Chapter 32

Using ChromEvol to Determine the Mode of Chromosomal Evolution

Marcial Escudero, Enrique Maguilla, José Ignacio Márquez-Corro, Santiago Martín-Bravo, Itay Mayrose, Anat Shafir, Lu Tan, Carrie Tribble, and Rosana Zenil-Ferguson

Abstract

The ChromEvol software was the first to implement a likelihood-based approach, using probabilistic models that depict the pattern of chromosome number change along a specified phylogeny. The initial models have been completed and expanded during the last years. New parameters that model polyploid chromosome evolution have been implemented in ChromEvol v.2. In recent years, new and more complex models have been developed. The BiChrom model is able to implement two distinct chromosome models for the two possible trait states of a binary character of interest. ChromoSSE jointly implements chromosome evolution, speciation, and extinction. In the near future, we will be able to study chromosome evolution with increasingly complex models.

Key words BiChrom, Dysploidy, ChromEvol, ChromoSSE, Chromosome count, ChromoPlus, Phylogeny, Polyploidy, Probabilistic model

1 Introduction

1.1 Background: The Importance of Chromosomal Evolution

Chromosome rearrangements foster speciation by intensifying reproductive isolation during the speciation process; thus, chromosome evolution is a major driver of diversification in eukaryotes [1]. Among the main clades of eukaryotes, chromosome numbers are extraordinarily diverse in plants, especially in ferns ($n = 4$ to $n = 720$, [2]) and angiosperms ($n = 2$ to $n = 320$, [3]). This suggests the importance of chromosome number evolution in plants. There are two main kinds of chromosome number transitions: the multiplication of a whole chromosome set which entails a whole-genome duplication, WGD [4], and gains and losses of single chromosomes which encompass different phenomena with various expected outcomes: (i) aneuploidy: duplication or loss of a chromosome including its DNA content (this has been thought to

Tony Heitkam and Sònia Garcia (eds.), *Plant Cytogenetics and Cytogenomics: Methods and Protocols*, Methods in Molecular Biology, vol. 2672, https://doi.org/10.1007/978-1-0716-3226-0_32,

have little evolutionary success), and (ii) dysploidy: chromosome fusion/fission that do not result in changes in DNA content [5]. The topic of chromosome number evolution and ancestral chromosome number in angiosperms has received much attention since the second half of the twentieth century to the present [6–13].

1.2 Before ChromEvol Modeling: The Limitations of Parsimony

Starting from the last quarter of the twentieth century and continuing today, the sequencing revolution has dramatically increased the availability of molecular phylogenies throughout the tree of life. In light of these inferred phylogenetic relationships, biologists attempted to infer the evolution of chromosome numbers in their clades of interest. Initially, researchers were making these inferences based on the parsimony principle—either by eyeballing [14–16] or via the use of computer software that implemented established maximum parsimony algorithms for general character state evolution [17–23]. Although maximum parsimony has been widely used to infer chromosome number evolution, there are several problems with using this approach [24]. (i) In parsimony, all types of transitions have the same weight, which is not realistic for chromosome number transitions (for example, chromosome number increase by polyploidy may have a different probability than an increase by a single chromosome number via ascending dysploidy). Although it is possible to set up different weights for each kind of transition in parsimony, a specific a priori weighting would be always subjective and unjustified. (ii) The maximum parsimony approach does not take into account the possibility of multiple or back transitions along a single branch, thus tending to underestimate the number of transitions. (iii) Maximum parsimony ignores any uncertainty in the assignments of ancestral states (reconstructed character states on internal nodes are treated in exactly the same way as observed chromosome numbers in extant species). (iv) There are often a number of equally parsimonious reconstructions which may be incongruent. (v) There is no way to objectively determine the accuracy of the reconstruction. Finally, (vi) maximum parsimony analysis ignores any information contained with the branch lengths of the tree, assuming that transitions are equally likely to occur along very long or very short branches.

1.3 A New Probabilistic Model to Infer Chromosome Number Evolution: ChromEvol v.1.0

The ChromEvol software v.1.0 [24] was the first to implement a likelihood-based approach, using a probabilistic model that depicts the pattern of chromosome number change along a specified phylogeny. The basic ChromEvol model accounts for three possible events: an increase by a single chromosome number (ascending dysploidy), a decrease by a single chromosome number (descending dysploidy), and duplications of the chromosome number (i.e., whole-genome duplication, or polyploidy). The model is

$$Q_{ij}=\begin{cases}\lambda & j=i+1,\\ \delta & j=i-1,\\ \rho & j=2i,\\ 0 & \text{otherwise,}\end{cases}$$

Fig. 1 ChromEvol model. The model is formulated as a continuous time Markov process, described by the instantaneous matrix *Q*, that specifies the transition rates from a genome with *i* haploid chromosomes to a genome with *j* haploid chromosomes. λ, δ, and ρ define the rates of ascending dysploidy, descending dysploidy, and genome duplications, respectively. The two additional rate parameters λ_l and δ_l allow the ascending and descending dysploidy rates to depend linearly on the current number of chromosomes (accounting for the possibility that a species with very small chromosome number has a lower probability for descending dysploidy than a species with a high chromosome number)

formulated as a continuous time Markov process, described by the instantaneous matrix Q, that specifies the transition rates from a genome with i haploid chromosomes to a genome with j haploid chromosomes, as shown in Fig. 1. λ, δ, and ρ define the rates of ascending dysploidy, descending dysploidy, and genome duplications, respectively. The two additional rate parameters λ_l and δ_l allow the ascending and descending dysploidy rates to depend linearly on the current number of chromosomes (accounting for the possibility that a species with very small chromosome number has a lower probability for descending dysploidy than a species with a high chromosome number). This basic model allows for polyploidy events that are exact duplications of the chromosome number (example: $n = 10 \rightarrow n = 20 \rightarrow n = 40$). However, a more complex model in ChromEvol v.1.0 allows for an additional parameter, defined as demi-duplication or demi-polyploidy (μ), which permits multiplications of the number of chromosomes by 1.5 (example: $n = 10 \rightarrow n = 15$). The use of this parameter allows modeling events such as the generation of hexaploid from a tetraploid lineage via the fusion of reduced and unreduced gametes. Notably, demi-polyploidization has been commonly inferred in lineages with frequent allopolyploidy in which the two parental species usually have different basic numbers. In ChromEvol v.1.0, it is possible to define models of two parameters (descending and ascending dysploidy) to six parameters (polyploidy, demi-polyploidy, and setting the dysploidy rates to linearly depend on the current chromosome number) representing chromosome transitions (Table 1). The best model is chosen by maximizing the likelihood of each examined model to the data and using a model selection criterion, such as the Akaike information criterion (AIC). Based on the best fitting model, ChromEvol allows us to infer the ancestral chromosome numbers using two different likelihood

Table 1
Models for chromosome number evolution implemented in ChromEvol v.1.0 (D_{ys}, $D_{ys}D_{up}$, $D_{ys}D_{up}D_{em}^{*}$, $D_{ys}D_{up}D_{em}$, D_{ys}^{Linear}, $D_{ys}^{Linear}D_{up}$, $D_{ys}^{Linear}D_{up}D_{em}^{*}$, $D_{ys}^{Linear}D_{up}D_{em}$) and new models in ChromEvol v.2.0. ($D_{ys}B_{num}$, $D_{ys}D_{up}B_{num}$)

Model	Number of free *p*arameters	Parameters	Possible transitions
D_{ys}	2	λ, δ	Gains and losses of single chromosomes (dysploidy)
$D_{ys}D_{up}$	3	λ, δ, ρ	Gains and losses of single chromosomes (dysploidy) and duplications (polyploidy)
$D_{ys}D_{up}D_{em}^{*}$	3	λ, δ, $\rho = \mu$	Gains and losses of single chromosomes (dysploidy) and duplications and demi-duplications (same rates $\rho = \mu$; polyploidy)
$D_{ys}D_{up}D_{em}$	4	λ, δ, ρ, μ	Gains and losses of single chromosomes (dysploidy) and duplications and demi-duplications (polyploidy)
D_{ys}^{Linear}	4	λ, δ, λ_1, δ_1	Gains and losses of single chromosomes (dysploidy) that may depend linearly on the current chromosome number
$D_{ys}^{Linear}D_{up}$	5	λ, δ, λ_1, δ_1, ρ	Gains and losses of single chromosomes (dysploidy) that may depend linearly on the current chromosome number and duplication (polyploidy)
$D_{ys}^{Linear}D_{up}D_{em}^{*}$	5	λ, δ, λ_1, δ_1, $\rho = \mu$	Gains and losses of single chromosomes (dysploidy) that may depend linearly on the current chromosome number and duplication and demi-duplication (same rates $\rho = \mu$; polyploidy)
$D_{ys}^{Linear}D_{up}D_{em}$	6	λ, δ, λ_1, δ_1, ρ, μ	Gains and losses of single chromosomes (dysploidy) that may depend linearly on the current chromosome number and duplication and demi-duplication (polyploidy)
$D_{ys}B_{num}$	3–4	λ, δ, ν, β	Gains and losses of single chromosomes (dysploidy) and multiplication of base chromosome number (polyploidy). Base number (β) can be estimated or fixed
$D_{ys}D_{up}B_{num}$	4–5	λ, δ, ρ, ν, β	Gains and losses of single chromosomes (dysploidy) and duplication and multiplication of base chromosome number (polyploidy). Base number (β) can be estimated or fixed

approaches: the marginal ancestral reconstruction [25] and the joint ancestral reconstruction [26]. The expected number of events for each type of transition (ascending dysploidy, descending dysploidy, polyploidy, and demi-polyploidy) along each branch is calculated over a large number of simulated evolutionary paths that are generated based on the best fitting model, and weighted

according to the joint probability of the two terminal nodes of each branch [24]. ChromEvol v.1.0 models have been implemented in the RevBayes platform [27] that allows for estimation of posterior distributions for these same parameters and ancestral state reconstruction based on marginal posterior probabilities at the nodes of the tree.

Since the publication of ChromEvol, evolutionary biologists have inferred patterns of chromosome evolution in different lineages (Mayrose et al.'s publication [24] has been cited ca. 200 times up to date January 24, 2022). One important study of diversification linked to polyploidy [28] used ChromEvol as a first step to (i) classify taxa in tips of a phylogeny as diploid and polyploid, effectively overcoming one of the major challenges in the study of polyploidy, the classification of ploidy using chromosome number and (ii) test whether or not those polyploid event happens before cladogenesis (internal branches) or after cladogenesis (external branches; see comments in [29, 30]). Other important examples of chromosomal evolution have been done in fungi [31] or fish [32]. These models have helped also to understand the evolutionary importance of dysploidy in angiosperms where, traditionally, polyploidy has received much more attention [5].

1.4 New Parameters to Model Polyploid Series from a Base Number: ChromEvol v.2.0

A new version of CromEvol (v.2.0) was published in 2014 [33]. This new version included, among other new features (*see* [33] for details), a new type of polyploid transition that accounts for general multiplication events of a base number (or monoploid number). Thus, for example, assuming a base number of $n = 7$, additions of 7, 14, or 21 chromosomes are allowed in a single step. This alleviates unrealistic scenarios where, for example, transitions from $n = 21$ to $n = 28$ are forced to include many dysploidy events. Base-number transitions are incorporated into the ChromEvol model using two additional parameters: β, the monoploid (base) number and ν, its respective transition rate (Table 1). The β parameter of a group of interest can be either inferred in the statistical framework via maximum likelihood or be fixed to a specified value by the user. Notably, base-number transitions can be included in the probabilistic model with or without other parameters that allow for simple multiplication events (i.e., by 2 or 1.5), thus leading to a rather complex model that better deals with clades that include polyploid series and many allopolyploidizations. Recently, chromosome evolution has been studied in the context of a mega phylogeny (10,000 tips) of angiosperms using the ChromEvol v.2.0 models along with an analysis of genome size evolution, which have shed more light on the evolutionary history of chromosome number in the whole clade of angiosperms [12].

1.5 Complex Models: Multiple Chromosome Number Evolution Models in the Phylogeny

The models described above assumed that the evolution of chromosome numbers is time-homogeneous, such that there are no changes in the mode and tempo of chromosome evolution throughout the entire clade under study. However, the potential influence of other traits linked to chromosome number change might modify rates of chromosome number change across time. Therefore, aside from inaccurate description of the evolutionary process, the simplification made by time-homogeneous models could result in inaccurate parameter estimations and erroneous inferences. Two modeling extensions allow for the possibility that rates of chromosome number change are associated with a binary organismal trait. The BiChrom model, incorporated within the Chromploid R package [34, 35], contains ten free parameters (Table 2): genome duplications, descending dysploidy and ascending dysploidy for each of the two binary states, transition rates between the two states, and two ancillary parameters to model chromosome transition for high and also rare haploid chromosome numbers in the dataset under study. Using this model, Zenil-Ferguson et al. [34] demonstrated that, in eudicots, rates of genome duplications are six times higher in herbaceous than in woody lineages. The model of Blackmon et al. [36], implemented in the chromePlus R package, similarly allows a binary trait to affect rates of chromosome number change and additionally allows for the binary trait to affect rates of speciation and extinction, following the BiSSE modeling framework [37].

There have been several attempts to test for a shift in the pattern of chromosome number change in different clades of a phylogeny. For example, Márquez-Corro et al. [38] used the

Table 2
Models and parameter of chromosome evolution in BiChrom and reduced BiChrom models as implemented in Chromploid R package

Model	Number of free *p*arameters	Parameters	Possible transitions
BiChrom	10	ρ_0, ρ_1, λ_0, λ_1, μ_0, μ_1, q_{01}, q_{10}, ε_0, ε_1	Duplication and gains and losses of single chromosomes for the trait state 0 and 1. Transitions from 0 to 1 and from 1 to 0. Two ancillary parameters for chromosome transitions for high and rare chromosome number in the data set
Reduced BiChrom	9	ρ, λ_0, λ_1, μ_0, μ_1, q_{01}, q_{10}, ε_0, ε_1	Duplication, the same for both trait states. Gains and losses of single chromosomes for the trait state 0 and 1. Transitions from 0 to 1 and from 1 to 0. Two ancillary parameters for chromosome transitions for high and rare chromosome number in the data set

"censored approach" for continuous traits from O'Meara et al. [39] to test for different rates of chromosome evolution in the phylogeny of the sedges (Cyperaceae; in the original publication O'Meara et al. compared a single Brownian Motion model for several clades, subtrees, of a given phylogeny versus several Brownian Motion models, one for each subtree of the phylogeny. The branches connecting the subtrees of the phylogeny were excluded from the analyses). The sedges phylogenetic tree was divided into five subtrees based on four different shifts in diversification rates that were previously detected [40, 41]. Márquez-Corro et al. [38] compared the hypothesis that each of the five subtrees is characterized by a different ChromEvol regime to the null hypothesis that rates of chromosome number evolution are similar across the phylogeny (i.e., a single model to the entire tree). The authors found that the best model was the one using four different regimes (two of the five subtrees shared the same chromosome model whereas the three remaining subtrees were each characterized by its own regime). In another study (this time using the "uncensored approach" described in [39]), Aparicio et al. [42] also found that the chromosome evolution of genus *Helianthemum* was better described using two different regimes of chromosome evolution versus using a single fixed rates model for the whole tree. Notably, in the above studies the phylogeny was partitioned to subclades prior to analysis, based on reasoning derived from external evidence. However, there is not yet an implementation that would allow users to automatically search for shifts in the pattern of chromosome number change, similar to models of continuous trait evolution, such as in BAMM, Brownie, Auteur, or Bayou [43–47].

1.6 Complex Models: Chromosome State Speciation and Extinction Model (ChromoSSE)

All the previous models assume that the chromosome number transitions occurred anagenetically, along branches of the phylogeny, but that they are not linked to cladogenesis. This is an important assumption that implies that chromosome changes do not affect reproductive isolation and speciation processes. To alleviate this assumption, Freyman and Höhna [48] proposed the ChromoSSE model (implemented in RevBayes, [27]) which is based on a statistical Bayesian approach, and allows for both anagenetic and cladogenetic chromosome number transitions, effectively linking the process of diversification to chromosome number change. The ChromoSSE model is a birth and death stochastic model with up to 13 parameters (Table 3). In the original publication of the model, Freyman and Höhna [48] estimated evidence of cladogenetic chromosome transitions in several lineages of angiosperms like *Aristolochia* (Aristolochiaceae), *Helianthus* (Asteraceae), *Carex* sect. *Spirostachyae* (Cyperaceae), *Primula* sect. *Aleuritia* (Primulaceae), and *Mimulus* (Phrymaceae). The ChromoSSE model is very promising but is computationally complex and requires long running times, prohibiting its applicability to large trees.

Table 3
ChromoSSE model parameters

	Parameter	Parameter symbol
Anagenetic	Chromosome gain rate	γ_a
	Chromosome loss rate	δ_a
	Polyploidization rate	ρ_a
	Demi-polyploidization rate	ρ_a
	Linear component of chromosome gain rate	γ_m
	Linear component of chromosome loss rate	δ_m
Cladogenetic	No change	ϕ_c
	Chromosome gain rate	γ_c
	Chromosome loss rate	δ_c
	Polyploidization rate	ρ_c
	Demi-polyploidization rate	ρ_c
Other	Root frequencies	π
	Relative extinction	r

2 Experimental Examples

2.1 Using ChromEvol v.1.0 and ChromEvol v.2.0 on Genus Centaurium (Gentianaceae)

The genus *Centaurium* Hill (Gentianaceae Juss.) has approximately 26 taxa (ca. 18 species and eight subspecies, [49, 50]) and its center of diversity in the Mediterranean basin, although the range of the genus includes Asia, Europe, North-Central Africa, and North America [49]. Polyploidy has been suggested to play an important role in the evolution of *Centaurium*, with 60% of the taxa being polyploids [51–53]. Ploidy levels in the genus seem to conform to a geographical pattern, with diploid species ($2n = 18$, 20) around the Mediterranean basin, tetraploids ($4x = 36$, 40) mainly in northern Europe and eastern Asia, and hexaploids ($6x = 54$, 56, 60) generally distributed in India, the Arabian Peninsula, and southwestern Mediterranean basin reaching the Canary Islands [52, 53]. The hexaploids (50% of polyploid taxa) have been proposed to be allopolyploids, while the tetraploids have been suggested to be autopolyploids [51, 52].

ChromEvol was used to study the evolution of chromosome numbers in the genus *Centaurium*. First, the eight models of chromosome number evolution implemented in ChromEvol v.1.0 were fitted to the data, which included a reconstructed phylogeny and chromosome numbers of 26 *Centaurium* taxa [53]. The best fitted model was $D_{ys}D_{up}D_{em}$ which contains fixed rates for

Table 4
Akaike's information criterion (AIC) values of all models of chromosome evolution tested in ChromEvol v.2.0 for *Centaurium*

Model	Log Likelihood	AIC
$D_{ys}D_{up}$	−82.6371	171.274
$D_{ys}D_{up}D_{em}{}^{*}$	−57.9445	121.889
$D_{ys}D_{up}D_{em}$	**−56.5147**	**121.029**
D_{ys}	−98.4156	200.831
$D_{ys}{}^{Linear}D_{up}$	−81.0634	172.127
$D_{ys}{}^{Linear}D_{up}D_{em}{}^{*}$	−57.5882	125.176
$D_{ys}{}^{Linear}D_{up}D_{em}$	−55.2665	122.533
$D_{ys}{}^{Linear}$	−91.7715	191.543
$D_{ys}B_{num}$	−56.2311	120.462
$D_{ys}D_{up}B_{num}$	**−53.4947**	**116.989**

Best fitting model is shown in bold letters. This information was extracted from Maguilla et al. 2021

chromosome doubling, demi-ploidy, and dysploidy. Two other models ($D_{ys}D_{up}D_{em}{}^{*}$ with same rates for chromosome doubling and demi-ploidy and $D_{ys}{}^{Linear}D_{up}D_{em}$ which assumes linear rates of dysploidy) were within two AIC units (Table 4), meaning that they are not significantly inferior to the best model. Using the best model, two ascending dysploidy events were inferred, zero descending dysploidy, 12 genome duplications, and four demipolyploidizations (Fig. 2). Most of the inferred transitions were polyploidization or demipolyploidizations occurring along external branches of the phylogeny (Fig. 2), although up to four polyploidizations were inferred on internal branches (including deep ones). Under this model, $n = 5$ was inferred as the most likely ancestral chromosome number at the root of the phylogeny, with posterior probability of 0.6 (the second most likely root state was $n = 9$, Fig. 2). Notably, this inference ($n = 5$) is questionable since it was not observed in any *Centaurium* species or in related genera and has not been suggested before for the genus *Centaurium* or for tribe Chironieae [52].

We then fitted two additional models—implemented in ChromEvol v.2.0 that allow for base-number transitions—to the same data ([53], Table 4). Both models were superior to all eight previous models. The best model was $D_{ys}D_{up}B_{num}$ that contains the possibility for base-number transitions (inferred as $\beta = 10$) along with dysploidy and genome duplications, and it obtained strong support compared to all other models (delta AIC larger than 4; Table 4). Using this model, the inferred transitions and ancestral

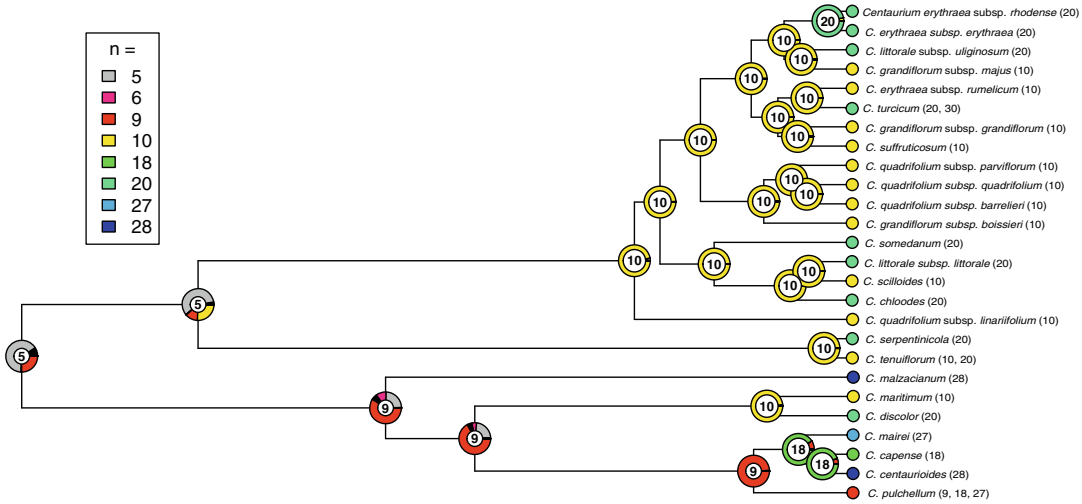

Fig. 2 Node charts represent the posterior probability of the inferred haploid chromosome number retrieved by ChromEvol v.2.0 [33] but only using the models implemented in ChromEvol v.1.0 [24]. Tip labels indicate species names and haploid chromosome number (*n*, in parenthesis and colored tip circles). (This information was extracted from Maguilla et al. [53])

states were markedly different from those inferred using $D_{ys}D_{up}D_{em}$. Specifically, four descending dysploidy events and no ascending dysploidy events were inferred (zero and two using $D_{ys}D_{up}D_{em}$, respectively) and 13 polyploidy events (either duplications or base-number transitions), compared to 16 events using $D_{ys}D_{up}D_{em}$ (Fig. 3). Under this model, $n = 10$ was inferred as the most likely ancestral chromosome number at the root of the phylogeny. A transition from $n = 10$ to $n = 9$ was inferred in an internal branch of the Widespread clade (Fig. 3). The two polyploid (duplication or base transition) events inferred in internal branches were located one in the Western clade and other in the Widespread clade (Fig. 3). Most of the remaining events were polyploidizations occurring along external branches of the phylogeny (Fig. 3). The three remaining descending dysploidy were inferred in hexaploid lineages of the Widespread clade (two in *C. malzacianum* $n = 28$ and one in *C. mairei* $n = 27$) to deal with the fact that these two species probably originated as the result of allopolyploidization between *Centaurium* species with $x = 10$ and $x = 9$ and not by polyploidization involving a single base chromosome number (10 or 9). Taken together, it seems that incorporating the possibility of base-number transitions reconstructs a more realistic scenario than that inferred using ChromEvol v.1.0 since $n = 10$ is the most common chromosome number in the genus *Centaurium* and is also very frequent in species of closely related genera. It is conceivable that a more complex model, which allows for multiple base numbers (here, 10 and 9) could provide even more realistic

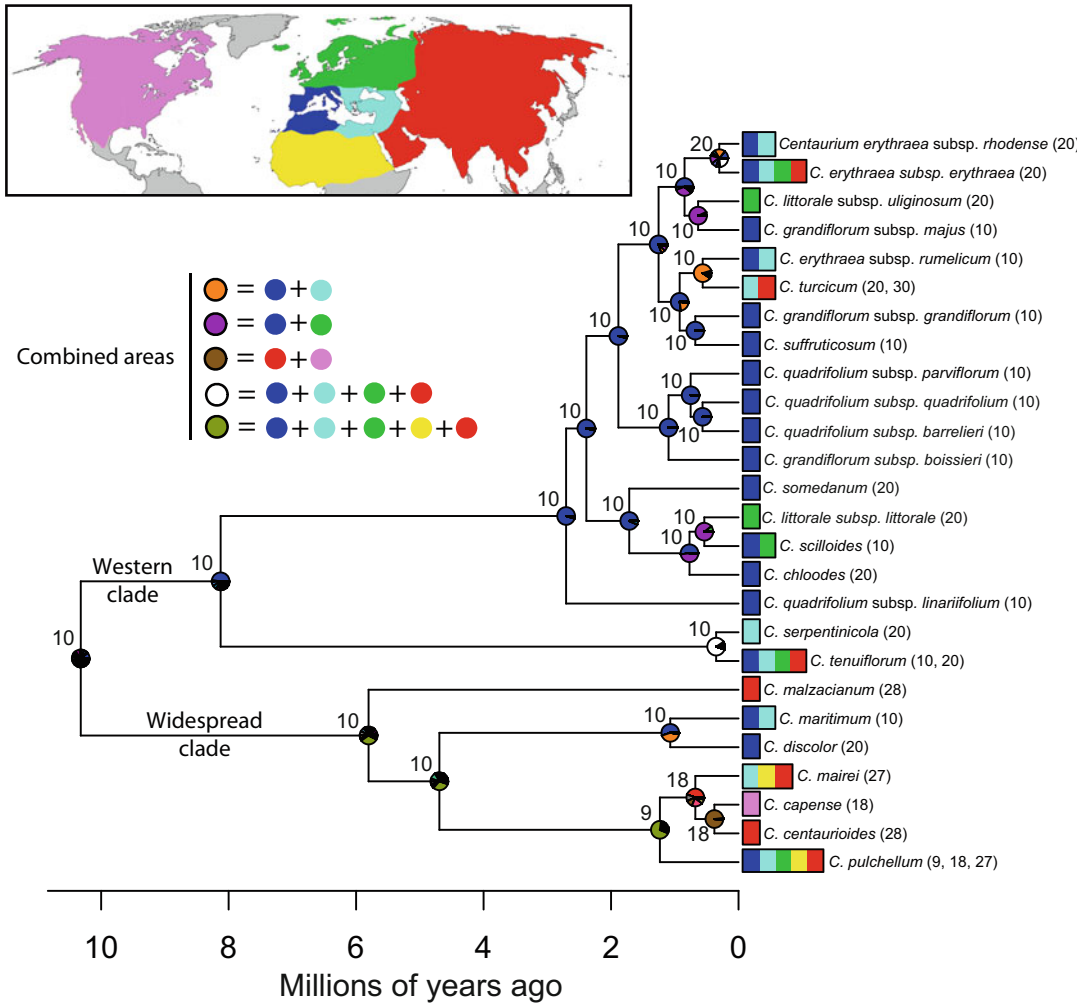

Fig. 3 Ancestral area reconstruction under constrained DEC model in *Centaurium* resulting from BioGeoBEARS analysis [62]. Node charts represent the probability of the occurrence of the most recent common ancestor (MRCA) of that node in the area represented by the same color on the map, and combined areas are represented by colors in the legend. Node numbers indicate the reconstructed haploid chromosome numbers retrieved by ChromEvol v.2.0 [33] using $D_{ys}D_{up}B_{num}$ model. Tip labels indicate species names and the area where each species occurs following colors in the map. Haploid chromosome number (*n*, in parenthesis) and ploidy levels for the species are also indicated. Lower time scale shows Millions of years ago, from the origin of the genus to the present (Figure from Maguilla et al. [53]. Reproduced with permit from Frontiers editorial)

reconstructions, allowing for allopolyploidy events involving lineages of two different base numbers.

2.2 Experimental Example on Genus Centaurium (Gentianaceae) Using Bichrom Model

The ploidy level in *Centaurium* is geographically distributed with diploids in the Mediterranean Basin (the area of origin), tetraploids in temperate areas toward the North, and hexaploids in more arid areas toward the South ([52], *see* Fig. 2 here). Maguilla et al. [53] tested for association between ploidy levels and geographic

Table 5
Rate parameters of chromosome number evolution inferred by the BiChrom and the reduced BiChrom models as implemented in the Chromploid R package fitted to *Centaurium*

Model	λ_0/λ_1	μ_0/μ_1	ρ_0/ρ_1	q_{01}/q_{10}	Log Lik	AIC	AICc
BiChrom	0.398/0.000	0.000/0.126	0.001/0.631	3.668/3.925	−88.2626	196.52	211.19
Reduced BiChrom	0.435/0.000	0.000/0.095	0.314	0.827/0.950	−88.7244	195.45	206.70

The subscripts 0 and 1 correspond to the two binary states examined (area of origin and colonized areas, respectively). λ = ascending dysploidy, μ = descending dysploidy, ρ = duplications, and q = transition rate between the two binary states

distribution using the correlated evolution model of Pagel [54]. This analysis indicated that the rate of transition from the area of origin to colonized areas (North and South regions) is independent of ploidy level but the rate of transition from diploid to polyploid was very low in the area of origin and 100 times faster in colonized areas. This analysis thus suggested that the probability of colonizing new areas is independent of ploidy level but after colonization the rate of polyploidization rapidly increases. Based on this evidence, we have fitted the BiChrom model (with two polyploidization rates, one for the area of origin and one for the colonized areas) and compared it to a reduced BiChrom model with a single polyploidization rate, assuming it is equal in both areas (Table 5). The fit of the reduced model was better than that of the BiChrom model (ΔAICc = 4.49) and is significantly higher than the AICc of the reduced BiChrom model, indicating that there is no significant evidence of two rates of chromosomal evolution. Still, the polyploidy inferred by the two models is markedly different, with the BiChrom model inferring a polyploidization rate ca. 600 times higher in colonized areas than in the area of origin, which is congruent with previous results reported by Maguilla et al. [53]. Thus, the non-significant support of the BiChrom model most probably stems from the small dataset size that is insufficient to support such a complex model.

2.3 Experimental Example on the Genus Centaurium (Gentianaceae) Using ChromoSSE Model

Polyploidy has been suggested to be a major driver of diversification in genus *Centaurium* [52, 53] but, to date, this has not been statistically tested. To test such hypotheses, we have run a simplified version of the ChromoSSE model in which only three parameters of chromosome number transitions are allowed for both anagenetic and cladogenetic changes (descending and ascending dysploidy and duplications, in addition to parameters that are not related to chromosome transitions: relative extinction, root frequencies, and no change parameters). The ChromoSSE model used here is similar to the $D_{ys}D_{up}$ ChromEvol model (Table 1) but also allows

chromosome-number transitions to be associated with cladogenetic events. As discussed in Experimental Example 1, we obtained unlikely reconstructions using a more complex model, $D_{ys}D_{up}D_{em}$, in comparison with those obtained using the $D_{ys}D_{up}B_{num}$ model. Therefore, it is expected that the simplified model we implemented in ChromoSSE will result in unrealistic reconstructions in the deep nodes of the phylogeny. Nevertheless, because of the already high number of parameters under consideration, our simplification of ChromoSSE helps us to illustrate its applicability to address questions of diversification linked to chromosome number changes despite the outcome of the reconstruction.

Based on the estimated model parameters using the reduced ChromoSSE model, the most frequent type of anagenetic transition in *Centaurium* is ascending dysploidy, with rates of duplications and descending dysploidy being substantially lower (Figs. 4 and 5). However, a markedly different pattern was obtained for cladogenesis transitions, where both rates of dysploidy transitions are near 0 while rate of genome duplications being much higher. This analysis thus indicates that polyploidy is linked to lineage splitting events in the genus *Centaurium*.

3 Future Perspectives

We have seen how the development of ChromEvol models and their more complex extensions have allowed us to investigate the mode and tempo of chromosomal evolution in many clades of eukaryotes using a robust quantitative approach. As with all evolutionary models, the models of chromosome number change we have discussed are simplifications of the underlying evolutionary process and there are multiple ways in which such models can be improved. First, we have discussed how the addition of a new parameter that allows polyploidy to be modeled as a multiplication of a base chromosome number can be critical at reconstructing polyploid series. However, this key parameter has not been taken into account in more complex models like BiChrom or ChromoSSE, perhaps due to the high dimensionality of these models, and its incorporation into future implementations could improve predictions. Furthermore, a potential model that allows for multiple base chromosome numbers, rather than one, may also improve inferences. Second, we still miss an implementation of a "split model," in which each subtree would be characterized by its own set of parameters. Such implementation would further allow an automatic search of shifting points along the phylogeny, although progress is being made on this critical issue. Third, it is well known that SSE models, such as BiSSE [37] and others, typically suffer elevated type I error rates [55]. To partly overcome this latter issue,

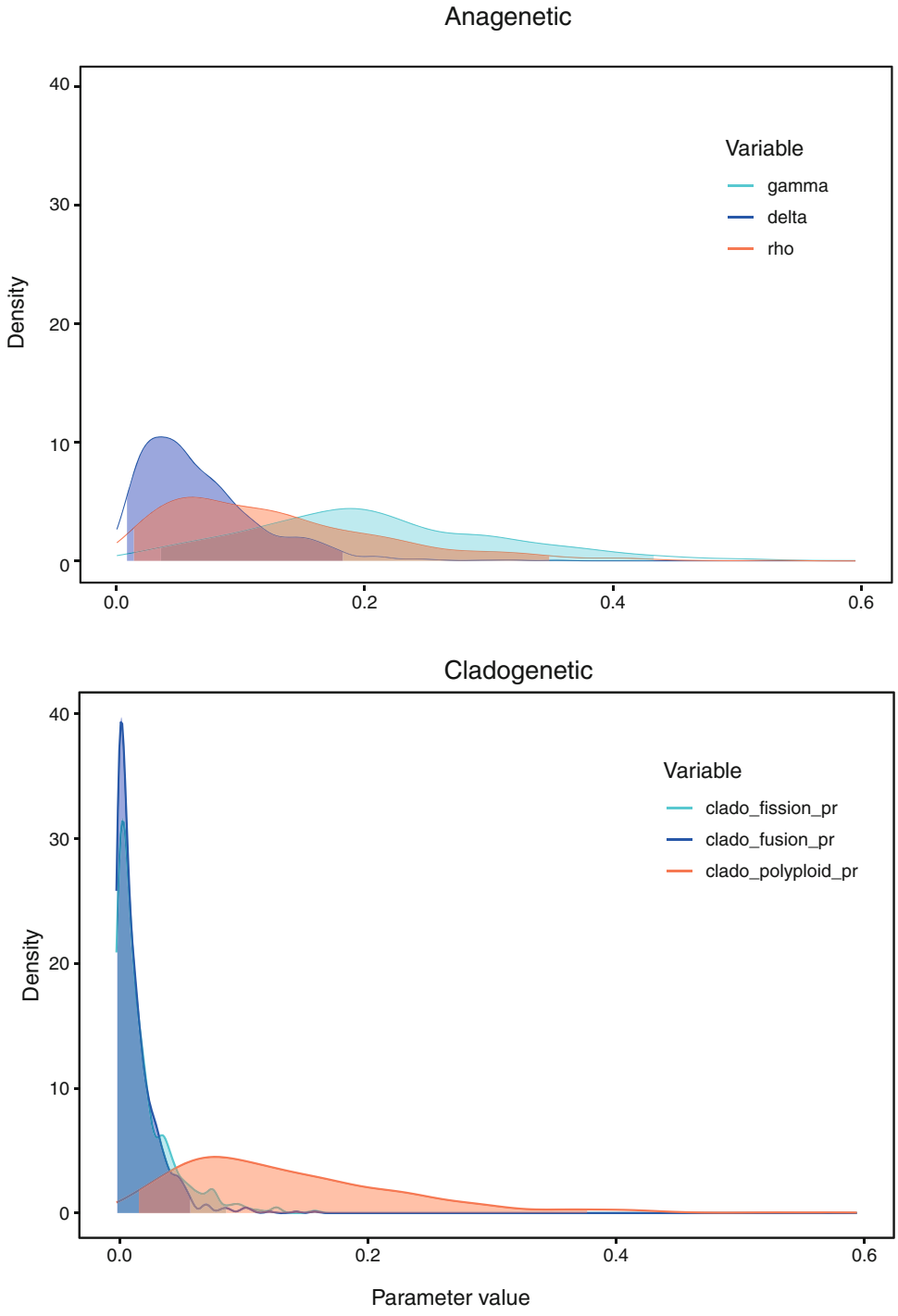

Fig. 4 Posterior distribution of estimated rate parameters using the ChromoSSE model [48] implemented in RevBayes [27]. The visualization and plotting of the results were made using RevGadgets [63]

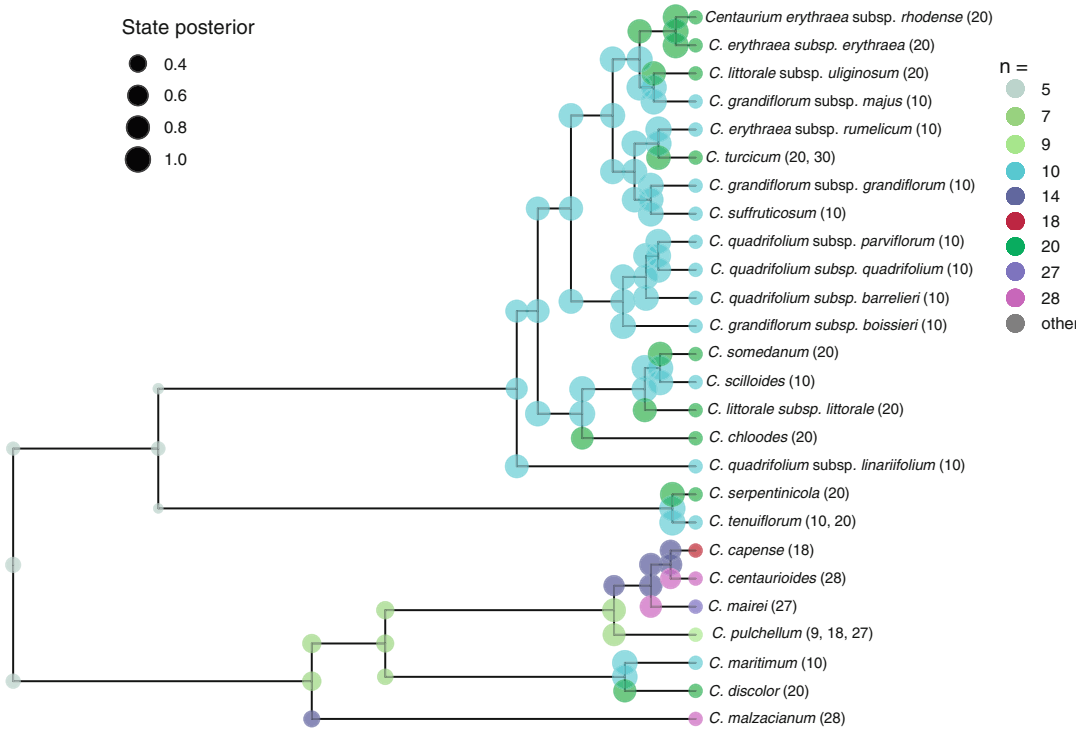

Fig. 5 Chromosome number and transition reconstructions using the ChromoSSE model [48] implemented in RevBayes [27]. The posterior probability of the chromosome number reconstruction is shown with the size of the circles (*see* legend State posterior). The reconstructed haploid chromosome numbers follow the color coding, shown at the right side of the phylogeny. The visualization and plotting of the results were made using RevGadgets [63]

the HiSSE model [56, 57] was introduced to incorporate unmeasured (or hidden) factors that could impact diversification rates aside from the trait of interest. The ChromoSSE model is expected to be as vulnerable as BiSSE to high false-positive rates. Hence, an implementation of hidden states in a ChromoSSE framework is critical to thoroughly test the contribution of chromosome number change to the diversification process in a clade.

Another important point is the time span in which models of chromosome numbers could provide reliable inferences. In their analysis of the entire angiosperm clade, Carta et al. [12] failed to infer any chromosome number duplications (whole-genome duplication) deep in the phylogeny, despite over 100 deep polyploid events inferred through comparative genomics data [58]. The difficulty of inferring these polyploid transitions based only on chromosome number information is rooted in the wondrous cycles of polyploidy in plants [13, 59] in which a polyploidization event is followed by a diploidization process (genome downsizing and chromosome fusion) that keeps genome size and chromosome number relatively stable throughout the evolution of angiosperms.

The use of informed models in which suspected polyploidization events are forced to occur along pre-specified branches of the phylogeny could aid with inferences. Finally, a more realistic model of chromosome number evolution can be achieved by adding to the ChromEvol framework new functions that better describe the dependencies between rates of polyploidy and dysploidy and the current number of chromosomes. In addition to the linear model that is implemented in ChromEvol, few other functions were proposed: a birth-death-like model, where the dysploidy rates are described as a linear function without an intercept term [60], and an exponential model of dependency on the current number of chromosomes [48]. All these models, however, ignore the possible dependency of the polyploidy rate on the current number of chromosomes, though polyploidy is less likely to occur in genomes with large numbers of chromosomes [61]. Moreover, each of the currently implemented functions has its own constraints and limitations. For example, while the linear models might be too simple and are largely constrained to keep the instantaneous Q matrix legitimate, the exponential function is less constrained, but can explode rapidly to unrealistic high values. These certainly represent fertile ground for future and more complex implementations that would result in more realistic inferences regarding the pathways by which the evolution of chromosome number proceeds.

Acknowledgments

This work was supported by a grant to ME (MICINN, PGC2018-099608-B-I00). We thank the Andalusian Scientific Information Technology Center (CICA, Seville, Spain) for providing computational and research resources.

References

1. Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L (2005) Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet 21:673–682
2. Khandelwal S (1990) Chromosome evolution in the genus *Ophioglossum* L. Bot J Linn Soc 102:205–217
3. Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A et al (2015) The Chromosome Counts Database (CCDB)–a community resource of plant chromosome numbers. New Phytol 206:19–26
4. Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C et al (2009) Polyploidy and angiosperm diversification. Am J Bot 96:336–348
5. Escudero M, Martín-Bravo S, Mayrose I, Fernández-Mazuecos M, Fiz-Palacios O, Hipp AL et al (2014) Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. PLoS One 9:e85266
6. Raven PH, Kyhos DW (1965) New evidence concerning the original basic chromosome number of angiosperms. Evolution 19:244–248
7. Stebbins GL (1971) Chromosomal evolution in higher plants. Chromosomal evolution in higher plants. Edward Arnold, London, UK
8. Grant V (1981) Plant speciation. Columbia University Press, NY
9. Leitch IJ, Bennett MD (1997) Polyploidy in angiosperms. Trends Plant Sci 2:470–476

10. Guerra M (2008) Chromosome numbers in plant cytotaxonomy: concepts and implications. Cytogenet Genome Res 120:339–350
11. Dodsworth S, Chase MW, Leitch AR (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? Bot J Linn Soc 180:1–5
12. Carta A, Bedini G, Peruzzi L (2020) A deep dive into the ancestral chromosome number and genome size of flowering plants. New Phytol 228:1097–1106
13. Escudero M, Wendel JF (2020) The grand sweep of chromosomal evolution in angiosperms. New Phytol 228:805–808
14. Vargas P, McAllister HA, Morton C, Jury SL, Wilkinson MJ (1999) Polyploid speciation in *Hedera* (Araliaceae): phylogenetic and biogeographic insights based on chromosome counts and ITS sequences. Plant Syst Evol 219:165–179
15. Wang W, Lan H (2000) Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. Mol Biol Evol 17:1326–1333
16. Ohi-Toma T, Sugawara T, Murata H, Wanke S, Neinhuis C, Murata J (2006) Molecular phylogeny of *Aristolochia* sensu lato (Aristolochiaceae) based on sequences of rbcL, matK, and phyA genes, with special reference to differentiation of chromosome numbers. Syst Bot 31: 481–492
17. Bena G, Prosperi JM, Lejeune B, Olivieri I (1998) Evolution of annual species of the genus *Medicago*: a molecular phylogenetic approach. Mol Phylogenet Evol 9:552–559
18. Cerbah M, Souza-Chies T, Jubier MF, Lejeune B, Siljak-Yakovlev S (1998) Molecular phylogeny of the genus *Hypochaeris* using internal transcribed spacers of nuclear rDNA: inference for chromosomal evolution. Mol Biol Evol 15:345–354
19. Watanabe K, Yahara T, Denda T, Kosuge K (1999) Chromosomal evolution in the genus *Brachyscome* (Asteraceae, Astereae): statistical tests regarding correlation between changes in karyotype and habit using phylogenetic information. J Plant Res 112:145–161
20. Aoki S, Ito M (2000) Molecular phylogeny of *Nicotiana* (Solanaceae) based on the nucleotide sequence of the matK gene. Plant Biol 2: 316–324
21. Bakker FT, Culham A, Pankhurst CE, Gibby M (2000) Mitochondrial and chloroplast DNA-based phylogeny of *Pelargonium* (Geraniaceae). Am J Bot 87:727–734
22. Martel E, Poncet V, Lamy F, Siljak-Yakovlev S, Lejeune B, Sarr A (2004) Chromosome evolution of *Pennisetum* species (Poaceae): implications of ITS phylogeny. Plant Syst Evol 249:139–149
23. Hansen AK, Gilbert LE, Simpson BB, Downie SR, Cervi AC, Jansen RK (2006) Phylogenetic relationships and chromosome number evolution in *Passiflora*. Syst Bot 31:138–150
24. Mayrose I, Barker MS, Otto SP (2010) Probabilistic models of chromosome number evolution and the inference of polyploidy. Syst Biol 59:132–144
25. Koshi JM, Goldstein RA (1996) Probabilistic reconstruction of ancestral protein sequences. J Mol Evol 42:313–320
26. Pupko T, Pe I, Shamir R, Graur D (2000) A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol Biol Evol 17: 890–896
27. Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR et al (2016) RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst Biol 65:726–736
28. Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP (2011) Recently formed polyploid plants diversify at lower rates. Sciences 333:1257–1257
29. Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV et al (2014) Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). New Phytol 202:1105–1117
30. Mayrose I, Zhan SH, Rothfels CJ, Arrigo N, Barker MS, Rieseberg LH, Otto SP (2015) Methods for studying polyploid diversification and the dead end hypothesis: a reply to Soltis et al. (2014). New Phytol 206:27–35
31. Albertin W, Marullo P (2012) Polyploidy in fungi: evolution after whole-genome duplication. Proc R Soc B: Biol Sci 279:2497–2509
32. Zhan SH, Glick L, Tsigenopoulos CS, Otto SP, Mayrose I (2014) Comparative analysis reveals that polyploidy does not decelerate diversification in fish. J Evol Biol 27:391–403
33. Glick L, Mayrose I (2014) ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Mol Biol Evol 31:1914–1922
34. Zenil-Ferguson R, Ponciano JM, Burleigh JG (2017) Testing the association of phenotypes with polyploidy: an example using herbaceous and woody eudicots. Evolution 71:1138–1148
35. Zenil-Ferguson R, Burleigh JG, Ponciano JM (2018) Chromploid: an R package for chromosome number evolution across the plant tree of life. Appl Plant Sci 6:e1037

36. Blackmon H, Justison J, Mayrose I, Goldberg EE (2019) Meiotic drive shapes rates of karyotype evolution in mammals. Evolution 73:511–523
37. Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. Syst Biol 56:701–710
38. Márquez-Corro JI, Martín-Bravo S, Spalink D, Luceño M, Escudero M (2019) Inferring hypothesis-based transitions in clade-specific models of chromosome number evolution in sedges (Cyperaceae). Mol Phylogenet Evol 135:203–209
39. O'Meara BC, Ané C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933
40. Escudero M, Hipp AL (2013) Shifts in diversification rates and clade ages explain species richness in higher-level sedge taxa (Cyperaceae). Am J Bot 100:2403–2411
41. Spalink D, Drew BT, Pace MC, Zaborsky JG, Starr JR, Cameron KM et al (2016) Biogeography of the cosmopolitan sedges (Cyperaceae) and the area-richness correlation in plants. J Biogeogr 43:1893–1904
42. Aparicio A, Escudero M, Valdés-Florido A, Pachón M, Rubio E, Albaladejo RG et al (2019) Karyotype evolution in *Helianthemum* (Cistaceae): dysploidy, achiasmate meiosis and ecological specialization in *H. squamatum*, a true gypsophile. Bot J Linn Soc 191:484–501
43. Eastman JM, Alfaro ME, Joyce P, Hipp AL, Harmon LJ (2011) A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65: 3578–3589
44. Stack JC, Harmon LJ, O'Meara B (2011) RBrownie: an R package for testing hypotheses about rates of evolutionary change. Methods Ecol Evol 2:660–662
45. Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, Alfaro ME (2013) Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat Commun 4:1–8
46. Uyeda JC, Harmon LJ (2014) A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Syst Biol 63:902–918
47. Ingram T, Mahler DL (2013) SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods Ecol Evol 4:416–425
48. Freyman WA, Höhna S (2018) Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. Syst Biol 67:195–215
49. Mansion G, Struwe L (2004) Generic delimitation and phylogenetic relationships within the subtribe Chironiinae (Chironieae: Gentianaceae), with special reference to *Centaurium*: evidence from nrDNA and cpDNA sequences. Mol Phylogenet Evol 32:951–977
50. Díaz-Lifante Z (2012) *Centaurium* in Flora Iberica 11, 49–81. Romero C, Quintanar A (eds). Real Jardín Botánico-CSIC Press, Madrid
51. Zeltner L (1970) Recherches de biosystématique Sur les genres *Blackstonia* Huds. et *Centaurium* Hill (Gentianaceae). Bull Soc Neuchâteloise Sci Nat 93:1–164
52. Mansion G, Zeltner L, Bretagnolle F (2005) Phylogenetic patterns and polyploid evolution within the Mediterranean genus *Centaurium* (Gentianaceae - Chironieae). Taxon 54:931–950
53. Maguilla E, Escudero M, Jiménez-Lobato V, Díaz-Lifante Z, Andrés-Camacho C, Arroyo J (2021) Polyploidy expands the range of *Centaurium* (Gentianaceae). Front Plant Sci 12: 650551
54. Pagel M (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc R Soc London Ser B: Biol Sci 255:37–45
55. Rabosky DL, Goldberg EE (2015) Model inadequacy and mistaken inferences of trait-dependent speciation. Syst Biol 64:340–355
56. Beaulieu JM, O'Meara BC, Donoghue MJ (2013) Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Syst Biol 62:725–737
57. Beaulieu JM, O'Meara BC (2016) Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst Biol 65:583–601
58. Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS (2018) Impact of whole-genome duplication events on diversification rates in angiosperms. Am J Bot 105: 348–363
59. Wendel JF (2015) The wondrous cycles of polyploidy in plants. Am J Bot 102:1753–1756
60. Hallinan NM, Lindberg DR (2011) Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. Genome Biol Evol 3:1150–1163
61. Zenil-Ferguson R, Ponciano JM, Burleigh JG (2016) Evaluating the role of genome downsizing and size thresholds from genome size

distributions in angiosperms. Am J Bot 103: 1175–1186

62. Matzke NJ (2013) Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Front Biogeogr 5:242–248

63. Tribble CM, Freyman WA, Lim JY, Landis MJ, Barido-Sottani J, Kopperud BT et al (2021) RevGadgets: an R Package for visualizing Bayesian phylogenetic analyses from RevBayes. *bioRxiv*