Detecting shifts in the mode of chromosomal speciation across the cosmopolitan plant lineage *Carex*

Carrie M. Tribble[1,*,†], José Ignacio Márquez-Corro[2,3,*], Michael R. May[4], Andrew L. Hipp[5], Marcial Escudero[6**], and Rosana Zenil-Ferguson[7**]

[1]*School of Life Sciences, University of Hawai'i at Mānoa, Honolulu, HI, 96822, USA*

[2]*Royal Botanic Gardens, Kew. Richmond, Surrey TW9 3AE, United Kingdom*

[3]*Department of Molecular Biology and Biochemistry Engineering, Universidad Pablo de Olavide. Sevilla 41013, Spain*

[4]*Department of Evolution and Ecology, University of California Davis, Davis, CA USA*

[5]*Herbarium and Center for Tree Science, The Morton Arboretum, Lisle, IL 60532 USA*

[6]*Department of Plant Biology and Ecology, University of Sevilla, Sevilla, Spain*

[7]*Department of Biology, University of Kentucky, Lexington KY 40506 USA*

[*]*Co-first authors*

[**]*Co-senior authors*

[†]*Corresponding author: ctribble09@gmail.com*

Abstract Recent studies of angiosperm diversification have focused on the role of polyploidy as a driver of diversification. However, we know far less about the effects of single changes in chromosome number—dysploidy, which can mediate lineage diversification by affecting recombination rates, linkage, and/or reproductive isolation. Modeling the effects of dysploidy on diversification is mathematically and computationally challenging because many states and parameters are required to track changes in individual chromosomes, especially in lineages that have high variability in chromosome number. Additionally, we expect the processes of diversification and dysploidy to vary across clades, which requires modeling process variation to disentangle the effects of the observed trait (chromosome number) from the effects of unobserved traits on diversification. In this work, we propose a new state-dependent diversification model of chromosome evolution that includes numerous character states and explicitly models heterogeneity in the diversification process. Our model includes parameters that functionally link diversification rates to dysploidy rates and differentiate between anagenetic and cladogenetic changes. We apply this model to *Carex* (Cyperaceae), a model lineage for understanding dysploidy and diversification, leveraging chromosome number information and the most recent time-calibrated phylogenetic tree for over 700 species and subspecies. We recover distinct modes of chromosomal speciation across *Carex*. In one mode, dysploidy occurs very frequently and drives faster diversification rates. In the other mode, dysploidy is rare and diversification is driven by other factors, unmeasured in our analysis. This study is the first to demonstrate that dysploidy drives diversification in plants while considering unmeasured factors affecting

diversification.

Significance Statement Evolutionary genomic rearrangements via the gain and/or loss of chromosomes without losing DNA may cause new species to form, but how this process plays out over a millions-of-years timescale is unclear. We test if there is a detectable effect of chromosome gains/losses on the rate of species formation in a large, diverse group of plants. We demonstrate decisively for the first time that these rearrangements sometimes drive the formation of new species through two evolutionary modes. In one mode, rearrangements happen rapidly and sometimes cause new species to form. In the other mode, rearrangements happen less frequently and are not associated with new species formation; instead, other factors—unmeasured in our analysis—likely drive the formation of new species.

[keywords—Chromosome number; diversification process variation; diversification; dysploidy; macroevolution; RevBayes; TensorPhylo]

Introduction

Unveiling the primary drivers of diversification remains one of the most important goals in evolutionary biology (Sauquet and Magallón, 2018). Hundreds of studies have focused on estimating changes in plant diversification processes through time (*e.g.*, Magallón and Castillo, 2009), across clades (*e.g.*, Magallón et al., 2019), or in association with trait evolution (*e.g.*, Helmstetter et al., 2023). Chromosome number changes and rearrangements are particularly likely to influence lineage diversification. Most plant diversification studies that address chromosome evolution focus on polyploidy, rather than on localized chromosomal rearrangements that do not involve large DNA content changes (*i.e.*, dysploidy, gains or losses of single chromosomes; Escudero et al., 2014; Mandáková and Lysak, 2018). A recent review of trait-dependent diversification in angiosperms, for example, cites seven studies linking speciation and polyploidy (Helmstetter et al., 2023), only one which considered disploidy linked to diversification (Freyman and Höhna, 2018). While dysploidy has the potential to influence lineage diversification through effects on recombination or reproductive isolation, far fewer macroevolutionary studies have focused on speciation and extinction as a result of dysploidy.

In a recent review, Lucek et al. proposed three contrasting models of chromosomal speciation (see Fig. 4 in Lucek et al., 2022). In the hybrid-dysfunction model, dysploidy is linked to speciation events. Under this model, dysploidy causes an immediate reproductive barrier, as reproduction between individuals with different chromosome numbers would cause problems during meiosis. Thus, most—if not all—dysploidy events across a phylogeny would occur cladogenetically. Alternatively, the recombination-suppression model posits that chromosomal rearrangements may become fixed in lineages via either drift or selection (as some rearrangements may physically link adaptive loci or locally reduce recombination). Under this model, dysploidy would evolve primarily anagenetically across the phylogeny. The two above-described models are not mutually exclusive. The authors also introduce a third option—the hybrid-dysfunction/recombination-suppression model, under which dysploidy evolves both anagenetically and cladogenetically. While populations may be able to continue interbreeding despite some dysploidy events (which then may be fixed in the lineage), other dysploidy events may cause speciation, either because of an accumulation of differences that eventually leads to incompatibility or because of the genomic signature of the dysploidy event itself. We describe additional theory on dysploidy and macroevolution in Supplemental Section S1.

Of the few applied macroevolutionary studies that address the role of dysploidy in lineage diversification, most have focused on monocentric chromosomes, which have a single centromere (*e.g.*, Ayala and Coluzzi 2005; Freyman and Höhna 2018, but see relevant work on Lepidoptera in de Vos et al. 2020, among others), even though approximately 15–20% of eukaryotic species may have holocentric chromosomes. Holocentric chromosomes, instead of having kinetochore activity concentrated in a single point (*i.e.*, at a single centromere in monocentric chromosomes), have centromeric regions distributed along the whole chromosome where the kinetochores assemble in most of the organisms with such chromosome type (Marques et al., 2015; Márquez-Corro et al., 2019b). In monocentric chromosomes, chromosome fissions are expected to result in a loss of genetic material during meiosis and inviable gametes, as chromosome

fragments without centromeres are unable to segregate normally, and fusions are generally inherited by combining two telocentric chromosomes (Robertsonian translocations; Robertson, 1916). Holocentric chromosomes, by contrast, allow chromosome fragments to segregate normally during meiosis (Faulkner, 1972). Holocentricity may promote chromosome number variation via fission and fusion, as these changes are expected to be neutral or nearly so in holocentric organisms (Márquez-Corro et al., 2019b). Thus, holocentric organisms provide a unique system in which to study chromosomal speciation (Lucek et al., 2022).

Holocentry is distributed broadly across the Tree of Life, including 18 different lineages in animals, plants, and rhizaria (Escudero et al., 2016b; Márquez-Corro et al., 2019b), and two particular holocentric lineages show extraordinary chromosome number variation: the insect order Lepidoptera (de Vos et al., 2020) and the angiosperm sedge family Cyperaceae ($2n = 4-224$; Márquez-Corro et al., 2021, 2019a). *Carex*—the largest genus in Cyperaceae—is particularly well suited for studying the effect of dysploidy on plant diversification because it is the third most species-rich monocot family—among the tenth in angiosperms—(POWO, 2023) and there are well-developed phylogenetic and chromosome number datasets for the genus—necessary for macroevolutionary studies—(Márquez-Corro et al., 2021; Martín-Bravo et al., 2019). In *Carex*, karyotype evolves mainly through fusion and fission events, in contrast to in other sedge lineages where karyotype evolves through both dysploidy and polyploidy (Elliott et al., 2022; Márquez-Corro et al., 2019a; Shafir et al., 2023). *Carex* also has exceptional variability in holocentric chromosome number, ranging from $2n = 10$ to $2n = 132$ (Márquez-Corro et al., 2021). *Carex* has experienced several rapid radiations (Martín-Bravo et al., 2019), and shifts in optimum chromosome number are thought to have played a role in some of these radiations (*e.g.*, *Carex* sect. *Cyperoideae*; Hipp, 2007; Márquez-Corro et al., 2021).

Here, we design a new model that incorporates process variation in chromosome number evolution and disentangles the effects of dysploidy from other unobserved factors that may also affect diversification rates (Fig. 1). Previous studies that have tested for a correlation between dysploidy and diversification (*e.g.*, Márquez-Corro et al., 2021) have relied on models that fail to account for alternative sources of variation in diversification rates— *e.g.*, morphological traits, climatic niche, or biotic interactions—that are not the study's focal trait (in our case, chromosome number; Beaulieu and O'Meara, 2016). Models that fail to account for alternative sources of variation in diversification rates have high type-I error rates (Rabosky and Goldberg, 2015) because the null hypothesis of those models assume that there is no variation in rates of diversification. This high error rate motivated the development of hidden-state models with null hypotheses that account for underlying diversification-rate variation (Beaulieu and O'Meara, 2016; Caetano et al., 2018). These hidden state models not only more accurately test for associations between the focal trait and diversification, but also provide an opportunity to model how focal evolutionary processes vary across the phylogeny. For example, does chromosome number evolution occur uniformly across *Carex*? Our proposed model differs from other SSE-type models with or without hidden states (*e.g.*, Helmstetter et al., 2023) in that we are testing differences in the speciation rates associated with types of transitions rather than with the states themselves. In the case of chromosomes, the goal is not to test whether $n = 15$ or $n = 16$ have different modes of diversification, but rather if the type of karyotype change (*i.e.*, increase: $n = 15$ to $n = 16$ vs. decrease: $n = 15$ to $n = 14$) carries significant diversification signal

compared to heterogeneity (or noise) in the diversification process. Thus, our approach addresses critical issues in the role of heterogeneity in the diversification process as discussed in Caetano et al. (2018). We apply our model to the most recent *Carex* time-calibrated phylogeny with chromosome number information that contains over 700 taxa and more than 50 states (Márquez-Corro et al., 2021; Martín-Bravo et al., 2019).

Modeling cladogenesis, chromosome number evolution, and hidden rate variation

We developed the ChromoHiSSE model (Chromosome number and Hidden State-dependent Speciation and Extinction model) as an extension of the ChromoSSE model (Freyman and Höhna, 2018). Under the ChromoSSE model, lineages evolve independently under a continuous-time Markov chain (CTMC) that describes changes in chromosome numbers, speciation events, and extinction events; each of these events occurs at a particular rate (interpreted as the expected number of events per lineage per unit time). The ChromoHiSSE model includes an additional hidden trait with m states. The states of this hidden trait correspond to different sets of ChromoSSE parameters, and lineages evolve among hidden states as a Markov process with a given rate.

Under the ChromoHiSSE model, the state of a lineage is both the chromosome number, n, and the hidden state, i. For numerical stability (Freyman and Höhna, 2018; Mayrose et al., 2010), we place an upper bound on the possible number of chromosomes (k; transitions to $n > k$ are prohibited, *i.e.*, have rate 0). We denote the two hidden states as i and ii. For example, the state for a lineage could be $10i$, indicating that the lineage has 10 chromosomes and is in the hidden state i.

The ChromoHiSSE model is a stochastic process that begins with two lineages at the root, which evolve independently forward in time. As the process evolves, a lineage can experience anagenetic events (changes in chromosome number or hidden state that do not involve speciation, denoted with superscript ^{a}), cladogenetic events (speciation events that may also involve changes in the number of chromosomes or hidden state for one of the daughter lineages, denoted with superscript ^{c}), and extinction events. As in all continuous time Markov chain models, each event happens at an instantaneous point in time and multiple events cannot happen concurrently. The anagenetic events are:

(1) n increases by one ($n + 1$ increasing dysploidy) but the hidden state stays the same, which occurs at rate γ_i^a;

(2) n decreases by one ($n - 1$ decreasing dysploidy) but the hidden state stays the same, which occurs at rate δ_i^a, and;

(3) n stays the same, but the hidden state changes (*e.g.*, i to ii), which occurs at rate χ^a.

Cladogenetic events produce two daughter lineages that then evolve independently. The states of the daughters depend on the type of event:

(4) both daughters inherit the state of the ancestor, which occurs at rate ϕ_i^c;

(5) one daughter inherits n, the other inherits $n + 1$ (increasing dysploidy), and both inherit the same hidden state i, which occurs with rate γ_i^c;

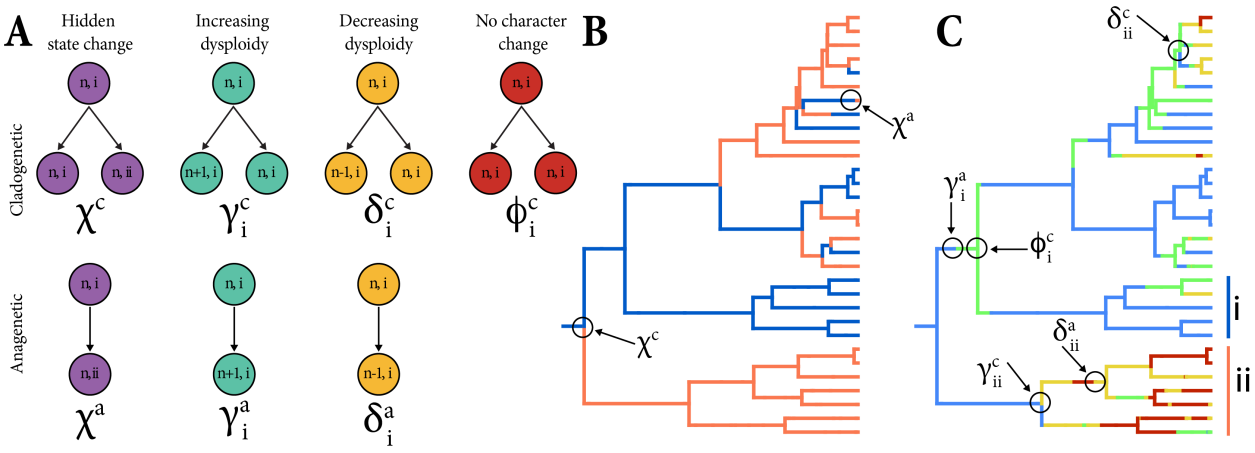

Figure 1: The ChromoHiSSE model. Panel (A) describes the event rates allowed in the model for both cladogenetic (top) and anagenetic (bottom) events. Panels (B) and (C) demonstrate those rates on a tree simulated under ChromoHiSSE. Panel (B) shows the anagenetic and cladogenetic changes in the hidden states, indicated by blue (*i*) vs. orange (*ii*). Panel (C) shows the anagenetic and cladogenetic gains and losses of chromosomes, as well as speciation with no corresponding change in chromosome numbers. More red colors in (C) correspond to more chromosomes. The vertical blue and orange bars in (C) indicate clades in the blue (*i*) vs. orange (*ii*) hidden states, displaying that chromosome number changes are less frequent in the blue hidden state than in the orange.

(6) one daughter inherits n, the other inherits $n-1$ (decreasing dysploidy), and both inherit the same hidden state i, which occurs with rate δ_i^c;

(7) both daughters inherit n from the ancestor, but one daughter changes hidden state (i to ii), which occurs at rate χ^c.

These events are depicted in Fig. 1. Additionally, all lineages go extinct at rate μ, independent of n or the hidden state. The lineages evolve forward in time until the present, at which point they are sampled independently with probability f. Extinct and unsampled lineages are pruned from the tree and the hidden state for sampled lineages is ignored; this produces a realization comprised of a reconstructed phylogeny relating the sampled lineages and a chromosome number for each sampled lineage. This stochastic process allows us to compute the probability of an observed dataset (*i.e.*, the probability that a realization under this process corresponds to our observed data) given a set of parameter values.

Note that the model presented here excludes polyploidization because previous analyses of *Carex* have found few polyploids in *Carex*, mostly concentrated in the small *Carex* subgenus *Siderostictae*—sister to the rest of *Carex*—and which has had no effect when modeling chromosomes and diversification previously (Márquez-Corro et al., 2021). We have also assumed that there are only two hidden states, and that rates of change between hidden states are symmetric; *i.e.*, the anagenetic rate of change from i to ii is the same as the rate of change from ii to i, and likewise for cladogenetic changes. Finally, we have assumed that the rate of extinction is constant among all lineages, regardless of the number of chromosomes or hidden state. We present the full ChromoHiSSE model that relaxes these assumptions and includes polyploidization (and demipolyploidization) in Supplementary Material S2.

Results

We recover two distinct "modes" of chromosome number change, which correspond to the two hidden states i, ii. The total speciation rate (λ) is very similar between these two modes (Fig. 2C), but dysploidy rates vary significantly (Fig. 2A,B,D). Hidden state i is characterized by fast rates of anagenetic and cladogenetic chromosome number change (Fig. 2, solid lines, and Table 1). Anagenetic increasing and decreasing dysploidy rates (γ_i^a and δ_i^a) are both remarkably fast, ~14 events per lineage per million years (E/L/Myr) (Table 1, Fig. 2B). Cladogenetic increasing and decreasing dysploidy (γ_i^c and δ_i^c) are significantly higher than 0 (Table 1) and faster than the cladogenetic rate of no character change (speciation without change in chromosome number, ϕ_i^c). While δ_i^c is slightly faster than γ_i^c, this difference is not significant (Table 1). The difference between speciation associated with chromosome number change and speciation without chromosome number change is greater than 0 E/L/Myr, indicating that there is a significant association between chromosome number change and faster speciation rates (Fig. 2D, blue).

In contrast, hidden state ii is characterized by slower rates of chromosome number change (Fig. 2, dashed lines, and Table 1). Anagenetic rates of change (γ_{ii}^a and δ_{ii}^a) are quite low, approximating 0 (Table 1, Fig. 2B), suggesting that little to no anagenetic change in chromosome number happens in hidden state ii. Cladogenetic rates are also quite low, and while the rate of decreasing dysploidy (δ_{ii}^c) is slightly higher than increasing dysploidy (γ_{ii}^c), this difference is not significant (Table 1). The difference between speciation associated with chromosome number change and speciation without chromosome number change is significantly less than 0 E/L/Myr, suggesting that speciation is actually slower when associated with a chromosome number change in hidden state ii (Fig. 2D, orange).

Table 1: Summary statistics of posterior distributions for parameter estimates for the ChromoHiSSE model. Light blue rows correspond to parameter estimates for hidden state i, and light orange rows correspond to parameter estimates for hidden state ii. All rate estimates are given in units of events per lineage per million years (E/L/Myr)

	Model parameter	Parameter Type	Hidden State	Median	2.5% Quantile	97.5% Quantile
χ^c	Hidden state change	Cladogenetic	i, ii	0.113	0.051	0.215
γ_i^c	Increasing dysploidy	Cladogenetic	i	0.330	0.021	0.772
γ_{ii}^c	Increasing dysploidy	Cladogenetic	ii	0.077	0.003	0.261
δ_i^c	Decreasing dysploidy	Cladogenetic	i	0.539	0.120	0.916
δ_{ii}^c	Decreasing dysploidy	Cladogenetic	ii	0.099	0.007	0.250
ϕ_i^c	No character Change	Cladogenetic	i	0.072	0.003	0.363
ϕ_{ii}^c	No character Change	Cladogenetic	ii	0.849	0.642	1.058
χ^a	Hidden state change	Anagenetic	i, ii	0.021	0.001	0.093
γ_i^a	Increasing dysploidy	Anagenetic	i	14.638	12.282	17.626
γ_{ii}^a	Increasing dysploidy	Anagenetic	ii	0.397	0.121	0.769
δ_i^a	Decreasing dysploidy	Anagenetic	i	14.191	11.903	17.246
δ_{ii}^a	Decreasing dysploidy	Anagenetic	ii	0.142	0.005	0.505

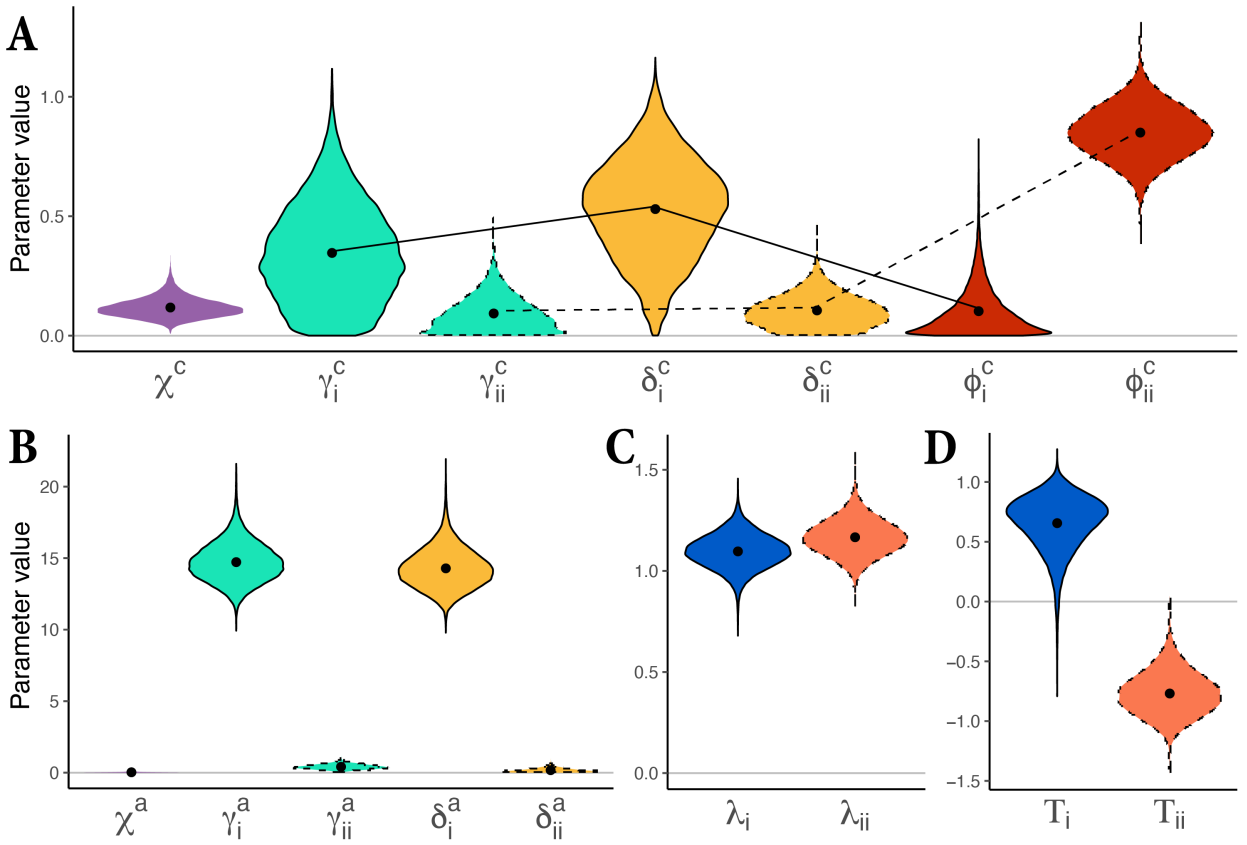

Figure 2: Posterior distributions of rate estimates. Solid lines indicate rates corresponding to hidden state *i* and dashed lines indicate hidden state *ii*. Panel (A) shows the posterior distributions of cladogenetic rates, panel (B) shows the anagenetic rates, panel (C) shows the total speciation rate per hidden state (λ), and panel (D) shows the total difference (TD) per hidden state between speciation associated with chromosome number change and speciation without a chromosome number change.

We also reconstructed the evolution of chromosome number and hidden state across the phylogeny of *Carex* (Fig. 3) using stochastic character mapping (Freyman and Höhna, 2019). Just over half (51.85%) of all branches showed a net change in chromosome number along the branch. Most clades in the phylogeny vary in chromosome number despite some shallow evolutionary divergences (Fig. 3A). For example, clade 2—which includes sect. *Pictae*, the Hirta Clade, and sect. *Praelongae*— shows substantial variation in chromosome number, from $n = 13$ to $n = 66$. However, a clade of hooked sedges (*Carex* sect. *Uncinia*; Fig. 3, clade 1 in grey), has a constant chromosome number ($n = 44$) across the 33 species included in this study, with one exception: *Carex perplexa* has a count of $n = 66$, suggesting that this species may be a demipolyploid (see Discussion). While both clades highlighted in Fig. 3 have similar crown ages, number of species, and average chromosome number ($n = 35.7$ vs. $n = 44.6$), they differ dramatically in karyotype variability and, consequently, in the inferred hidden state (clade 1 = *ii*, clade 2 = *i*, Fig. 3C). Overall, the dominant pattern of chromosome number in *Carex* is one of frequent but gradual change; 75.05% of per-branch net chromosome number changes are 3 or fewer gains or losses.

Discussion

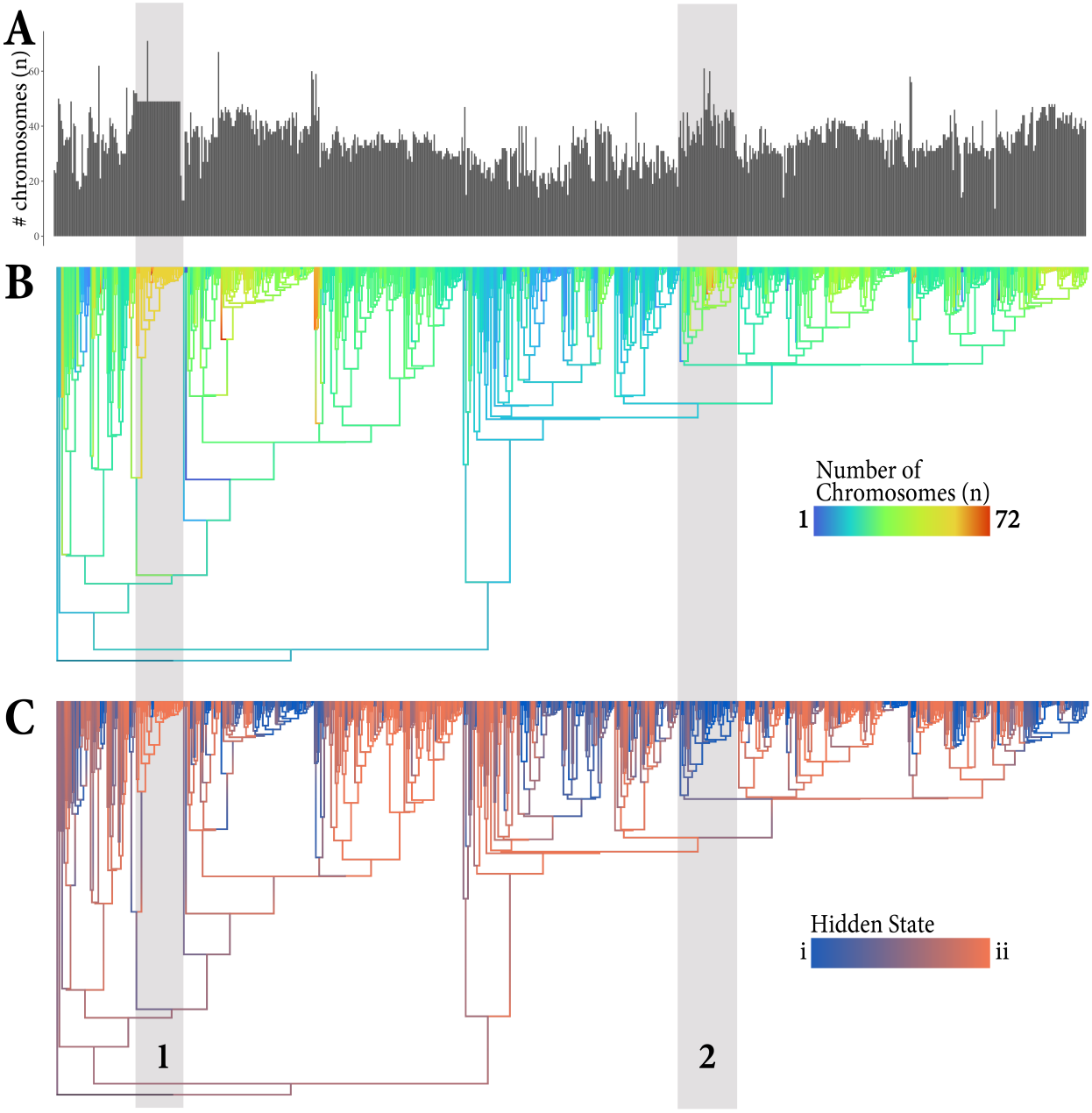

Figure 3: Reconstruction of chromosome numbers (A and B) and hidden states (C) on the *Carex* phylogeny. Panel (A) shows the distribution of haploid chromosome numbers for all extant taxa included in the analysis. Panel (B) shows the reconstructed evolution of chromosome number along branches of the phylogeny, where lighter colors indicate more chromosomes. Panel (C) shows the reconstructed evolution of the hidden state along branches of the phylogeny, where blue indicates strong statistical support for state A, orange indicates strong support for state B, and intermediary colors indicate uncertainty in the estimates. Grey bars highlight two clades (labeled 1 and 2) that are discussed in the main text.

We demonstrate that dysploidy plays an important yet complex role in diversification in *Carex*. Although the

impact of dysploidy on *Carex* diversification has been addressed previously (*e.g.*, Faulkner, 1972; Hipp, 2007; Márquez-Corro et al., 2021, 2019b), this is the first study to jointly model chromosome number change and diversification, and to demonstrate an association between higher speciation rates and dysploidy in parts of the phylogeny despite heterogeneity in the process of diversification. While gains and losses in chromosome number spur diversification along parts of the phylogeny (hidden state *i*), they have the opposite effect elsewhere in the tree (hidden state *ii*, see Fig. 3). We propose that these discrepancies may be due to the nature of holocentric chromosomes, where a single dysploidy event in isolation may not be enough to trigger reproductive isolation (Escudero et al., 2016a; Hipp et al., 2010; Lucek et al., 2022; Márquez-Corro et al., 2019b; Whitkus, 1988). However, the accumulation of sufficient dysploidy events in a lineage may form a reproductive barrier and thus trigger speciation (Escudero et al., 2016a; Whitkus, 1988). This theory is described as the recombination suppression/hybrid dysfunction chromosomal speciation model (Figure 4 in Lucek et al., 2022). Below, we discuss the evidence supporting this model and provide suggestions for future research.

Chromosomal speciation modes in *Carex*

Our new ChromoHiSSE model allows us to identify variation in modes of chromosomal evolution across the tree. In one mode, we observe high rates of anagenetic dysploidy, which indicates that dysploidy does not always create an immediate reproductive barrier. In this mode, every estimated dysploidy-related cladogenetic rate was significantly higher than non-dysploidy rates (*e.g.*, speciation with no state change, Fig. 2A), which suggests that speciation happens more frequently when associated with changes to karyotype. The results of our model show that while most dysploidy does not lead to speciation, most speciation is associated with dysploidy. This conclusion is also supported by additional evidence from prior studies, discussed below. The other mode, by contrast, is characterized by lower rates of cladogenetic dysploidy, and anagenetic dysploidy rates are also quite low. However, the overall total speciation rate (λ) between the two modes is very similar (Fig. 2C), demonstrating that—in parts of the tree characterized by this latter mode—additional traits may be contributing to the fast speciation rates.

We can see these these different modes of macroevolution illustrated in specific clades of our *Carex* phylogeny. For example, *Carex* sect. *Uncinia* (hook sedges, highlighted as clade 1 in Fig. 3) is characterized by hooked utricles that allow for long-distance dispersal through epizoochory (García-Moro et al., 2022). The hook sedges exemplify a *Carex* clade that is characterized by low-to-zero dysploidy but fast rates of speciation as an apparently young radiation (Martín-Bravo et al., 2019)—though this was not formally tested in our study—and thus may be a particularly appealing candidate for future studies focusing on potential adaptive traits (not related to chromosome number) that led to such biodiversity. This lineage is characterized by karyotypes with numerous chromosomes (Márquez-Corro et al., 2021), mostly reported from New Zealand, where ca. half of the section diversified. However, this only gives partial information on their evolutionary history, as the karyotypes of the Andean relatives have not been studied in detail yet. Most of the New Zealand chromosome counts correspond to a single, old report and there is some variation in the few existing counts from South America, so these results may be taken with caution. If the South American species show

wider karyotypic variation than currently detected in comparison to the New Zealand lineage, there may be different drivers enhancing speciation on each side of the South Pacific.

Some *Carex* species are known to have striking chromosome-number polymorphism, even within populations (Escudero et al., 2023a, 2013b; Luceño and Castroviejo, 1991; Whitkus, 1988). For example, *Carex scoparia* varies from $n = 28$ to $n = 35$ (Escudero et al., 2013b), and individuals with different chromosome numbers are able to reproduce and exchange alleles, maintaining gene flow despite chromosome number differences. This polymorphism suggests that some chromosome number differences are not sufficient to create reproductive barriers (Hipp et al., 2009), which is supported by our results; often, inferred chromosome number changes do not result in an immediate speciation event on the phylogeny. The question remains, why does dysploidy result in speciation in some cases and not in others? One possibility is that an accumulation of changes may eventually lead to reproductive isolation, where one "last straw" dysploidy event triggers speciation (referred to as the last-straw hypothesis). We discuss potential model developments that could test this theory below in the *Modeling chromosomal speciation* section. Another possibility is that rearrangements in some parts of the genome are more stable than others, and the genomic architecture—where the fragmentation or fusion occurs in the genome—determines the evolutionary effects of dysploidy.

The genomic architecture of clades might point to why dysploidy has a stochastic effect on speciation. A recent study (Cornet et al., 2023) argues that patterns of repeat DNA—*e.g.*, LINEs, LTRs, and Helitrons— are significantly different when comparing fast vs. slow chromosome evolving *Carex* lineages—while this was only tested with a small number of species, those species represent all major *Carex* clades—and, interestingly, repeat DNA regions often serve as hotspots for genome rearrangements (Escudero et al., 2023b; Höök et al., 2023). These repeat regions are significantly more common in the genomes of rapidly evolving lineages (Cornet et al., 2023). It may well be that holocentry and dysploidy evolution together form an adaptive mechanism that enables the evolution of adaptive linkage blocks. Comparative genomics may also help to decipher the relationships between dysploidy, diversification, and supergenes in *Carex*. Synteny in holocentric sedge chromosomes is more conserved than we would expect if chromosome evolution were unconstrained, even in comparisons between species that span deep nodes in the phylogeny (Escudero et al., 2023b). This is surprising given the high rates of chromosome evolution inferred in this study. The massive diversity of sedges is likely an outcome of both processes: recombination and reproductive isolation evolving as chromosomes split and fuse.

An important caveat to our study is the presence of missing data—particularly for tropical lineages—both in terms of sequenced taxa represented in the phylogeny and for available chromosome counts. While the phylogeny used in this study was assembled with a HybSeq backbone and three DNA regions for ca. 1400 out of 2000 species, phylogenetic relationships in some areas of the genus are still tenuous (Jiménez-Mejías et al., 2016; Martín-Bravo et al., 2019; Roalson et al., 2021), and sequencing more individuals from the uncertain clades may help resolve those relationships. Additionally, there are only chromosome counts for $\sim$ 700 species in the phylogeny. Filling gaps in our karyological knowledge—especially for tropical taxa—will promote a better understanding of chromosome evolution across the phylogeny of *Carex* and its relationship with other adaptive traits (Spalink et al., 2018). Thus, molecular and karyolog-

ical studies can still continue to improve our knowledge of how dysploidy is taking part in *Carex* evolution (Escudero et al., 2023a, 2013a,b; Márquez-Corro et al., 2021). This is particularly true within a few noteworthy clades. One of the most species-diverse lineages within *Carex*—the Decora Clade/*Carex* sect. *Indicae* (Roalson et al., 2021)—only includes five species with reported chromosome numbers (Márquez-Corro et al., 2021). This pantropical lineage alone would constitute ca. 10% of the genus (Roalson et al., 2021). Also, we include only a few New Zealand species from *Carex* sect. *Uncinia* (clade 1 in Fig. 3), as there are few known karyotypes for the South American representatives of the lineage. Other traits such as hooked utricles and epizoochory might be as or more important for diversification as karyotype in this clade, but we cannot tease these effects apart without additional chromosome count data. While *Carex* biodiversity is greatest in temperate regions including the Global North, this bias in data availability impedes our ability to detect evolutionary patterns of the genus as a whole.

Modeling chromosomal speciation

The methodological innovations presented here—at the cutting edge of macroevolutionary model development—allow us to model process variation in chromosome number evolution, despite the computational challenges associated with the huge state space of such a model. Process variation is fundamental to our conclusions; only by incorporating multiple modes do we discover that—in some parts of the phylogeny—dysploidy is rare and does not lead to faster speciation. In fact, when we implement our model without process variation, our results suggest a strong, uniform boost in speciation rate associated with dysploidy (results presented in Supplemental Section S3). Future work that builds off of our novel modeling approach will allow us to pursue promising avenues for innovative research; we highlight two examples below.

First, like all birth-death models, our ChromoHiSSE model operates with species as the fundamental unit of analysis (the tips in the tree) and thus does not formally model chromosome number polymorphism that is present in some *Carex* lineages. Second, while ChromoHiSSE tests for the effect of single changes in chromosome number on diversification rates, it cannot test for the effect of an accumulation of changes (the last-straw hypothesis). However, ChromoHiSSE could be modified to include tip-state polymorphism (*e.g.*, a dysploid series) as additional hidden states (*e.g.*, a particular tip either has 12, 13, or 14 chromosomes). Additionally, Goldberg and Foo (2020) describe a mechanism for modeling memory (thus the accumulation of chromosome number changes) in a macroevolutionary framework using hidden states, which could be applied to test the last-straw hypothesis.

Conclusions

Our work demonstrates the important role of dysploidy on diversification in *Carex*, a model lineage for understanding how karyotype rearrangements via dysploidy affect speciation and macroevolutionary dynamics. Our results—using the new ChromoHiSSE model—paint a complex picture of how dysploidy affects speciation in a clade characterized by high species diversity, high morphological disparity, and holocentric chromosomes (Martín-Bravo et al., 2019), and

our results support the recombination suppression/hybrid dysfunction chromosomal speciation model, in which only some karyotype rearrangements trigger reproductive isolation and thus speciation. Future work on the underlying genomic mechanisms of chromosomal speciation via comparative genomics will be particularly powerful for linking across scales, from molecules to lineages. Ultimately, our novel modeling approach also serves as a critical step towards even more complex and powerful macroevolutionary analyses that incorporate intraspecific chromosome number variation and track the accumulation of change through time.

Materials and Methods

To test the role of chromosome number changes on speciation and extinction patterns in the genus *Carex*, we developed the ChromoHiSSE model (Fig. 1). This model estimates cladogenetic and anagenetic rates linked to dysploidy, plus an unmeasured variance that might explain the observed data (*i.e.*, hidden state).

Chromosome and phylogenetic data

We performed all analyses on a large dataset, which includes a phylogeny of *Carex* with 755 taxa ($\sim$ 40% of extant diversity), representing all *Carex* subgenera and most of the sections for which chromosome counts have been reported. Our dataset also includes haploid chromosome number (n) for all tips in the tree; both the tree and chromosome number data come from Márquez-Corro et al. (2021). The original tree, based on a HybSeq backbone and three DNA regions (ITS, ETS and *mat*K) from ca. 1400 species out of 2000 *Carex* species, was published by Martín-Bravo et al. (2019). Tips without chromosome number information were pruned for the current study. Chromosome number in *Carex* evolves through dysploid events, except for in the small subgenus *Siderostictae* (13 species in our tree)—sister to the rest of the genus—which includes species with different reported ploidy levels (Márquez-Corro et al., 2021, 2019a). To avoid modeling rare polyploidy events that occur only in a small part of the tree, we removed *Siderostictae* from the primary analysis (though see Supplemental Section S3 for results with this clade included). For the remaining taxa, we coded chromosome numbers as the most frequent haploid number or the lowest haploid cytotype in polyploid lineages from the Márquez-Corro et al. (2021) dataset for those tips. The final data matrix included n-values ranging from 5 to 66 chromosomes for 742 taxa.

Model implementation

We implemented ChromoHiSSE in `RevBayes` (Höhna et al., 2016), a software for specifying Bayesian probabilistic graphical models primarily for phylogenetics and phylogenetic comparative methods. Due to the large state space of our model and cladogenesis, we used the newly-developed `TensorPhylo` plugin (May and Meyer, 2022) to accelerate likelihood calculations and thus achieve convergence in a more reasonable time frame. We ran two chains of the analysis and assessed convergence in the R programming language (R Core Team, 2013). We processed the traces and

removed 10% of the generations per chain as burnin using `RevGadgets`. We then calculated the effective sample size (ESS) value for each parameter in each chain using `coda` (Plummer et al., 2006) and verified that the harmonic mean of the ESS values of each chain was greater than 200. We additionally visually inspected all model parameters across both chains in Tracer (Rambaut et al., 2018). For any parameters that appeared to have strikingly non-normal posterior distributions, we also estimated a transformed ESS following Vehtari et al. (2021). We subsequently combined both runs for all downstream analyses, discarding 10% of the total generations per chain as burnin prior to combining.

Analysis and postprocessing

We summarized posteriors and plotted results in R (R Core Team, 2013) using the R package `RevGadgets` (Tribble et al., 2022). We additionally transformed model parameters to produce two types of useful summary statistics. First, we calculated the total speciation rate in hidden state i vs. ii as $\lambda = \chi^c + \gamma^c + \delta^c + \phi^c$. Second, we wanted to know if the rate of speciation concurrent with changes in chromosome number ($\gamma^c + \delta^c$) is greater or less than the rate of speciation with no change in chromosome number ($\chi^c + \phi^c$). We thus estimated an additional metric: $TD = (\gamma^c + \delta^c) - (\chi^c + \phi^c)$, such that positive values of TD indicate more speciation with chromosome number change and negative values indicate less speciation with chromosome number change. We estimate TD for each hidden state to estimate how chromosome number changes associated with speciation vary between modes of the model.

All code for implementing and running the model and processing and plotting the results is available at Zenodo DOI: 10.5281/zenodo.8320249.

Acknowledgements

JIM-C was granted by the 'Next Generation EU' funding, the Recovery Plan, Transformation and Resilience and the Ministry of Universities, under the grants 'Margarita Salas' for the requalification of the Spanish university system 2021–2023 called by the Universidad Pablo de Olavide, Seville. ME was supported by the MICINN-FEDER, Project DiversiChrom (PID2021-122715NB-I00). CMT was supported by the School of Life Sciences at the University of Hawai'i at Mānoa funded through the Office of the Vice President of Research. This material is based upon work supported by the NSF Postdoctoral Research Fellowships in Biology Program under Grant No. 2109835 to CMT. RZF was supported by NSF-DEB 2323170.

References

Ayala, F. J. and Coluzzi, M. (2005). Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proceedings of the National Academy of Sciences*, 102:6535–6542.

Beaulieu, J. M. and O'Meara, B. C. (2016). Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic biology*, 65(4):583–601.

Caetano, D. S., O'Meara, B. C., and Beaulieu, J. M. (2018). Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution*, 72(11):2308–2324.

Cornet, C., Pablo, M., Augustijnen, H., Nguyen, P., Escudero, M., and Lucek, K. (2023). Holocentric repeat landscapes: from microevolutionary patterns to macroevolutionary associations with karyotype evolution. *Molecular Ecology*, 000(000):000–000.

de Vos, J. M., Augustijnen, H., Bätscher, L., and Lucek, K. (2020). Speciation through chromosomal fusion and fission in Lepidoptera. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1806):20190539.

Elliott, T. L., Zedek, F., Barrett, R. L., Bruhl, J. J., Escudero, M., Hroudová, Z., Joly, S., Larridon, I., Luceño, M., Márquez-Corro, J. I., Martín-Bravo, S., Muasya, A. M., Šmarda, P., Thomas, W. W., Wilson, K. L., and Bureš, P. (2022). Chromosome size matters: genome evolution in the cyperid clade. *Annals of Botany*, 130(7):999–1014.

Escudero, M., Arroyo, J. M., González-Ramírez, S., and Jordano, P. (2023a). Founder events and subsequent genetic bottlenecks underlie karyotype evolution in the ibero - north african endemic *Carex helodes*. *Annals of Botany*, 000:000–000.

Escudero, M., Hahn, M., Brown, B. H., Lueders, K., and Hipp, A. L. (2016a). Chromosomal rearrangements in holocentric organisms lead to reproductive isolation by hybrid dysfunction: the correlation between karyotype rearrangements and germination rates in sedges. *American Journal of Botany*, 103(8):1529–1536.

Escudero, M., Maguilla, E., and Luceño, M. (2013a). Selection by climatic regime and neutral evolutionary processes in holocentric chromosomes (*Carex* gr. *laevigata*: Cyperaceae): A microevolutionary approach. *Perspectives in Plant Ecology, Evolution and Systematics*, 15(2):118–129.

Escudero, M., Marques, A., Lucek, K., and Hipp, A. L. (2023b). Genomic hotspots of chromosome rearrangements explain conserved synteny despite high rates of chromosome evolution in a holocentric lineage. *Molecular Ecology*, 000(000):000–000.

Escudero, M., Márquez-Corro, J. I., and Hipp, A. L. (2016b). The phylogenetic origins and evolutionary history of holocentric chromosomes. *Systematic Botany*, 41(3):580–585.

Escudero, M., Martín-Bravo, S., Itay, M., Fernández-Mazuecos, M., Fiz-Palacios, O., Hipp, A. L., Pimentel, M., Valcárcel, V., Vargas, P., and Luceño, M. (2014). Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PloS One*, 9(1):e85266.

Escudero, M., Weber, J. A., and Hipp, A. L. (2013b). Species coherence in the face of karyotype diversification in holocentric organisms: the case of a cytogenetically variable sedge (*Carex scoparia*, Cyperaceae). *Annals of Botany*, 112(3):515–526.

Faulkner, J. (1972). Chromosome studies on *Carex* section *Acutae* in north-west Europe. *Botanical Journal of the Linnean Society*, 65(3):271–301.

Freyman, W. A. and Höhna, S. (2018). Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. *Systematic Biology*, 67(2):195–215.

Freyman, W. A. and Höhna, S. (2019). Stochastic character mapping of state-dependent diversification reveals the tempo of evolutionary decline in self-compatible Onagraceae lineages. *Systematic Biology*, 68(3):505–519.

García-Moro, P., Otero, A., Benítez-Benítez, C., Costa, L., Martín-Bravo, S., Naczi, R. F., Reznicek, A., Roalson, E. H., Starr, J. R., and Jiménez-Mejías, P. (2022). Biogeography and systematics of *Carex* subgenus *Uncinia* (Cyperaceae): A unique radiation for the genus *Carex* in the Southern Hemisphere. *Taxon*, 71(3):587–607.

Goldberg, E. E. and Foo, J. (2020). Memory in trait macroevolution. *The American Naturalist*, 195(2):300–314.

Helmstetter, A. J., Zenil-Ferguson, R., Sauquet, H., Otto, S. P., Méndez, M., Vallejo-Marin, M., Schönenberger, J., Burgarella, C., Anderson, B., de Boer, H., Glémin, S., and Käfer, J. (2023). Trait-dependent diversification in angiosperms: Patterns, models and data. *Ecology Letters*, 26(4):640–657.

Hipp, A., Rothrock, P., and Roalson, E. (2009). The evolution of chromosome arrangements in *Carex* (Cyperaceae). *The Botanical Review*, 75:96–109.

Hipp, A. L. (2007). Nonuniform processes of chromosome evolution in sedges (*Carex*: Cyperaceae). *Evolution*, 61(9):2175–2194.

Hipp, A. L., Rothrock, P. E., Whitkus, R., and Weber, J. A. (2010). Chromosomes tell half of the story: the correlation between karyotype rearrangements and genetic diversity in sedges, a group with holocentric chromosomes. *Molecular Ecology*, 19:3124–3138.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., and Ronquist, F. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, 65(4):726–736.

Höök, L., Näsvall, K., Vila, R., Wiklund, C., and Backström, N. (2023). High-density linkage maps and chromosome level genome assemblies unveil direction and frequency of extensive structural rearrangements in wood white butterflies (*Leptidea* spp.). *Chromosome Research*, 31(1):2.

Jiménez-Mejías, P., Hahn, M., Lueders, K., Starr, J. R., Brown, B. H., Chouinard, B. N., Chung, K.-S., Escudero, M., Ford, B. A., Ford, K. A., Gebauer, S., Gehrke, B., Hoffmann, M. H., Jin, X.-F., Jung, J., Kim, S., Luceño, M., Maguilla, E., Martín-Bravo, S., Míguez, M., Molina, A., Naczi, R. F. C., Pender, J. E., Reznicek, A. A., Villaverde, T., Waterway, M. J., Wilson, K. L., Yang, J.-C., Zhang, S., Hipp, A. L., and Roalson, E. H. (2016). Megaphylogenetic Specimen-Level Approaches to the *Carex* (Cyperaceae) Phylogeny Using ITS, ETS, and *mat*K Sequences: Implications for Classification The Global *Carex* Group). *Systematic Botany*, 41(3):500–518.

Lucek, K., Augustijnen, H., and Escudero, M. (2022). A holocentric twist to chromosomal speciation? *Trends in Ecology & Evolution*, 37(8):655–662.

Luceño, M. and Castroviejo, S. (1991). Agmatoploidy in *Carex laevigata* (Cyperaceae). Fusion and fission of chromosomes as the mechanism of cytogenetic evolution in Iberian populations. *Plant Systematics and Evolution*, 117:149–159.

Magallón, S. and Castillo, A. (2009). Angiosperm diversification through time. *American journal of botany*, 96(1):349–365.

Magallón, S., Sánchez-Reyes, L. L., and Gómez-Acevedo, S. L. (2019). Thirty clues to the exceptional diversification of flowering plants. *Annals of Botany*, 123(3):491–503.

Mandáková, T. and Lysak, M. A. (2018). Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. *Current Opinion in Plant Biology*, 42:55–65.

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., Pellino, M., Fuchs, J., Ma, W., Kuhlmann, M., Brandt, R., Vanzela, A. L. L., Beseda, T., Šimková, H., Pedrosa-Harand, A., and Houben, A. (2015). Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. *Proceedings of the National Academy of Sciences*, 112(44):13633–13638.

Márquez-Corro, J. I., Martín-Bravo, S., Jiménez-Mejías, P., Hipp, A. L., Spalink, D., Naczi, R. F., Roalson, E. H., Luceño, M., and Escudero, M. (2021). Macroevolutionary insights into sedges (*Carex*: Cyperaceae): The effects of rapid chromosome number evolution on lineage diversification. *Journal of Systematics and Evolution*, 59(4):776–790.

Márquez-Corro, J. I., Martín-Bravo, S., Spalink, D., Luceño, M., and Escudero, M. (2019a). Inferring hypothesis-based transitions in clade-specific models of chromosome number evolution in sedges (Cyperaceae). *Molecular Phylogenetics and Evolution*, 135:203–209.

Márquez-Corro, J. I., Martín-Bravo, S., Pedrosa-Harand, A., Hipp, A. L., Luceño, M., and Escudero, M. (2019b). *Karyotype evolution in holocentric organisms*, pages 1–7. John Wiley & Sons, Ltd.

Martín-Bravo, S., Jiménez-Mejías, P., Villaverde, T., Escudero, M., Hahn, M., Spalink, D., Roalson, E. H., Hipp, A. L., Group, G. C., Benítez-Benítez, C., et al. (2019). A tale of worldwide success: Behind the scenes of *Carex* (Cyperaceae) biogeography and diversification. *Journal of Systematics and Evolution*, 57(6):695–718.

May, M. R. and Meyer, X. (2022). TensorPhylo RevBayes plugin. https://bitbucket.org/mrmay/tensorphylo/src/master.

Mayrose, I., Barker, M. S., and Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology*, 59(2):132–144.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news*, 6(1):7–11.

POWO (2023). Plants of the World Online, Facilitated by the Royal Botanic Gardens, Kew. http://www.plantsoftheworldonline.org/. Retrieved: 2023-08-06.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabosky, D. L. and Goldberg, E. E. (2015). Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology*, 64(2):340–355.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology*, 67(5):901–904.

Roalson, E. H., Jiménez-Mejías, P., Hipp, A. L., Benítez-Benítez, C., Bruederle, L. P., Chung, K.-S., Escudero, M., Ford, B. A., Ford, K., Gebauer, S., Gehrke, B., Hahn, M., Hayat, M. Q., Hoffmann, M. H., Jin, X.-F., Kim, S., Larridon, I., Léveillé-Bourret, E., Lu, Y.-F., Luceño, M., Maguilla, E., Márquez-Corro, J. I., Martín-Bravo, S., Masaki, T., Míguez, M., Naczi, R. F. C., Reznicek, A. A., Spalink, D., Starr, J. R., Uzma, Villaverde, T., Waterway, M. J., Wilson, K. L., and Zhang, S.-R. (2021). A framework infrageneric classification of *Carex* (Cyperaceae) and its organizing principles. *Journal of Systematics and Evolution*, 59(4):726–762.

406 Robertson, W. (1916). Chromosome studies. I. Taxonomic relationships shown in the chromosomes of Tettigidae and Acrididae:
407 V-shaped chromosomes and their significance in Acrididae, Locustidae, and Gryllidae: chromosomes and variation. *Journal of*
408 *Morphology*, 27:179—-331.

409 Sauquet, H. and Magallón, S. (2018). Key questions and challenges in angiosperm macroevolution. *New Phytologist*, 219(4):1170–1187.

410 Shafir, A., Halabi, K., Escudero, M., and Itay, M. (2023). A non-homogeneous model of chromosome-number evolution to reveal shifts
411 in the transition patterns across the phylogeny. *New Phytologist*, 238(4):1733–1744.

412 Spalink, D., Pencer, J., Escudero, M., Hipp, A. L., Roalson, E. H., Starr, J. R., Waterway, M., Bohs, L., and Sytsma, K. J. (2018). The
413 spatial structure of phylogenetic and functional diversity in the United States and Canada: An example using the sedge family
414 (Cyperaceae). *Journal of Systematics and Evolution*, 56(5):449–465.

415 Tribble, C. M., Freyman, W. A., Landis, M. J., Lim, J. Y., Barido-Sottani, J., Kopperud, B. T., Höhna, S., and May, M. R. (2022). RevGad-
416 gets: an R Package for visualizing Bayesian phylogenetic analyses from RevBayes. *Methods in Ecology and Evolution*, 13(2):314–323.

417 Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An
418 improved R for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16(2):667–718.

419 Whitkus, R. (1988). Experimental hybridizations among chromosome races of *Carex pachystachya* and the related species *C. macloviana*
420 and *C. preslii* (Cyperaceae). *Systematic Botany*, 13(1):146–153.