which computes the per-site log Likelihoods for the contrasted topologies. The per-site log Likelihoods were analyzed with CONSEL [87] to obtain the S-H statistic values. BF test was used to assess alternative phylogenetic hypothesis in a Bayesian framework [85,86]. The BF test quantifies the support for one hypothesis versus another given the data. We also used this approach, implemented in Tracer 1.4 [88] to test significant differences between the unconstrained and constrained Bayesian analyses of AGT1 and cpDNA. Stationarity and convergence of analyses were assessed in Tracer after discarding the first 10% of sampled generations as burn-in. Marginal likelihoods, their standard errors (estimated using 1000 bootstrap replicates) and BF's were calculated. We considered  $2\ln\text{BF}(H_1 \text{ vs. } H_0) - 2$  to  $-6$  as positive evidence against  $H_1$  in favor of  $H_0$ ;  $2\ln\text{BF}(H_1 \text{ vs. } H_0) - 6$  to  $-10$  as strong evidence against  $H_1$  in favor of  $H_0$ ; and  $2\ln\text{BF}(H_1 \text{ vs. } H_0) < -10$  as very strong evidence against  $H_1$  in favor of  $H_0$  [89].

### Species Tree Inference

After excluding potential hybrids (to not violate the species tree model assumptions), we used the allelic data (and >1 individual per species in some cases, see Table S1) to estimate the species tree with the \*BEAST (StarBeast) method [20] implemented in BEAST v.1.6.2. [89]. Allelic data were included in three data partitions with unlinked genealogies: (i) ITS sequences, (ii) AGT1 sequences and (iii) combined plastid (*rpl32-trnL<sup>UAG</sup>* and *trnS-trnG*) sequences. We used Sutton's species delimitation [45], but additionally recognizing *L. almiyarensis* Campo & Amo [64] (one population). The prior probability of the divergence time between *Linaria* and *Antirrhinum* was constrained to 20 Ma  $\pm 4$  as a normal distribution, following date estimates obtained for the tribe Antirrhineae (Vargas *et al.*, unpublished, see "Gene trees estimation and calculation of dates" section). A Birth-Death process [90] was employed as the species tree branching prior. We used an uncorrelated lognormal relaxed clock model, with the prior probability for the substitution rate uniformly distributed, with ranges of  $5 \times 10^{-4}$ - $5 \times 10^{-2}$  and  $1 \times 10^{-4}$ - $1 \times 10^{-2}$  substitutions per site per Ma (s/s/Ma) for the nuclear loci and the plastid locus respectively. These rate constraints include previous estimates for

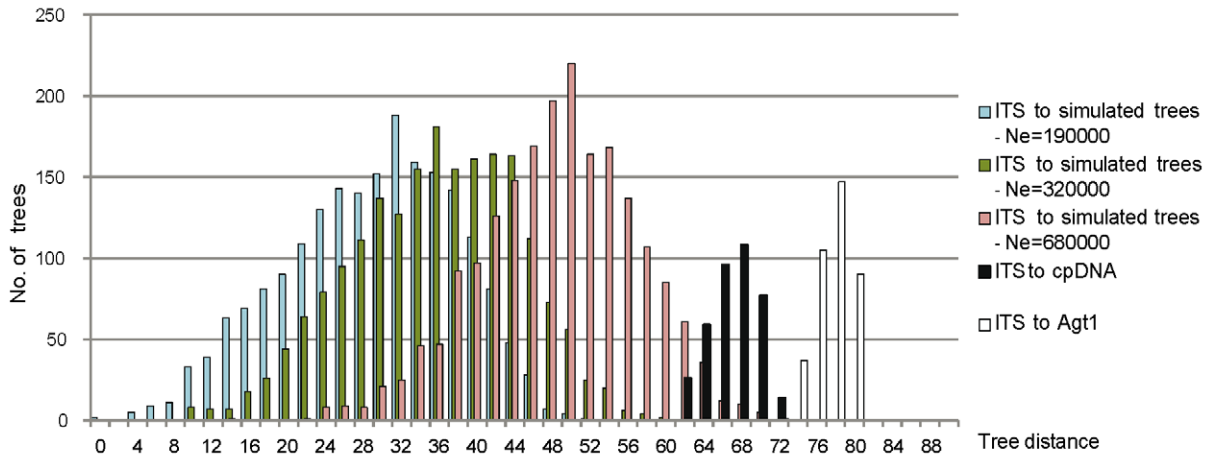
herbaceous plant ITS rates ( $1.7$ - $8.3 \times 10^{-3}$  s/s/Ma) [91] and chloroplast rates ( $1.0$ - $3.0 \times 10^{-3}$  s/s/Ma) [92]. Nuclear synonymous substitution rates, being nearly neutral, may approximate nuclear intron rates. The former rates have been found in other plants to lie within the range we used (e.g., 48 *Gossypium* genes,  $3.5$ - $7.3 \times 10^{-3}$  s/s/Ma, [93]; 39 legume genes, mean of  $5.2 \times 10^{-3}$  s/s/Ma, [94]). Six MCMC analyses were run for 300 million generations each, with a sample frequency of 1000. Analysis with Tracer v.1.5 [88] confirmed convergence of analyses and adequate sample sizes, with ESS values above 200. Analyses were combined using LogCombiner v.1.6.2 after discarding the first 10% generations of each run as burn-in. Trees were summarized in a maximum clade credibility tree using TreeAnnotator v.1.6.2. After combination of the six log files from the analyses, the standard deviation of the uncorrelated lognormal relaxed clock (ucl.d.stdev) and the coefficient of variation (CoV) in the three genes were not close to 0: cpDNA ucl.d.stdev = 0.94, cpDNA Cov = 0.97; AGT1 ucl.d.stdev = 0.806, AGT1 CoV = 0.854; ITS ucl.d.stdev = 0.685, ITS CoV = 0.702. This branch rate heterogeneity indicated that the uncorrelated lognormal relaxed clock was appropriate.

### Multilabelled Species Tree

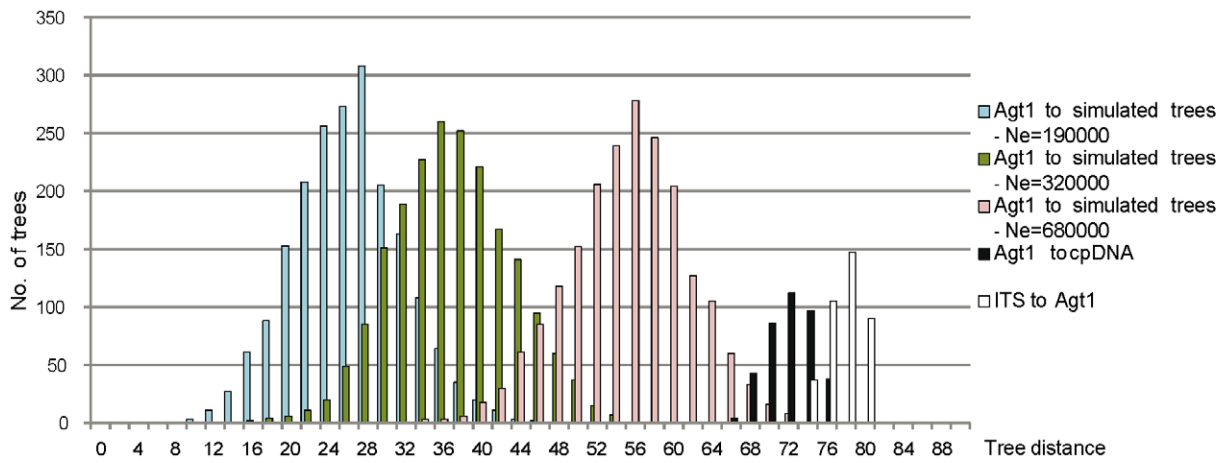
A multilabelled species tree was inferred to retrieve the origin of the parental lineages of individuals affected by reticulation processes. We inferred a second species tree but this time including allelic data from potential hybrids. We recalculated the best-fitting model of sequence evolution with jModeltest 0.1.1 [78,79], while the remaining priors were set as in the species tree analysis. The multilabelled species tree was built by assigning the two most congruent genes to one label (tip, or terminal species branch) and the remaining gene to a second label (see Table S3) while using missing data for the gene not assigned in the label. Thus, the two labels of a potential hybrid species (L1 and L2) were treated as different "species" in \*BEAST analysis in order to show which two hybridizing lineages have contributed to a lineage of hybrid origin. The analysis therefore treated the differences between the two most congruent genes as being caused by incomplete lineage sorting alone, whereas our multilabelling approach allowed the differences between the most incongruent positions to be due to hybridization without violating the assumptions of the \*BEAST model. The key concept is that a lineage of hybrid origin has two sources of parental contribution to its genome. These origins are best represented in a tree diagram by including two labels rather than just one (as is the case for lineages without a hybrid origin). This approach is novel, as far as we know, but has similarities to the approach used by Pirie *et al.* [95]. Four MCMC analyses were run for 100 million generations each, with a sample frequency of 10000. Analysis with Tracer v.1.5 [88] also confirmed convergence of analyses and adequate sample size, with ESS values above 200. We combined the analyses and summarized the tree as indicated above.

In order to contrast the results of the multilabelled species tree with other procedures widely used in phylogenetic studies, we also

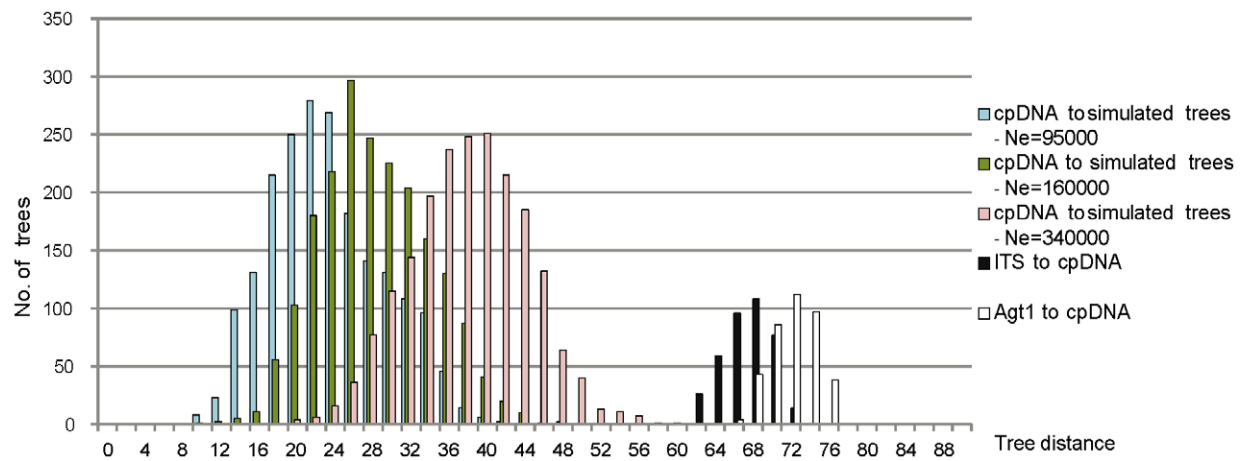
A. ITS



B. AGT1



C. cpDNA



**Figure 2. Baseline and observed distributions of tree distances.** Frequency distribution of tree-to-tree distances between 20 representative trees from the stable posterior distribution of the Bayesian analysis (ITS (A), AGT1 (B) and cpDNA (C)) and 100 simulated gene trees obtained by coalescent simulations (baseline distributions). Blue, green and red bars represent baseline distributions under *L. glacialis*, *L. elegans* and *L. simplex*  $N_e$  estimates respectively. Black and white bars represent the distances between gene trees (observed distributions). doi:10.1371/journal.pone.0039089.g002

performed a \*BEAST species tree analysis and a total evidence analysis, both with potential hybrids included.

## Results

### Haplotype Data

The Arlequin analysis gave us the two most probable haplotypes from the unphased genotypes of AGT1 and ITS sequences. For AGT1, we obtained haplotypes of 50 individuals with posterior probabilities (PP) above 0.95 and haplotypes of four individuals with PP below 0.95. For ITS, we obtained haplotypes of 34 individuals with PP above 0.95 and haplotypes of 20 individuals with PP below 0.95. The AGT1 phased data retrieved for the four individuals with low PP were empirically confirmed by amplicon cloning, recovering exactly the same allelic data that Arlequin inferred. As ITS is a multi-copy locus marker, there would be more than two copies for each unphased ITS genotype. This may have affected the haplotype detection, thus giving low support for

the ITS haplotypes obtained. But (i) as one haplotype with low probability and differential position in the ITS allele-tree has been confirmed empirically (*L. bubanii*, 0.41 PP) and (ii) highly differentiated alleles have not been obtained in the Arlequin analyses (excluding *L. bubanii*), being all sister or closely-related in allelic-gene trees, we then considered that the two ITS haplotypes detected by Arlequin were good representatives of the existing ITS alleles per sample.

### Recombination Test

Recombination could not be detected in ITS by any of the five methods used. AGT1 showed one recombination event affecting several sequences that was detected by SiScan (Av.  $p$ -value =  $3.712 \times 10^{-2}$ ) but when contrasting the UPGMA trees of the recombinant and non-recombinant regions it showed almost the same topology with both potential parents separated in the tree from the potential recombinants. Additionally, evidence for recombination was not considered convincing if it only was

**Table 2. Effect of taxa exclusion on the differences between base line (from simulated trees) and observed distributions of tree distances, numbers indicate steps while negative (-) and positive (+) values indicate approximation and separation between distributions, respectively.**

Taxa with incongruent position in gene trees	Effect after taxa deletion (steps)									Average effect
	ITS baseline distribution to ITS-cpDNA observed distribution			AGT1 baseline distribution to AGT1-cpDNA observed distribution			cpDNA baseline distribution to ITS-cpDNA observed distribution			
	A	B	C	A	B	C	A	B	C	
<i>L. glauca</i> ssp. <i>olcadium</i> *	-4 <sup>§</sup>	-2	-4	-3	-2	-2	-2	-2	-2	-2.56
<i>L. orbensis</i> *	-4	0	-4	+1	+1	+1	-2	-2	-2	-1.22
<i>L. amethystea</i> ssp. <i>amethystea</i> *	-3	-1	-3	-1	0	+1	-1	-1	-1	-1.11
<i>L. cuartanensis</i> *	+1	-1	-1	-1	-1	-1	-1	-2	-2	-1.00
<i>L. tursica</i> *	+2	+2	0	-2	-2	-1	-2	-2	0	-0.56
<i>L. oblongifolia</i> ssp. <i>oblongifolia</i> *	0	0	0	-2	-1	-1	0	0	0	-0.44
<i>L. alpina</i> *	0	+2	0	-1	-1	+1	0	-2	-2	-0.33
<i>L. filicaulis</i> *	0	0	0	-1	-1	-1	0	0	0	-0.33
<i>L. saturejoides</i> ssp. <i>saturejoides</i> *	0	0	0	-2	0	0	0	0	0	-0.22
<i>L. propinqua</i> *	-2	+1	-2	-1	+1	+1	0	0	0	-0.22
<i>L. supina</i> ssp. <i>supina</i>	0	+2	-2	-1	+1	+1	0	0	0	+0.11
<i>L. bubanii</i>	0	+2	-2	+1	-1	+1	0	-1	+1	+0.11
<i>L. bipunctata</i> ssp. <i>bipunctata</i>	-2	+2	0	-1	0	0	0	+2	0	+0.11
<i>L. saxatilis</i>	-2	+3	-2	+1	0	+1	0	0	0	+0.11
<i>L. badalii</i>	-1	+3	-1	-1	-1	-1	+1	+1	+1	+0.11
<i>L. almijarensis</i>	+1	+3	+1	-1	-1	-1	+1	0	0	+0.33
<i>L. mumbyana</i>	+1	+3	+1	-1	0	-1	+1	0	0	+0.44
<i>L. aegyptiaca</i>	+1	+1	+1	0	+1	+1	0	-1	0	+0.44

**A** *L. glacialis*  $N_e$ : Nuclear  $N_e$  = 190000, Plastid  $N_e$  = 95000.

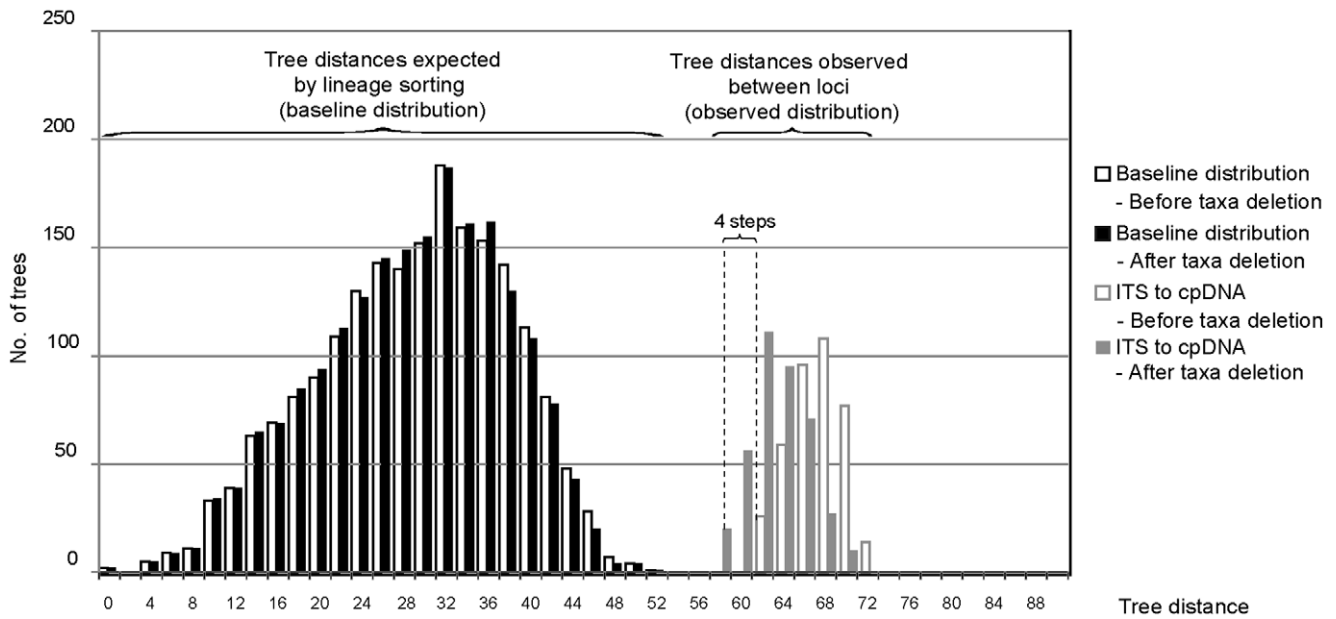
**B** *L. elegans*  $N_e$ : Nuclear  $N_e$  = 320000, Plastid  $N_e$  = 160000.

**C** *L. simplex*  $N_e$ : Nuclear  $N_e$  = 680000, Plastid  $N_e$  = 340000.

\*Individuals of putative hybrid origin that were excluded from the analysis in Figure 4.

<sup>§</sup>Calculation plotted as an example in Fig. 3.

doi:10.1371/journal.pone.0039089.t002



**Figure 3. Detection of potential hybrids.** An example illustrating the method used for the detection of potential hybrids. It is shown the effect of the exclusion of *L. glauca* ssp. *olcadium* on the differences between base line and observed distributions of tree distances.  
doi:10.1371/journal.pone.0039089.g003

detected by a single method as done in *Poke et al.* [96]. Therefore, we proceeded without removing the sequences under discussion.

**Gene Tree Inference**

ITS phylogenetic analysis supported monophyly for section *Supinae* sister to a group formed by four species of sect. *Diffusae* (*L. laxiflora*, *L. warionis*, *L. haelava*, *L. joppensis*). In ITS, relationships within *Supinae* were not clearly related to morphological features (Fig. 1A). The AGT1 region did not support monophyly of the section, as species of sect. *Diffusae* and sect. *Versicolores* were grouped together with sect. *Supinae*. The three *Supinae* groups detected in AGT1 were also not obviously correlated with morphological characters (Fig. 1B). The cpDNA dataset did not support monophyly of the section, as there were two clearly separated groups of *Supinae* species, however, this locus showed three well-supported groups within *Supinae* associated with corolla sizes and seed shape (Fig. 1C).

**Coalescent Simulations**

When using the small and medium (*L. glacialis* and *L. elegans*, respectively)  $N_e$  estimates (Table S2), the pairwise distances of gene trees lay outside the base line distribution for either gene (Fig. 2). Contrastingly, when using the largest  $N_e$  values (from widespread *L. simplex*), the pairwise distances of gene trees lay inside the base line distribution of ITS and AGT1 genes (Fig. 2). As we expected a high overestimation of the population size when using *L. simplex*  $N_e$ , these results reflected that the degree of incongruence in the three gene trees was difficult to explain by incomplete lineage sorting alone when applying Maureira-Butler’s test [6].

**Detection of Potential Hybrids**

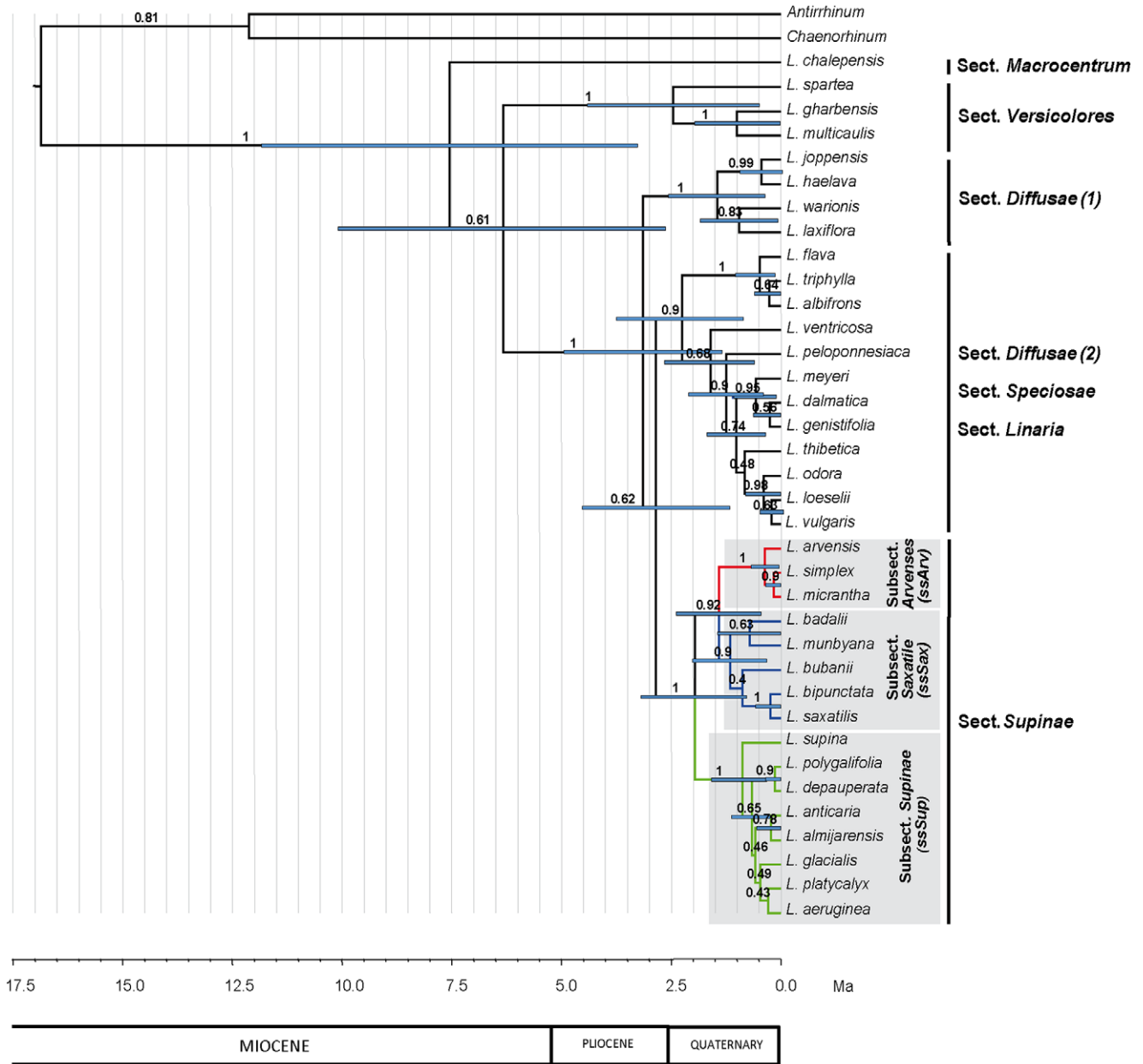
When using simulations obtained with medium  $N_e$  values (*L. elegans*), only one individual needed to be removed in order to retrieve overlapping baseline and observed distributions (not shown), and therefore only one potential hybrid could be considered robustly detected. In contrast, when using simulations

**Table 3.** Results of Shimodaira-Hasegawa (S-H) test and Bayes Factors (BF) test with observed log-likelihood difference obtained in Maximum Likelihood analyses, S-H test statistics, mean values of marginal likelihood of the Bayesian analyses and BF test statistics (2xlnBF) for the unconstrained analysis ( $H_0$ ) and the analysis with monophyly of *Supinae* constrained in AGT1 and cpDNA datasets ( $H_1$ ).

Gene tree	Hypothesis (H)	S-H test		BF test	
		Observed log-likelihood difference	SH statistic	Marginal likelihood (lnP(model   data)) ± SE	2xlnBF (H vs. H0)
AGT1	$H_0$	–	–	–3513.183 ± 0.27	–
	$H_1$ : monophyly of sect. <i>Supinae</i>	35.9	0.01*	–3547.955 ± 0.33	–69.546**
cpDNA	$H_0$	–	–	–4461.125 ± 0.27	–
	$H_1$ : monophyly of sect. <i>Supinae</i>	35.4	0.01*	–4488.746 ± 0.26	–55.242**

\*≤0.05, support for rejection of  $H_1$ .  
\*\*≤10, very strong evidence for rejection of  $H_1$ .  
doi:10.1371/journal.pone.0039089.t003





**Figure 4. Species tree of *Linaria*.** Maximum clade credibility tree obtained in the \*BEAST species tree analysis after excluding potential hybrids and using allelic data of ITS, AGT1 and cpDNA datasets. Node bars represent the 95% highest posterior density intervals for the divergence time estimates of nodes with posterior probabilities above 0.50. Values above branches indicate Bayesian posterior probabilities. *Linaria* sections following Sutton (1988) are shown. Colors and clade labels represent the systematic nomenclature for *Supinae* as suggested in this paper. doi:10.1371/journal.pone.0039089.g004

**Table 4. Morphological key traits of the subsections proposed for section *Supinae* regarding the results obtained in the \*BEAST species tree analysis of ITS, AGT1 and cpDNA sequences (Figure 4).**

	<b>Subsect. <i>Arvenses</i></b>	<b>Subsect. <i>Saxatile</i></b>	<b>Subsect. <i>Supinae</i></b>
<b>Corolla size</b>	<b>Small (2.5–9 mm)</b>	<b>Medium (6–18 mm)</b>	<b>Large (16–31 mm)</b>
Seed wing shape	Thick-wide	Thick-wide/narrow (or absent)	Membranous-wide
Life-form	Annual	Annual/Perennial	Perennial

doi:10.1371/journal.pone.0039089.t004

**Table 5.** Divergence dates of clades of *Linaria* sect *Supinae*, presented as mean crown ages and 95% highest posterior density (HPD) intervals based on the \*BEAST species tree analysis (Figure 4).

Clade/Lineage	Mean age of divergence (Ma)	95% HPD interval
Genus <i>Linaria</i>	7.55	3.57–12.14
Sect. <i>Supinae</i>	1.97	0.87–3.28
Sect. <i>Supinae</i> subsect. <i>Arvenses</i>	0.36	0.08–0.72
Sect. <i>Supinae</i> subsect. <i>Saxatile</i>	1.16	0.39–2.08
Sect. <i>Supinae</i> subsect. <i>Supinae</i>	0.87	0.31–1.58

doi:10.1371/journal.pone.0039089.t005

with the smallest  $N_e$  values (*L. glacialis*), even after removing all the individuals with incongruent positions, we still had non-overlapping distributions (not shown), and consequently all species with incongruent positions (17 spp.) were identified as potential hybrids. Therefore, our  $N_e$  estimates showed all possible scenarios: (i) gene tree incongruence is explained by incomplete lineage sorting alone (*L. simplex*  $N_e$ ), (ii) gene tree incongruence is explained by both incomplete lineage sorting and hybridization (*L. elegans*  $N_e$ ) and (iii) gene tree incongruence is explained by hybridization alone (*L. glacialis*  $N_e$ ). These results clearly illustrated the high dependence on  $N_e$  estimates in order to obtain the exact number of individuals of hybrid origin. We assumed that a reliable number of potential hybrids lay between the two extreme values obtained in (ii) and (iii).

The effect of the deletion of each incongruent individual on both the observed and base line distributions is shown in Table 2. We considered that individuals with the highest probability of hybrid origin were those individuals that, after deletion, decreased (on average) the differences between the base line and the observed distributions (in number of steps, see an example in Fig. 3). Ten of the 17 incongruent individuals decreased the differences among distributions and consequently were considered to be potential hybrids or to have a hybrid history in the broadest sense.

### Testing Monophyly of *Supinae*

After excluding putative hybrids, the S-H tests indicated that the constrained topologies for AGT1 and cpDNA had significantly worse likelihood scores than the unconstrained topologies (Table 3), thus monophyly of *Supinae* for these genes was statistically rejected. The BF test (Table 3) also recovered decisive (very strong) support ( $2 \times \ln \text{BF} < -10$ ) for rejection of monophyly of *Supinae* in the AGT1 and cpDNA. As monophyly of *Supinae* was recovered in ITS (Fig. 1A), topological incongruence in concert with S-H and BF test suggested that processes other than hybridization and inference limitations were also responsible for the topological incongruence among genes.

### Species Tree Inference

The \*BEAST species tree analysis (potential hybrids excluded) (Fig. 4) retrieved four well supported groups within *Linaria*: (i) sect. *Versicolores* (1 PP), (ii) four species of sect. *Diffusae* (1 PP), (iii) a group formed by: three species of sect. *Diffusae*, four species of sect. *Speciosae* and five species of sect. *Linaria* (0.9 PP); and (iv) all sect. *Supinae* species (1 PP). Therefore sect. *Supinae* was retrieved as a monophyletic group with high support and was divided in three clades: one clade was represented by three annual species (*L. arvensis*, *L. simplex*, *L. micrantha*; 1 PP) with small corollas (2.5–9 mm) and a thick-wide seed wing (subsect. *Arvenses*, hereafter *ssArv*). A second clade was represented by five annual or perennial

species (*L. badalii*, *L. munbyana*, *L. bubanii*, *L. bipunctata*, *L. saxatilis*; 0.90 PP) with medium-sized corollas (6–18 mm) and a thick-wide seed wing or narrow wing (marginal ridge) (subsect. *Saxatile*, hereafter *ssSax*). The third clade contained eight perennial species (*L. supina*, *L. polygalifolia*, *L. depauperata*, *L. anticaria*, *L. almijarensis*, *L. glacialis*, *L. platycalyx*, *L. aeruginea*; 1 PP) with large corollas (16–31 mm) and a membranous-wide seed wing (subsect *Supinae*, hereafter *ssSup*) (see Table 4).

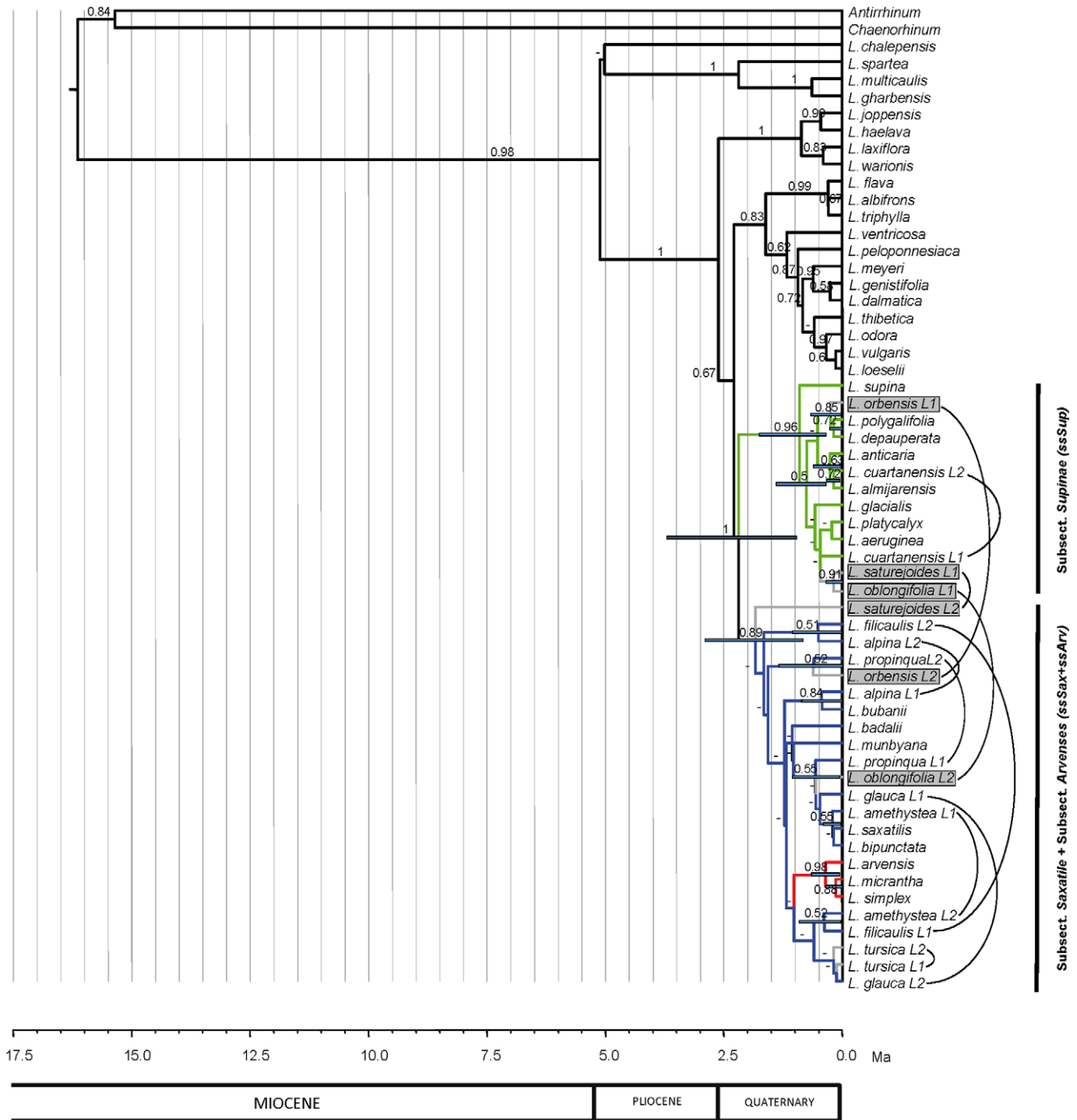
The \*BEAST species tree detected that incomplete lineage sorting has affected all gene trees analyzed. In the ITS dataset we detected deep coalescence at medium depth branches (see *L. bubanii* position in the ITS tree and \*BEAST species tree); from the AGT1 dataset we detected deep coalescence at medium depth branches (*L. munbyana*, *L. badalii*) and at deeper branches (*L. polygalifolia*, *L. depauperata*, *L. orbensis*, *L. anticaria*, *L. almijarensis*, *L. aeruginea*, *L. glacialis* and *L. platycalyx*); in cpDNA we also detected deep coalescence at the deepest branches (*L. badalii*, *L. bubanii*, *L. munbyana*, *L. bipunctata* and *L. saxatilis*).

The time to the most recent common ancestor (TMRCA) of *Supinae* was placed in the late Pliocene-early Pleistocene (0.87–3.28 Ma), the TMRCA of *ssArv* was located in the middle-late Pleistocene (Ionian-Tarantian) (0.08–0.72 Ma), the TMRCA of *ssSax* in the early-middle Pleistocene (Gelasian-Calabrian-Ionian) (0.39–2.08 Ma) and the TMRCA of *ssSup* in the early-middle Pleistocene (Calabrian-Ionian) (0.31–1.58 Ma) (see Table 5).

### Multilabelled Species Tree

The multilabelled species tree (Fig. 5) retrieved a well supported clade (0.96 PP, *ssSup*) and a clade with moderate support (0.89 PP, *ssSax+ssArv*) within *Supinae*. Out of ten reticulation events that have been presumed to occur, one was produced within the *ssSup* clade, six within the *ssSax+ssArv* clade and three between these two clades. One of the six potential hybridization events within *ssSax+ssArv* clade is reflected in *L. tursica*, a species with morphological traits typical from both *ssSax* and *ssArv* clades (Fig. 4): wingless seed (some species of *ssSax* present narrow to marginal seed wings) and small corolla (*ssArv*). The three reticulation events inferred between *ssSup* and *ssSax+ssArv* produced three species with morphological traits typical of both clades (*L. orbensis*, *L. saturejoides* and *L. oblongifolia*, see Fig. 5 and Table 6).

We estimated the timing of the hybridization events by looking at the divergence time of parental lineages of putative hybrids. As hybridization could not take place prior to divergence of parental lineages, divergence time for the most recent lineage constituted the maximum age of each hybridization event. Despite the topological uncertainty at the tips, we found that all but one maximum age of the presumed hybridization episodes occurred during the Pleistocene (Fig. 5 and Table 7). In a single case, *L. tursica*, the 95% HPD overlapped the Middle Pliocene, although the mean estimate remained within the Pleistocene (Table 7).



**Figure 5. Reticulate evolution in *Supinae*.** Maximum clade credibility tree obtained in the multilabelled \*BEAST species tree analysis by including the presumed hybrids connected in two labels (L1 and L2) representing the two parental lineages of hybrid species. Node bars represent the 95% highest posterior density intervals for the divergence time estimates of nodes with posterior probabilities above 0.50 (only divergence time estimates for *Supinae* lineages are shown). Values above branches indicate Bayesian posterior probabilities. A hyphen (-) indicates posterior probability below 0.50. Colors and tree labels represent the systematic nomenclature for *Supinae* as established in this paper. Species labels of putative hybrids produced by the cross of the two main *Supinae* clades are highlighted in grey. doi:10.1371/journal.pone.0039089.g005

**Discussion**

**Using a Coalescent Framework to Disclose the Evolutionary History of *Supinae***

Systematics of *Linaria* and specifically sect. *Supinae* has been subject to various interpretations in numerous taxonomic treat-

ments in the last two centuries. Historical disagreement occurred when discerning the naturalness of the section and its internal classification [43–51] (see Table 1). To disclose the evolutionary history of *Supinae*, we sampled genetic data from 46 *Linaria* species, including sequences from three presumably unlinked genes. Because of the highly supported incongruence among trees based

**Table 6.** Morphological key traits of species with putative hybrid origin produced by the cross between subsect *Saxatile* + subsect *Arvenses* (*ssSax+ssArv*) and subsect *Supinae* (*ssSup*) parental lineages based on the results obtained in the \*BEAST multilabelled species tree analysis (Figure 5).

	<i>L. orbensis</i>	<i>L. saturejoides</i>	<i>L. oblongifolia</i>
<b>Corolla size</b>	<b>Medium (11–15 mm)</b>	<b>Medium (12–17 mm)</b>	<b>Medium-large (15–22 mm)</b>
Seed wing shape	Membranous-wide	Membranous-wide	Membranous-wide
Life-form	Annual	Annual	Annual

doi:10.1371/journal.pone.0039089.t006

on separate analysis of the three genes, difficulty in the systematic reconstruction of *Supinae* at this stage of analysis was patent, the naturalness of the section remained unclear and the infra-sectional classification was still controversial.

In the last few years the incorporation of the coalescent model into phylogenetic analysis has greatly improved the theoretical basis for inferring species trees from gene trees via a mixed model—the multispecies coalescent (e.g., BEST [28]; \*BEAST [20]; [97]). One key practical challenge is to include only data that meet the assumptions of the current implementations. Of significant concern is to properly handle sequences, individuals or taxa with multiple histories, such as by excluding recombinants or hybrids prior to species tree inference.

Here, we performed simulations under coalescence following the method of Maureira-Butler *et al.* [6] to estimate whether the gene tree incongruence detected among genes could be explained by incomplete lineage sorting without hybridization. The test exposed that with small and medium  $N_e$  values used in the simulations, the topological variation generated by incomplete lineage sorting was not as high as the incongruence observed between the three genes (Fig. 2), whereas with high  $N_e$  (*L. simplex*  $N_e$ ), the variation generated by incomplete lineage sorting alone could explain the totality of incongruence observed between genes (Fig. 2). We considered that the high  $N_e$  greatly overestimated the general  $N_e$  of *Linaria*, as only 9 out of 150 *Linaria* species have a similar wide range size [45] (and presumably similar  $N_e$ ). Hence, the results of Maureira-Butler's test suggested that incongruence among genes was difficult to explain by incomplete lineage sorting alone, indicating that hybridization may also account for the gene tree inconsistency. However, the exact number and

identity of individuals that may have hybrid histories is not clearly established here, because of the sensitivity of the test to  $N_e$  estimation. We consider, instead, that the test has provided a probable set of individuals that may adversely affect the \*BEAST analysis and that a cautious approach (removing these individuals before the analysis) is preferred here, rather than risking a spurious species tree inference.

The hybrid detection test (Table 2) and the multilabelled \*BEAST species tree (Fig 5) was also contrasted with a \*BEAST species tree including all potential hybrids (not shown). After six runs with 30 million generations, convergence could not be reached and some ESS values (of population size parameters) remained under 200, which illustrated that the inclusion of potential hybrids may be violating assumptions of the \*BEAST analysis. Our approach was also contrasted with an additional analysis of the three datasets concatenated in a total evidence approach (see Fig. S1). Results of both approaches (our multilabelled species tree with hybrids excluded vs. the total evidence analysis) gave highly conflicting results. These discordant results were expected, as it is known that concatenation of data from multiple loci may lead to biased phylogenetic estimates under widespread incomplete lineage sorting and/or hybridization [19]. Results presented here highlight the paramount importance of (i) analyzing multiple loci datasets in a multispecies coalescent approach in order to find a more realistic species tree and (ii) the requirement of additional analytical tools to identify and to disclose the origin of species affected by historical hybridization. We note that our multilabelled species tree still allows the possibility of observing congruent placements for each label of the same individual. That is, we are not forcing different placements with this approach, but instead allowing them, if preferred by the data. Therefore, this

**Table 7.** Divergence dates of parental lineages of hybrid species presented as mean age of divergence and 95% highest posterior density (HPD) intervals based on the \*BEAST multilabelled species tree analysis (Figure 5).

Hybrid taxa	Mean age of divergence from most recent parental lineage (Ma)	95% HPD interval	Mean age of divergence from 2 <sup>nd</sup> parental lineage (Ma)	95% HPD interval
<i>L. glauca</i> ssp. <i>olcadium</i>	0.61	0.17–1.17	1.53	0.62–2.68
<i>L. orbensis</i>	0.29	0.03–0.67	0.51	0.00–1.35
<i>L. amethystea</i> ssp. <i>amethystea</i>	0.17	0.02–0.38	0.36	0.00–0.89
<i>L. cuartanensis</i>	0.10	0.00–0.28	0.61	0.23–1.09
<i>L. tursica</i>	1.53	0.62–2.68	1.53	0.62–2.68
<i>L. oblongifolia</i> ssp. <i>oblongifolia</i>	0.12	0.00–0.36	0.61	0.17–1.17
<i>L. alpina</i>	0.35	0.00–0.86	0.35	0.00–1.04
<i>L. filicaulis</i>	0.35	0.00–1.04	0.36	0.00–0.89
<i>L. saturejoides</i> ssp. <i>satirejoides</i>	0.12	0.00–0.36	1.53	0.62–2.68
<i>L. propinqua</i>	0.51	0.00–1.35	0.61	0.17–1.17

doi:10.1371/journal.pone.0039089.t007

**Table 8.** Previous phylogenetic studies of Mediterranean plants with highly supported incongruence among gene trees, we indicate those articles that claim hybridization and/or incomplete lineage sorting as major causes of topological inconsistency.

Mediterranean plant group	Suggested cause of incongruence		Reference
	Hybridization	Incomplete lineage sorting	
<i>Antirrhinum</i>	✓	✗	[102]
<i>Euphorbia</i> sect. <i>Aphyllis</i>	✓	✗	[110]
<i>Anthemis</i>	✓	✗	[111]
<i>Centaurium</i>	✓	✗	[112]
<i>Heliosperma</i>	✓	✗	[11]
<i>Reseda</i> sect. <i>Glaucosceda</i>	✗	✓	[100]
<i>Ptilostemon</i>	✗	✓	[113]
<i>Hordeum</i>	✗	✓	[114]
<i>Amarillidaceae</i> (Mediterranean clade)	✗	✓	[115]
<i>Achillea</i>	✓	✓	[116]
<i>Senecio</i> sect. <i>Senecio</i>	✓	✓	[117]
<i>Arenaria</i> sect. <i>Plinthine</i>	✓	✓	[118]
<i>Phlomis crinita/lychnitis</i> complex	✓	✓	[119]
<i>Medicago</i>	✓	✓*	[6]

\*Although not explicitly discussed, incongruence due to incomplete lineage sorting is also apparent among gene trees in this paper.  
doi:10.1371/journal.pone.0039089.t008

approach appears to combine the ideals of utilizing the available comparable data sets (including hybrids) while also appropriately accommodating processes that may cause incongruence (incomplete lineage sorting) and could otherwise lead to spurious tree inference.

### Systematics and Drivers of Evolution in *Supinae*

The *Linaria* \*BEAST species tree retrieved three well supported clades that agreed with previous classifications (Fig 4): (i) Sect. *Versicolores*, (ii) four species of Sect. *Diffusae* and (iii) Sect. *Supinae*. It also retrieved a group that was incongruent with previous taxonomic treatments. This latter group contained three species of Sect. *Diffusae*, four species of Sect. *Speciosae* and five species of Sect. *Linaria*. In this analysis *Supinae* was monophyletic, as found in the ITS phylogeny. Furthermore, *Supinae* was divided into three morphologically-based subclades consistent with life-form, corolla size and seed wing shape (Table 4), as found in the cpDNA phylogeny: subsect. *Supinae* (*ssSup*), subsect. *Arvenses* (*ssArv*) and subsect. *Saxatile* (*ssSax*). These results are strikingly consistent with some earlier hypotheses, despite the incongruence observed among gene trees. *ssSup* contained eight species that were grouped together in several previous morphological classifications, *ssArv* contained three species that were also previously grouped in a taxonomical entity, whereas *ssSax* contained five species that were historically placed in several distinct taxonomic groups (Systematic proposal in Table 1, diagnostic characters in Table 4). Corolla size and seed wing shape were also previously used as diagnostic characters in a morphological taxonomic revision of winged-seeded *Linaria* species [46]. This author considered *Arvenses* (*ssArv*) (small flowers) as an independent section and divided *Supinae* in three subsections according to life form and seed wing shape: (i) subsect. *Supinae* (*ssSup*): perennial plants with membranous seed wings, (ii) subsect. *Amethystea*: annual plants with thick seed wings and (iii) subsect. *Saxatile*: annual or perennial plants with somewhat thin wings.

Reproductive biology and interaction with pollinators may have played an important role in differentiation within *Supinae*. This is supported by the fact that the species with very low investment in flower structures (small corollas, *ssArv*) are all self-compatible,

whereas species with a high investment in flower formation (large corollas, *ssSup*) are all self-incompatible, mainly pollinated by large bees and with low pollinator diversity (Blanco-Pastor & Vargas, unpublished). Geography appears to have played a role in structuring the diversity within *Supinae* as the diversity of *ssSax* is located in the northern part of the Iberian Peninsula (three out of the five species are northern Iberian endemics), whereas the diversity of *ssSup* is located in southern Iberia (five out of eight species are southern Iberian endemics). The timing of divergence of the three subclades (crown nodes, Table 5) indicates that diversification occurred during the Quaternary, after the establishment of the Mediterranean climate regime [39], when species had to tolerate the climatic oscillations occurring in that period [98,99]. This pattern of geographical differentiation driven by Quaternary interglacial fragmentation has been previously identified in many Iberian plants [36,37,100], including the closely-related genus *Antirrhinum* [101,102].

### Hybridization during the Quaternary Glaciations

We found that historical hybridization has been likely during the course of *Supinae* evolution. Our analyses identified 10 out of 17 individuals with incongruent positions in gene trees that were difficult to reconcile with incomplete lineage sorting (Table 2). Simple introgression (that is, recurrent horizontal gene flow toward one parental species without formation of new species) can explain the observed gene tree incongruence in those individuals. But the observed pattern could have been also generated by homoploid hybrid speciation (all *Linaria* species analyzed here are diploid ( $2n = 12$ ) excluding *L. chalepensis* ( $2n = 24$ )). Despite speciation via homoploid hybridization has been historically hard to detect (as it could present a similar signal to simple introgression or incomplete lineage sorting) [31], recent studies have suggested that it might be an important mechanism for plant speciation [58,103,104]. Our analyses do not validate speciation via homoploid hybridization, but this process must not be discarded as potential generator of diversity in *Supinae*.

The multilabelled \*BEAST species tree analysis (Fig. 5) recovered, to some degree, the origin of the parental alleles of

individuals affected by historical hybridization. There is bound to be a loss of power, because of the reduced number of loci available to place the multilabelled species as well as the need to use missing data. Even so, out of ten potential hybridization events detected, our analyses suggested that one occurred within the *ssSup* lineage, three between two distant parental lineages (*ssSax+ssArv* and *ssSup*) and six within the *ssSax+ssArv* lineage. Crosses between the two distant parental lineages retrieved in the analysis (*ssSax+ssArv* and *ssSup*) were also supported by morphology, given that those three taxa (*L. orbensis*, *L. oblongifolia* and *L. saturejoides*) presented morphological key traits from both clades (Table 6). All hybridization events inferred here were also supported by the results obtained in experimental crosses performed by Valdés [53]. In that study, this author obtained fruits in one of the four crosses performed among *ssSup* species, three of the four crosses between *ssSax+ssArv* and *ssSup* species and four of the seven crosses among *ssSax+ssArv* species (note that here we only accounted for crosses produced between species used in this study thus a higher number of total successful crosses were produced, see [53]).

The maximum age of a hybridization event was considered here to be the maximum age of the origin of the most recent parental lineage. Those ages were circumscribed between 0.28–1.35 Ma in nine of the ten potential hybrids (Table 7). Only in *L. tursica* did the maximum age of hybridization surpass 2.5 Ma (2.68 Ma). Taking into account the effect of low phylogenetic resolution that obscured the detection of ages in parental lineages (thus considering the maximum age of hybridization at deeper nodes), the present results lead us to affirm that all potential hybridization events detected but one may have occurred during the Pleistocene climatic oscillations. During the Quaternary, hybrid zones were established in contact zones (Pyrenees, Alps, Central Europe and Scandinavia) of interglacial northward colonization routes from the temperate regions of Europe [105,106]. In the Iberian Peninsula, where ice effects were less severe, subsequent patterns of contraction, fragmentation, persistence, expansion and admixture during altitudinal migrations may have repeatedly produced multiple hybrid zones [33,34,99]. The complex Iberian orography may have allowed partial differentiation of lineages in allopatry but subsequent secondary contacts of differentiated genomes from close locations [34]. That may have been the framework for *Linaria* and many other southern European plant groups (Table 8). Clearly, the investigation of hybridization in Mediterranean plant groups is vital for the accurate inference of species trees, as well as to understand the role of hybridization in the generation of new genetic combinations and morphological differentiation. However, we have shown in this example that existing tools, although limited, can nonetheless provide valuable insights in these areas.

### Incomplete Lineage Sorting as a Significant Process in Mediterranean Plants

Several studies have claimed incomplete lineage sorting as a major cause of gene tree incongruence and non-monophyly in Mediterranean plants (Table 8). Failure of gene lineages to coalesce occurs when the time between speciation events is very short and/or when the effective population size of the ancestral populations is very large [2]. We detected incomplete lineage sorting in all independent loci analyzed for *Linaria*. In this genus, population size estimates obtained by using three *Linaria* species (*L. glacialis*, *L. elegans*, *L. simplex*) suggested that ancestral populations may have not been extremely large (see [107] for comparison). Conversely, extremely rapid divergence of ancestral populations

seems more likely. *Linaria* has diversified since the late Miocene-early Pliocene (3.57–12.14 Ma) (crown node of the genus, Table 5) to recent times in the late Quaternary (Table 5, Fig. 4). During its evolutionary history, this Mediterranean group may have experienced drastic climatic events such as the Messinian Salinity Crisis (5.96 Ma) [108], the catastrophic flood that caused the refilling of the Mediterranean Sea (5.33 Ma) [109], the progressive establishment of the Mediterranean rhythm with dry summers (3.2 Ma) and the Quaternary type oscillations with glacial and interglacial stages (2.3 Ma) [39]. These extreme climatic changes coupled with the irregular mountain ranges of the Mediterranean basin might have promoted rapid diversification driven by isolation in reduced areas causing rapid allopatric speciation. The secondary contacts occurring during the climatic oscillations seem to have promoted historical hybridization between closely related *Linaria* species, but also the high number of species in the Mediterranean (104 spp.) [45] and its recent origin suggest that this group is likely to have undergone rapid diversification. Additional analyses not performed here are proposed to confirm rapid speciation as the cause for incomplete lineage sorting in *Linaria*.

The basis underlying phylogenetic incongruence may vary depending on the plant group under study, but the flora of the Mediterranean is formed, in part, by many genera that similarly display numerous species generated in short periods of time that also may have suffered secondary contacts in short term cycles (20,000–100,000 yr.). In these groups incomplete lineage sorting and hybridization appear to be the rule rather than the exception.

### Supporting Information

**Figure S1 Total evidence analysis.** The 50% majority-rule consensus tree obtained in the Bayesian analysis of the concatenated ITS, AGT1 and cpDNA datasets. Numbers above branches are Bayesian posterior probabilities. Colors represent the systematic nomenclature for *Supinae* as suggested in this paper (see Fig. 4). Species with intermediate key traits are represented in grey. (TIF)

**Table S1 List of taxa included with localities, collector's numbers and Genbank accession numbers.** (DOCX)

**Table S2 Effective population size estimates ( $N_e$ ) used in the coalescent simulations.** (DOCX)

**Table S3 Assignment of genes to label 1 (L1) or label 2 (L2) in the multilabelled species tree analysis (Fig. 5).** (DOCX)

**Methods S1 Supplemental methods.** (DOCX)

### Acknowledgments

We would like to thank M. Fernández-Mazuecos for his encouraging support, I. Liberal for his help in the laboratory and B. Oxelman, A. Petri, T. Marcussen, E. Sjökvist, F. de Sousa and Y. Bertrand for their comments and helpful discussions. We also thank the MA, RNG, E herbaria and the Flora Iberica Project for providing us with plant material.

### Author Contributions

Conceived and designed the experiments: JLB-P PV BEP. Performed the experiments: JLB-P BEP. Analyzed the data: JLB-P BEP. Contributed reagents/materials/analysis tools: JLB-P PV BEP. Wrote the paper: JLB-P PV BEP.

## References

- Doyle JJ (1992) Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144–163.
- Maddison WP (1997) Gene Trees in Species Trees. *Systematic Biology* 46: 523–536.
- Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular biology and evolution* 5: 568–583.
- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Review Genetics* 3: 380–390.
- Wendel JF, Doyle JJ (1998) Phylogenetic incongruence: Window into genome history and molecular evolution. Pages 265–296. In: Soltis DE, Soltis PS, Doyle JJ, editors. *Molecular systematics of plants II: DNA Sequencing*. Norwell: Kluwer Academic Publishers.
- Maureira-Butler IJ, Pfeil BE, Muangprom A, Osborn TC, Doyle JJ (2008) The reticulate history of *Medicago* (Fabaceae). *Systematic Biology* 57: 466–482.
- Buckley TR, Cordeiro M, Marshall DC, Simon C (2006) Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (Maoricicada Dugdale). *Systematic Biology* 55: 411–425.
- Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). *Evolution* 61: 1992–2006.
- van der Niet T, Linder HP (2008) Dealing with incongruence in the quest for the species tree: A case study from the orchid genus *Satyrium*. *Molecular Phylogenetics and Evolution* 47: 154–174.
- Joly S, McLenachan PA, Lockhart PJ (2009) A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist* 174: E54–E70.
- Frajman B, Eggens F, Oxelman B (2009) Hybrid origins and homoploid reticulate evolution within *Heliosperma* (Sileneae, Caryophyllaceae)—A multigenic phylogenetic approach with relative dating. *Systematic Biology* 58: 328–345.
- Stone RD, Andreasen K (2010) The Afro-Madagascan genus *Warneckea* (Melastomataceae): Molecular systematics and revised infrageneric classification. *Taxon* 59: 83–92.
- Wong SY, Boyce PC, Sofiman bin Othman A, Pin LC (2010) Molecular phylogeny of tribe Schismatoglottideae (Araceae) based on two plastid markers and recognition of a new tribe, Philonotieae, from the neotropics. *Taxon* 59: 117–124.
- Link P, rez MA, Watson LE, Hickey RJ (2011) Redefinition of *Adiantopsis* Fee (Pteridaceae): Systematics, diversification, and biogeography. *Taxon* 60: 1255–1268.
- Martínez-Azorín M, Crespo MB, Juan A, Fay MF (2011) Molecular phylogenetics of subfamily Ornithogaloideae (Hyacinthaceae) based on nuclear and plastid DNA regions, including a new taxonomic arrangement. *Annals of Botany* 107: 1–37.
- Liu ZW, Zhou J, Liu ED, Peng H (2010) A molecular phylogeny and a new classification of *Pyrola* (Pyroleae, Ericaceae). *Taxon* 59: 1690–1700.
- Emadzade K, Lehnbach C, Lockhart P, Hörandl E (2010) A molecular phylogeny, morphology and classification of genera of Ranunculaceae (Ranunculaceae). *Taxon* 59: 809–828.
- Roxanne Steele P, Friar LM, Gilbert LE, Jansen RK (2010) Molecular systematics of the neotropical genus *Psiguria* (Cucurbitaceae): Implications for phylogeny and species identification. *American Journal of Botany* 97: 156–173.
- Kubatko LS, Degnan JH (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* 56: 17–24.
- Heled J, Drummond AJ (2010) Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution* 27: 570–580.
- Yu Y, Cuong T, Degnan JH, Nakhleh L (2011) Coalescent Histories on Phylogenetic Networks and Detection of Hybridization Despite Incomplete Lineage Sorting. *Systematic Biology* 60: 138–149.
- Bloomquist EW, Suchard MA (2010) Unifying Vertical and Nonvertical Evolution: A Stochastic ARG-based Framework. *Systematic Biology* 59: 27–41.
- Holland B, Benthin S, Lockhart P, Moulton V, Huber K (2008) Using supernetworks to distinguish hybridization from lineage-sorting. *BMC Evolutionary Biology* 8: 202.
- Ané C, Larget B, Baum DA, Smith SD, Rokas A (2007) Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24: 412–426.
- Edwards SV, Liu L, Pearl DK (2007) High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences* 104: 5936–5941.
- Carstens BC, Knowles LL (2007) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from *Melanoplus* grasshoppers. *Systematic Biology* 56: 400–411.
- Liu L, Pearl DK, Brumfield RT, Edwards SV (2008) Estimating species trees using multiple-allele DNA sequence data. *Evolution* 62: 2080–2091.
- Liu L, Yu L, Pearl DK, Edwards SV (2009) Estimating Species Phylogenies Using Coalescence Times among Sequences. *Systematic Biology* 58: 468–477.
- Fan HH, Kubatko LS (2011) Estimating species trees using approximate Bayesian computation. *Molecular Phylogenetics and Evolution* 59: 354–363.
- Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV (2009) Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution* 53: 320–328.
- Mallet J (2007) Hybrid speciation. *Nature* 446: 279–283.
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359: 183–195.
- Gómez A, Lunt D (2006) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N, editors. *Phylogeography of Southern European Refugia*. The Netherlands: Springer. 155–188.
- Feliner GN (2011) Southern European glacial refugia: A tale of tales. *Taxon* 60: 365–372.
- Belfiore NM, Liu L, Moritz C (2008) Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia: Geomyidae). *Systematic Biology* 57: 294–310.
- Fiz-Palacios O, Vargas P, Vila R, Papadopulos AST, Aldasoro JJ (2010) The uneven phylogeny and biogeography of *Erodium* (Geraniaceae): radiations in the Mediterranean and recent recurrent intercontinental colonization. *Annals of Botany* 106: 871–884.
- Valente LM, Savolainen V, Vargas P (2010) Unparalleled rates of species diversification in Europe. *Proceedings of the Royal Society B: Biological Sciences* 277: 1489–1496.
- Guzmán B, Lledó MD, Vargas P (2009) Adaptive Radiation in Mediterranean *Cistus* (Cistaceae). *Plos One* 4: e6362.
- Suc JP (1984) Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature* 307: 429–432.
- Fernández-Mazuecos M, Vargas P (2011) Historical Isolation versus Recent Long-Distance Connections between Europe and Africa in Bifid Toadflaxes (*Linaria* sect. *Versicolores*). *Plos One* 6.
- Hsu KJ, Montadert L, Bernoulli D, Cita MB, Erickson A, et al. (1977) History of the Mediterranean salinity crisis. *Nature* 267: 399–403.
- Vargas P, Rossello JA, Oyama R, Guemes J (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: 151–172.
- Rothmaler W (1943) Zur Gliederung der Antirrhineae. *Feddes Repertorium Specierum Novarum Regni Vegetabilis* 52: 16–39.
- Sáez L, Bernal M (2009) *Linaria* Mill. In: Castroviejo S, Herrero A, Benedi C, Rico E, Guemes J, editors. *Flora Ibérica* 13. Madrid: CSIC. 232–324.
- Sutton DA (1988) A revision of the tribe Antirrhineae. London. UK: Oxford University Press.
- Valdés B (1970) Revisión de las especies europeas de *Linaria* con semillas aladas. Sevilla. Spain: Anales de la Universidad Hispalense. 1–288 p.
- Wettstein (1895) Scrophulariaceae. In: Prantl E, editor. *Die Natürlichen Pflanzenfamilien*. 39–107.
- Bentham G (1846) Scrophulariaceae In: Candolle A-L-P-Pde, editor. *Prodromus Systematis Universalis Regni Vegetabilis*. 266–288.
- Viano J (1978) Les linaires à graines aptères du bassin méditerranéen occidental. 1. *Linaria* sect. *Versicolores*. *Candollea* 33: 33–88.
- Viano J (1978) Les linaires à graines aptères du bassin méditerranéen occidental. 2. *Linaria* sect. *Elegantae, Bipunctatae, Diffusae, Speciosae, Repentes*. *Candollea* 33: 209–267.
- Chavannes E (1833) *Monographie des Antirrhinées*. Paris et Lausanne.
- Valdés B (1970) Taxonomía experimental del género *Linaria* III. Cariología de algunas especies de *Linaria*, *Cymbalaria* y *Chaenorhinum*. *Boletín de la Real Sociedad Española de Historia Natural (Biología)* 67: 243–256.
- Valdés B (1970) Taxonomía experimental del género *Linaria* V. Hibridación interespecífica. *Acta Phytotaxonomica Barcinonensis* 4.
- Viano J (1978) Croisements expérimentaux interspécifiques au sein du genre *Linaria*. *Caryologia* 31(4): 383–425.
- Rouy G (1909) des Tribus et des Genres de la famille des Scrophulariacées. In: Deyrolle LFdÉ, editor. *Flore de France ou description des plantes qui croissent spontanément en France, en Corse et en Alsace-Lorraine*. Paris.
- Druce GC (1925) *L. repens* x *supina* nov. hybr. *Botanical Society and Exchange Club of the British Isles* 7: 998.
- Fournier P (1946) Les Quatre Flores de la France. Paris: Corse Comprise.
- Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology and Evolution* 19: 198–207.
- Liepmann AH, Olsen IJ (2001) Peroxisomal alanine : glyoxylate aminotransferase (AGT1) is a photorespiratory enzyme with multiple substrates in *Arabidopsis thaliana*. *The Plant Journal* 25: 487–498.
- White TJ, Bruns T., Lee S. and Taylor J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In "PCR Protocols. A Guide to Methods and Applications"; M.A. Innis DHG, J.J. Sminsky and T. J. White, Eds., editor. San Diego. USA: Academic Press.
- Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8: 521–523.
- Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. *American Journal of Botany* 94: 275–288.

63. Sáez L, Crespo MB (2005) A taxonomic revision of the *Linaria verticillata* group (Antirrhineae, Scrophulariaceae). *Botanical Journal of the Linnean Society* 148: 229–244.
64. del Campo P, del Amo M (1855) Especies de plantas nuevas descubiertas por D. Pedro del Campo y descritas por D. Mariano del Amo. *Revista de los Progresos de las Ciencias Exactas, Físicas y Naturales* 5: 55–58.
65. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.
66. Excoffier L, Laval G, Balding D (2003) Gametic phase stimation over large genomic regions using an adaptive window approach. *Human Genomics* 1:7–9.
67. Scheen A-C, Pfeil BE, Petri A, Heidari N, Nylinder S, et al. (2011) Use of allele-specific sequencing primers is an efficient alternative to PCR subcloning of low-copy nuclear genes. *Molecular Ecology Resources* 12: 128–135.
68. Heath L, Van Der Walt E, Varsani A, Martin DP (2006) Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *Journal of Virology* 80: 11827–11832.
69. Martin D, Rybicki E (2000) RDP: Detection of recombination amongst aligned sequences. *Bioinformatics* 16: 562–563.
70. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* 265: 218–225.
71. Smith JM (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* 34: 126–129.
72. Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses* 21: 98–102.
73. Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16: 573–582.
74. Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176: 1035–1047.
75. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* 98: 13757–13762.
76. Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
77. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9: 286–298.
78. Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular biology and evolution* 25: 1253–1256.
79. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307–321.
80. Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-revisited. *American Journal of Botany* 97: 1296–1303.
81. Sanderson MJ (2003) r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19: 301–302.
82. Heled J, Drummond A (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* 8: 289.
83. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475: 493–496.
84. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114–1116.
85. Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* 90: 773–795.
86. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution* 18: 1001–1013.
87. Shimodaira H, Hasegawa M (2002) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246–1247.
88. Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from BEAST Software website: <http://beast.bio.ed.ac.uk/Tracer>. Accessed November 2011.
89. Drummond A, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.
90. Gernhard T (2008) The conditioned reconstructed process. *Journal of Theoretical Biology* 253: 769–778.
91. Kay KM, Whittall JB, Hodges SA (2006) A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology* 6.
92. Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences* 84: 9054–9058.
93. Senchina DS, Alvarez I, Cronn RC, Liu B, Rong JK, et al. (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular Biology and Evolution* 20: 633–643.
94. Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. *Systematic Biology* 54: 441–454.
95. Pirie MD, Humphreys AM, Barker NP, Linder HP (2009) Reticulation, Data Combination, and Inferring Evolutionary History: An Example from Danthonioidae (Poaceae). *Systematic Biology* 58: 612–628.
96. Poke FS, Martin DP, Steane DA, Vaillancourt RE, Reid JB (2006) The impact of intragenic recombination on phylogenetic reconstruction at the sectional level in *Eucalyptus* when using a single copy nuclear gene (cinnamoyl CoA reductase). *Molecular Phylogenetics and Evolution* 39: 160–170.
97. Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 24: 332–340.
98. Postigo-Mijarra JM, Morla C, Barrón E, Morales-Molino C, García S (2010) Patterns of extinction and persistence of Arctotertiary flora in Iberia during the Quaternary. *Review of Palaeobotany and Palynology* 162: 416–426.
99. Hewitt G (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica* 139: 617–638.
100. Martín-Bravo S, Valcárcel V, Vargas P, Luceño M (2010) Geographical speciation related to Pleistocene range shifts in the western mediterranean mountains (*Reseda* sect. *Glaucoreseda*, Resedaceae). *Taxon* 59: 466–482.
101. Vargas P, Carrió E, Guzmán B, Amat E, Güemes J (2009) A geographical pattern of *Antirrhinum* (Scrophulariaceae) speciation since the Pliocene based on plastid and nuclear DNA polymorphism. *Journal of Biogeography*
102. Wilson Y, Hudson A (2011) The evolutionary history of *Antirrhinum* suggests that ancestral phenotype combinations survived repeated hybridizations. *Plant Journal* 66: 1032–1043.
103. Abbott RJ, Hegarty MJ, Hiscock SJ, Brennan AC (2010) Homoploid hybrid speciation in action. *Taxon* 59: 1375–1386.
104. Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH (2000) The likelihood of homoploid hybrid speciation. *Heredity* 84: 441–451.
105. Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68: 87–112.
106. Hewitt G (2000) The genetic legacy of the quaternary ice ages. *Nature* 405: 907–913.
107. Gossmann TI, Song B-H, Windsor AJ, Mitchell-Olds T, Dixon CJ, et al. (2010) Genome Wide Analyses Reveal Little Evidence for Adaptive Evolution in Many Plant Species. *Molecular Biology and Evolution* 27: 1822–1832.
108. Krijgsman W, Hilgen FJ, Raffi I, Sierro FJ, Wilson DS (1999) Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400: 652–655.
109. Garcia-Castellanos D, Estrada F, Jiménez-Munt I, Gorini C, Fernández M, et al. (2009) Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature* 462: 778–781.
110. Barres L, Vilatersana R, Molero J, Susanna A, Galbany-Casals M (2011) Molecular phylogeny of *Euphorbia* subg. *Esula* sect. *Aphyllis* (Euphorbiaceae) inferred from nrDNA and cpDNA markers with biogeographic insights. *Taxon* 60: 705–720.
111. Presti RML, Oppolzer S, Oberprieler C (2010) A molecular phylogeny and a revised classification of the mediterranean genus *Anthemis* s.l. (Compositae, Anthemideae) based on three molecular markers and micromorphological characters. *Taxon* 59: 1441–1456.
112. Mansion G, Zeltner L, Bretagnolle F (2005) Phylogenetic patterns and polyploid evolution within the Mediterranean genus *Centaureum* (Gentianaceae - Chironieae). *Taxon* 54: 931–950.
113. Vilatersana R, Garcia-Jacas N, Garnatje T, Molero J, Sonnante G, et al. (2010) Molecular phylogeny of the genus *Ptilostemon* (Compositae: Cardueae) and its relationships with *Cynara* and *Lamyropsis*. *Systematic Botany* 35: 907–917.
114. Jakob SS, Blattner FR (2006) A chloroplast genealogy of *Hordeum* (Poaceae): Long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Molecular Biology and Evolution* 23: 1602–1612.
115. Meerow AW, Francisco-Ortega J, Kuhn DN, Schnell RJ (2006) Phylogenetic relationships and biogeography within the Eurasian clade of Amaryllidaceae based on plastid ndhF and nrDNA ITS sequences: Lineage sorting in a reticulate area? *Systematic Botany* 31: 42–60.
116. Guo YP, Ehrendorfer F, Samuel R (2004) Phylogeny and systematics of *Achillea* (Asteraceae-Anthemideae) inferred from nrITS and plastid trnL-F DNA sequences. *Taxon* 53: 657–672.
117. Comes HP, Abbott RJ (2001) Molecular phylogeography, reticulation, and lineage sorting in mediterranean *Senecio* sect. *Senecio* (Asteraceae). *Evolution* 55: 1943–1962.
118. Valcárcel V, Vargas P, Feliner GN (2006) Phylogenetic and phylogeographic analysis of the western Mediterranean *Arenaria* section *Plinthine* (Caryophyllaceae) based on nuclear, plastid, and morphological markers. *Taxon* 55: 297–312.
119. Albaladejo RG, Aguilar JF, Aparicio A, Feliner GN (2005) Contrasting nuclear-plastidial phylogenetic patterns in the recently diverged Iberian *Phlomis crinita* and *P. lychnitis* lineages (Lamiaceae). *Taxon* 54: 987–998.