

# Endemism in Recently Diverged Angiosperms Is Associated With Chromosome Set Duplication

**Sara Villa**

University of Milan: Università degli Studi di Milano

**Matteo Montagna**

Università degli Studi di Milano

**Simon Pierce** (✉ [simon.pierce@unimi.it](mailto:simon.pierce@unimi.it))

Università degli Studi di Milano: Università degli Studi di Milano <https://orcid.org/0000-0003-1182-987X>

---

## Research Article

**Keywords:** adaptative radiation, apo-endemics, diversity creation, genome doubling, neo-endemism, speciation mechanism

**Posted Date:** June 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-550165/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Chromosome set duplication (polyploidy) drives instantaneous speciation and shifts in ecology for angiosperms, and is frequently observed in neo-endemic species. However, the extent to which chromosome set duplication is associated with endemism throughout the flowering plants has not been determined. We hypothesised that across the angiosperms polyploidy is more frequent and more pronounced (higher evident ploidy levels) for recent endemics. Data on chromosome counts, molecular dating and distribution for 4210 species belonging to the major clades of angiosperms were mined from literature-based databases. As all clades include diploid taxa, with polyploids representing a possible 'upper limit' to the number of chromosomes over evolutionary time, upper boundary regression was used to investigate the relationship between the number of chromosomes and time since taxon divergence, both across clades and separately for families, with endemic and non-endemic species compared. A significant negative exponential relationship between the number of chromosomes and taxon age was evident across angiosperm clades ( $R^2_{\text{adj}}=0.48$  with endemics and non-endemics considered together,  $R^2_{\text{adj}}=0.46$  for endemics;  $R^2_{\text{adj}}=0.44$  for non-endemics;  $p \leq 0.0001$  in all cases), which was three times stronger for endemics (decay constant=0.12, *cf.* 0.04 for non-endemics). The majority of families exhibited this relationship, with a steeper regression slope for endemic Campanulaceae, Compositae, Leguminosae, Poaceae, Caryophyllaceae and Rosaceae, *cf.* non-endemics. In conclusion, chromosome duplication is more frequent and extensive in recent angiosperms, particularly young endemics, supporting the hypothesis of polyploidy as a key driver of neo-endemism. However, as young endemics may also be diploid, polyploidy is not an exclusive driver of neo-endemism.

# Introduction

Why endemic species occur spontaneously in restricted geographic ranges is a key question in biogeography and ecology with implications for the conservation of rare and nationally important species (Olivieri et al. 2015). For many endemics, reasons for range restriction are evident. For example, they may be limited to a small area because they can occupy a limited ecological niche (Williams et al. 2009), or have originated by hybridization and occur only in the area of contact between the progenitor species (Grünig et al. 2021), they may be relict species remaining from a much wider past range (paleo-endemics; Petrova et al. 2015) or are recent species (neo-endemics, *sensu* Stebbins and Major 1965) yet to disperse beyond the centre of origin (Behroozian et al. 2020). An additional explanation may lie with genome duplication (polyploidy), which effectively represents an 'instantaneous', sympatric speciation mechanism (Mayr 1963), and can be a major driver of plant evolution (Levin 1983; Otto and Whitton 2000; Soltis et al. 2014; Van de Peer 2017). Polyploidy occurs when the cell division cycle includes chromosome duplication but lacks the subsequent stages of chromosome and cell component separation, resulting in cells with multiple copies of chromosomes. This can occur during mitosis (in a process known as somatic doubling) or meiosis (non-reduction during sporogenesis), either within populations of a single species (autopolyploidy), or subsequent to interspecific hybridization (allopolyploidy) (Ramsey and Schemske 1998). Multiple copies of chromosomes require larger nuclei and cells, often

altering the physiology and morphological traits of the offspring with respect to parental species: indeed, these events are usually linked with key innovations in phenotype that effect fitness (for example, polyploids exhibit a greater tendency to use vegetative reproduction; Herben et al. 2017; see also Soltis and Soltis 2016). Altered phenotypes in polyploids are known to change the ecology of these species with respect to diploid progenitors, for example larger cell size and larger organs (the 'giga effect') produce larger flowers of different colours and scents that favour different pollinator species (see examples in Rezende et al. 2020). Exhibiting different sets of chromosomes, polyploids are often reproductively isolated from mother plants (Husband and Sabara 2004; Laport et al. 2016; Lavanaia 2020). Ancient events of genome doubling are thus associated with increased rates of speciation (Husband et al. 2013; Soltis and Soltis 2016). Polyploidy occurs frequently in plants, and particularly in angiosperms (Soltis and Soltis 2016; Lavanaia 2013, 2020): it has been determined that all angiosperms show evidence of multiple events of genome doubling, except for basal *Amborella trichopoda* Baill. (Soltis et al. 2009; Jiao et al. 2011; Lavanaia 2020). Polyploid neo-endemics have been shown to exhibit a particularly wide range of functional traits and capacity to adapt, with geographic range constrained simply by limited time for dispersal (Behroozian et al. 2020).

However, despite being widely recognised as an important process for plant evolution and ecology, the extent to which chromosome set duplication represents a general mechanism in the emergence of endemic species throughout the angiosperms has yet to be quantified.

Understanding the extent to which chromosome set duplication is associated with endemism is complicated by the fact that endemism is defined, on a case-by-case basis, in terms of geographical regions or political boundaries, which are extremely variable and often biologically irrelevant. Furthermore, the distinction between 'neo-endemic' and 'paleo-endemic' may be based on perceived past distributions inferred from current distributions, rather than objective measures of events over time. Additionally, it can also be difficult to define the range of an endemic species, which can be relatively extensive (i.e. "continental endemic") or comparatively restricted (i.e. "local endemic") (Coelho et al. 2020). Moreover, geographical range determined from occurrence records and species distribution models does not always consider local compliance with the ecological requirements of the species, the degree of habitat preservation or the presence of highly **anthropized** areas, and thus should be interpreted carefully (Brookes et al. 2010; Beck et al. 2014; Fithian et al. 2014). The highest richness of endemics is found in biodiversity hotspots (Cañadas et al. 2014; Noroozi et al. 2018), which have been identified in 36 areas around the globe (Conservation International, [www.conservation.org](http://www.conservation.org); Critical Ecosystem Partnership Fund, [www.cepf.net](http://www.cepf.net)), and range from 18,972 Km<sup>2</sup> (New Caledonia) to 2,373,057 Km<sup>2</sup> (Indo-Burma region). Such heterogeneity often requires the identification of smaller, higher-concentration areas within these regions ("hotspots within hotspots") to effectively target conservation efforts (Cañadas et al. 2014; Noroozi et al. 2018).

A further complication may arise from the fact that ploidy level is not always easy to define, even when the basic number of chromosomes is known for a taxon. Indeed, this can be the result of ancient autopolyploidy or allopolyploidy (Parisod et al. 2010; Lavanaia 2013; Zozomová-Lihová et al. 2014),

sometimes followed by diploidization events and thus chromosome loss (i.e. diploidized paleo-polyploids; Tamayo-Ordóñez et al. 2016; Qiao et al. 2019). Moreover, some taxa show intraspecific variability, with multiple chromosome counts and different ploidy levels, which is expected when polyploidy events occur relatively frequently (Husband et al. 2013). While polyploidy multiplies sets of chromosomes (and thus has striking effects during karyotype evolution), a range of processes can subtly rearrange single chromosomes, altering chromosome number or characteristics such as size and DNA content. These include insertion, deletion or duplication, inversion, and intra- or inter-chromosomal reciprocal translocation, which can be particularly evident for paleo-species that have accumulated changes over time (reviewed by Schubert and Lysak 2011). Essentially, change in the overall set of chromosomes occurs due to polyploidy, further mediated by complicating processes that can both increase and decrease the number of chromosomes; the relationship between the number of chromosomes and polyploidy is not necessarily straightforward. With this *caveat* in mind, in the present study it is assumed that entire genome multiplication in the sporophyte generation ( $2n$ ) is principally affected by polyploidy, and the number of chromosomes is used as a quantitative measure to represent the net result of karyotype evolution (see Methods for details).

Despite the considerable attention given to the classification of endemic species in terms of when and how they originated (Favarger and Contandriopoulos 1961; Stebbins and Major 1965; Maers and Giller 2013), an age threshold differentiating paleo- from neo-endemics remains undefined. Indeed, the term paleo-endemic refers more to a process than to a particular time or period *per se* (i.e. endemism by restriction or fragmentation of a previously extensive range). However, when comparing speciation events and the evolution of taxa it may also be useful to refer to the absolute framework of geological time and the climatic and geographic context that this provides. While species younger than 1 million years are clearly neo-endemics (Kraft et al. 2010), the issue becomes complex for less recent species. Even without providing a precise criterion, numerous attempts have been made to identify a clear discriminating principle. Ferreira and Boldrini (2011), for example, addressed the problem by suggesting the combination of a dated phylogeny (i.e. estimated age and degree of systematic isolation) with environmental context (based on stratigraphy and the age of underlying rocks). Lazarina et al. (2019) considered reproductive and geographic isolation, while Mishler et al. (2014) proposed a Categorical Analysis of Neo- And Paleo-Endemism (CANAPE) based on the Relative Phylogenetic Endemism index, to distinguish centres of neo- and paleo-endemism. However, these methods are too unwieldy to be used for a prompt distinction between neo- and paleo-endemics in large datasets, and an age threshold fixed *a priori*, despite being relatively imprecise, provides a consistent and readily applicable approach. Among the recently formed endemics, particular relevance has been given to apo-endemics, or polyploids diverged from diploid progenitors (Stebbins and Major 1965). However, the origin of a polyploid and divergence in the case of sympatric speciation is not always possible to date (Doyle and Egan 2010). Indeed, dating polyploidy events and their role in creating new taxa has so far been limited to the timing of major clade emergences (Wood et al. 2009), lacking sufficient detail to compare particular species within families or genera. In the present study, rather than entering into the debate regarding what

constitutes a neo- as opposed to a paleo-endemic, we refer simply to the time period elapsed since the divergence of the taxon.

In summary, the comparison of estimated taxon age (Ma since divergence) with the number of chromosomes will allow investigation of whether the chromosome complement is higher in recent species than in ancient ones; information on the occurrence range of species will allow assessment of whether the phenomenon is general within the angiosperms or relatively prevalent in endemics. Based on these data, the principal objective of the present study is to assess whether polyploidization events are principal drivers of the emergence of new endemic species. Specifically, it is hypothesized that, despite a prevalence of diploid taxa throughout evolutionary time: 1). the highest chromosome counts are evident for angiosperm taxa that have diverged recently, 2). higher numbers of chromosomes are particularly evident for recent endemic (*cf.* non-endemic) taxa, 3). the character of the ploidy level/divergence time relationship is consistent throughout the angiosperms, from ancient to recently diverged clades.

## Materials And Methods

### *Data mining*

The relationship between genome duplication and neo-endemism for angiosperms involved analysis of sporophyte chromosome count data, with neo-endemism (reflecting both restricted geographic distribution and taxon age) being represented by 'time since divergence' and geographic presence/absence data. These data were mined from databases containing values from the scientific literature, and directly from the literature itself, and aimed to broadly represent both endemic and non-endemic taxa across the angiosperms. Thus data mining was performed for major clades of angiosperms following the recent phylogeny of Leebens-Mack et al. (2019), starting with ANA-grade taxa (represented by Nymphaeaceae – other families in this clade are too under-represented in terms of both chromosome number and taxon age data), and following the phylogeny via the monocots (represented by Poaceae and Orchidaceae), Magnoliids (Magnoliaceae), Ranunculids (Ranunculaceae), Caryophyllales (Caryophyllaceae), Asterids (Apiaceae, Campanulaceae, Compositae, Ericaceae, Primulaceae), Saxifragales (Saxifragaceae) and Core Rosids (Brassicaceae, Euphorbiaceae, Leguminosae, Rosaceae, Violaceae). Taxonomic name standardization was indirectly ensured using data from The Chromosome Counts Database (CCDB, v 1.47: [ccdb.tau.ac.il/browse](http://ccdb.tau.ac.il/browse)), which is based on the automatic taxonomic name resolution software Taxonome (Kluyver and Osborne 2013) and the consensus database The Plant List (v 1.1: [www.plantlist.org](http://www.plantlist.org)). Species with both "accepted" and "unresolved" taxonomic status were included in the analysis: subspecies and varieties were discarded.

The diploid number of chromosomes for the sporophyte generation of each species was attained from the CCDB (last access: October 2020), as a quantitative proxy of ploidy level. Only one count per species was included in the analysis, except when different counts were equally reported in the database. When multiple counts were reported for a species, the modal value was retained; when a species exhibited multiple modal values, all of these were retained (e.g.  $2n=25$ ,  $2n=30$ ,  $2n=36$  for *Paphiopedilum victoria-*

*mariae*, Orchidaceae). Missing sporophyte values were obtained by doubling the gametophyte counts. B chromosomes were not considered. Negative values, as well as 0 and 1 were considered errors (being biologically improbable) and discarded. When possible, the source material used for the counting was checked: usually the mitotic counts were made using the root-tip squash method (Miller 1961), while meiotic counts were made from floral buds (reviewed by Windham et al. 2020). Chromosome counts for each taxon are presented in Table S1 (Online Resource 1).

The age for each taxon was obtained from the public database *Timetree: The Timescale of Life* (TTOF, [www.timetree.org](http://www.timetree.org); last access: December 2020). Molecular dating has been applied to an increasing number of species (the largest dated phylogenetic tree for the angiosperms comprises more than 36,000 species, belonging to ~8,400 genera, 426 families and all orders; Janssens et al. 2020), but heterogeneity in datasets, sequences, calibrations and the software used can yield different estimates for the same species, often hindering comparison between the results of different studies (Pulquério and Nichols 2007). TTOF provides a comprehensive synthesis of data published between 1987 and 2013 (2,274 studies) for 50,632 species, of which 14,465 angiosperms, and offers data uniformity, rapid data access and a robust foundation in the scientific literature (Hedges et al. 2006, 2015; Kumar and Hedges 2011; Kumar et al 2017). Divergence time between taxa is estimated through a hierarchical average linkage method; the maximum consistency with single original time-tree is ensured by testing and updating topological partitions (Hedges et al. 2015). A preliminary check of data included in the TTOF estimates was made from original chronograms in specific published papers, cited by TTOF. The discretional value of 0.001 Ma was attributed to extremely recent nodes, when a specific “estimated time” was not indicated (e.g. *Adenocarpus hispanicus* (Leguminosae), *Anemone hepatica*, *A. narcissiflora* (Ranunculaceae), *Magnolia coco*, *M. obovata* and *M. officinalis* (Magnoliaceae) (see Table S1 for estimated divergence times for all study species; Online Resource 1).

Here, we classify species as endemic on a case-by-case basis using a combination of quantitative data (geographic range limited to below a set threshold) guided by qualitative information such as designation as ‘endemic’ in national and regional floras. The geographical range for each species was obtained from public databases (last access: December 2020): Global Biodiversity Information Facilities (GBIF: <https://www.gbif.org/>), the Plants of the World Online portal ([www.plantsoftheworldonline.org](http://www.plantsoftheworldonline.org)), taxon-specific databases (i.e. the Global Compositae Database - GCD: [www.compositae.org](http://www.compositae.org); the Campanula portal: [campanula.e-taxonomy.net](http://campanula.e-taxonomy.net)) or specific papers (i.e. for Campanulaceae: Kandemir 2007 for *Campanula coriacea*; Crowl et al. 2015 for *C. delicatula*; for Compositae: Zhang et al. 2011 for the genera *Sorozeris*, *Stebbinsia* and *Syncalathium*). Species with a geographical range not exceeding 600,000 km<sup>2</sup> were classified as endemic. This threshold was chosen in order to include the remaining vegetation of the 36 Biodiversity Hotspots (see Table S2, Online Resource 2) according to Conservation International. Since the extension of some hotspots (i.e. Indo-Burma region, Brazil’s Cerrado, or Mediterranean Basin) exceeds 2,000,000 km<sup>2</sup>, and also includes urban areas, only the extension of the remaining vegetation was considered. Classification as endemic or not was also checked by determining whether species are recognised as endemic in national or local floras (for example: New Zealand Plant Conservation Network,

<http://www.nzpcn.org.nz>; Cellinese et al. 2009 for endemic Campanulaceae of Crete; Brochmann et al. 1997 for endemics of Cape Verde).

### ***Data analysis***

The coverage rate of the available data for each family was calculated as a percentage ratio between the number of species included in the analysis and the total number of species (both accepted and unresolved) reported in the consensus database The Plant List. Three separate analyses were performed both on the totality of data collected (referred to as “angiosperms”) and on subsets for single families, further sub-divided into analysis of all species (endemics and non-endemics), and endemics and non-endemics treated separately. For Nymphaeaceae only one analysis was performed, due to lack of data on endemic species, while for Magnoliaceae, Rosaceae, Saxifragaceae and Violaceae analysis of endemic species was not performed, due to insufficient data (10 spp. or less). Analyses were performed using the statistical software R (version 3.5.1; R Core Team 2018). Data were plotted according to the estimated time since divergence (x axis) and the number of chromosomes (y axis), using the *ggplot2* package (Wickham et al 2020).

In order to investigate the maximum number of chromosomes exhibited by taxa over geological time (time since taxon emergence), an upper boundary regression was applied to the data, fitting the regression curve only to the highest values of the data set. Boundary functions are widely used in ecological studies to highlight the maximum effect of a process, otherwise concealed by the weight of the mean values (Pierce 2014 and references therein). To remove the effects of redundant chromosome counts within each family, age values along the x axis were divided into periods (‘bins’) of 1 million years, and regression was fitted to the five highest y values within each bin. The function applied was an exponential decay, with the formula:  $y = Ae^{(-kx)} - c$ ; where: y = sporophyte number of chromosomes; A = initial quantity (estimated y value for x = 0); k = decay constant; x = estimated time since divergence; c = lowest y value for each family. Exponential decay was chosen for modelling the relationship between number of chromosomes and age, since polyploids originate from progressive duplications of the genome (Brysting et al. 2007). The c parameter was introduced in order to obtain a horizontal asymptote equal to the lower chromosome count and avoid curves tending to zero, as zero chromosomes would be biologically unrealistic.

Finally, the percentage ratio between the number of polyploids (*sensu* Wood et al. 2009) and the total number of species for each family included in the analysis was calculated in order to test whether taxonomic groups are differentially predisposed to polyploidy, in terms of both formation and establishment. All data are available in a Microsoft Excel spreadsheet format (Table S1; Online Resource 1).

## **Results**

Analyses were performed on a total number of 4530 records, corresponding to 4210 species, 1270 (30.2%) of which were found to be endemic. For each family, the coverage rate of collected data generally did not exceed 4% of taxa recorded for the family (see Table S3, Online Resource 3), with the exception of Ranunculaceae (5.8%), Apiaceae (6.2%), Magnoliaceae (9.1%) and Campanulaceae (12.6%). According to the estimated crown age (Table S3, Online Resource 3), the oldest families (>100 Ma) are Rosaceae (106 Ma, 95% C.I.: 41-161 Ma), Orchidaceae (105 Ma, 95% C.I.: 97-113 Ma) and Magnoliaceae (104 Ma, 95% C.I.: 95-113 Ma), while Brassicaceae (46 Ma, 95% C.I.: 19-71 Ma), Caryophyllaceae (52 Ma, 95% C.I.: 39-62 Ma) and Apiaceae (54 Ma, 95% C.I.: 29-57 Ma) are the most recent. However, uncertainty related to the estimated crown age was sometimes substantial: in Rosaceae, the extreme case, the confidence interval differed by 120 million years. Leguminosae (71-80 Ma), Violaceae (67-77 Ma) and Ranunculaceae (75-88 Ma) exhibited the least variable estimates. The Saxifragaceae family showed the greatest difference between the oldest species and the estimated crown age for the entire family (65.3 million years), while the lowest difference (20 million years) occurred in Leguminosae. Indeed, the oldest species included in the analysis (56 Ma) belong to the genus *Sophora*: *S. affinis*, *S. flavescens*, *S. microphylla* and *S. secundiflora*. Most data fell in the last 10 million years, with the exception of Magnoliaceae and Nymphaeaceae, which showed spikes in species divergence between 28.3 and 30 Ma, and at 11.3 Ma, respectively. In some families (Euphorbiaceae, Leguminosae, Orchidaceae, Poaceae, Primulaceae), available data for endemic species are relatively recent, compared to non-endemic species. However, only eight endemic species (*Helleborus orientalis*, *Lobelia physaloides*, *Magnolia guatemalensis*, *M. kobus*, *M. minor*, *M. salicifolia*, *Malus tschonoski*, *Pimpinella siifolia*), or approx. 0.6% of the total number, are older than 20 Ma. Recurrent numbers of chromosomes were evident in all families, which sometimes represent high percentages of data, for example  $2n=22$  in Apiaceae (58%),  $2n = 34$  in Campanulaceae (38.8%),  $2n = 38$  in Magnoliaceae (67.7%),  $2n = 32$  in Ranunculaceae (54.9%) and  $2n = 48$  in Violaceae (39%). Model parameters (i.e.  $k$  = decay constant;  $A$  = number of chromosomes estimated for age = 0 Ma) and statistics (i.e.  $R^2_{adj}$  and p-value) were extracted from model elaboration in the R environment. Hereafter, the adjusted  $R^2$  values for each general analysis (total species (T) representing endemics (E) plus non-endemics (N)) are indicated with  $R^2_{adjT}$ , while for each analysis performed on endemics and non-endemics separately variance is indicated with  $R^2_{adjE}$ , and  $R^2_{adjN}$ , respectively.

Upper boundary regression applied to all of the angiosperms in the dataset (Fig 1) showed a significant exponential decay trend ( $R^2_{adjT} = 0.48$ ;  $R^2_{adjE} = 0.46$ ;  $R^2_{adjN} = 0.44$ ; p always lower than 0.0001) between the number of chromosomes and the estimated time since divergence, which was three times steeper in endemic species ( $k = 0.12$ ) with respect to non-endemics ( $k = 0.04$ ). In this analysis, half-life for the decline in maximum number of chromosomes over time was 12.9, 6.2 and 19.8 Ma for angiosperms in general, and for endemic and non-endemics, respectively.

In analyses performed on single families, results were varied, but it was possible to distinguish common patterns. The overall negative relationship between the number of chromosomes and taxon age was confirmed in the majority of the families analysed, with differing degrees of significance. In Campanulaceae (Fig 2 a-c), Compositae (Fig 2 d-f), Leguminosae (Fig 2 g-i), Poaceae (Fig 2 j-l)

Caryophyllaceae (Fig S1 a-c, Online Resource 4), Rosaceae (Fig S1 d-f, Online Resource 4), the relationship was significant (e.g. for Campanulaceae:  $R^2_{adjT} = 0.42$ ;  $R^2_{adjE} = 0.22$ ;  $R^2_{adjN} = 0.32$ ; for Compositae:  $R^2_{adjT} = 0.46$ ;  $R^2_{adjE} = 0.44$ ;  $R^2_{adjN} = 0.24$ ; with  $p$  always lower than 0.0001), but similar exponential decay patterns were only evident as non-significant trends for endemic Ranunculaceae (Fig S1 g-i, Online Resource 4;  $R^2_{adjE}$  with a negative value and  $p = 0.376$ ) and Primulaceae (Fig S1 j-l, Online Resource 4;  $R^2_{adjE} = 0.065$ ;  $p = 0.116$ ). The relationship was stronger for endemics, as confirmed by the higher values of  $k$  (e.g. Compositae,  $k = 0.15$  and  $0.04$  for endemics vs. non-endemics, respectively; Caryophyllaceae,  $k = 0.21$  and  $0.15$  for endemics vs. non-endemics, respectively), and the entire family (e.g  $k = 0.11$  for Compositae, or  $k = 0.15$  for Caryophyllaceae). The negative trend was generally less significant for Apiaceae ( $R^2_{adjT} = 0.47$ ;  $R^2_{adjE} = 0.04$ ;  $R^2_{adjN} = 0.39$ ;  $p = 0.07$  for endemics, and lower than 0.0001 for the entire family and for non-endemics; Fig S2 a-c, Online Resource 4), Orchidaceae ( $R^2_{adjT} = 0.19$ ;  $R^2_{adjE} = -0.02$ ;  $R^2_{adjN} = 0.14$ ;  $p = 0.56$  in endemics and lower than 0.001 for the family and non-endemics; Fig S2 d-f, Online Resource 4), Violaceae ( $R^2_{adjT} = 0.05$ ;  $R^2_{adjN} = 0.03$ ;  $p$  always higher than 0.1; Fig S2 g-i, Online Resource 4) and Magnoliaceae (negative values for  $R^2_{adjT}$  and  $R^2_{adjN}$ ;  $p$  always higher than 0.6; Fig S2 j-l, Online Resource 4), and the slope for endemics was gentler ( $k = 0.04$  and  $k = 0.03$  for Apiaceae and Orchidaceae, respectively) than in the non-endemics ( $k = 0.12$  and  $k = 0.06$ , respectively) and in general for the family ( $k = 0.1$  and  $k = 0.08$ , respectively).

For Ericaceae, regressions were not significant ( $R^2_{adj}$  always lower than 0.01;  $p$  always higher than 0.1) and trends suggested variable and contrasting results (Fig 3): a non-significant exponential decay trend for chromosome counts with increasing taxon age was suggested for endemics ( $k = 0.05$ ). When non-endemics were taken into account, the regression exhibited a positive slope ( $k = -0.002$  at the family level, and  $k = -0.01$  for non-endemics).

In contrast, Brassicaceae (Fig 4) and Euphorbiaceae (Fig S3, Online Resource 4) exhibited statistically significant negative slopes in non-endemics and the family as a whole, while slopes for endemics were not significant (negative values for  $R^2_{adj}$  and  $p \sim 0.7$ ) with increasing tendencies with taxon age ( $k = -0.03$  and  $k = -0.009$ , respectively).

Finally, Nymphaeaceae (Fig 5) showed an overall, significant positive relationship, with  $k = -0.04$ ,  $R^2_{adjT} = 0.29$  and  $p = 0.03$ . A similar, but non-significant, tendency was shown in Saxifragaceae ( $k = -0.02$  for the family and  $k = -0.01$  for non-endemics, Fig S4 a-c, Online Resource 4), with negative values for  $R^2_{adj}$  and  $p$  always higher than 0.3.

The proportion of polyploid taxa within each family was found to differ substantially between families (Fig S5, Online Resource 4): polyploidy was evident for the majority of taxa in Violaceae (92%), Primulaceae (76%) and Campanulaceae (68%), while it occurred in less than 10% of species of the Leguminosae (8%) and Orchidaceae (6%).

## Discussion

The results broadly support the hypotheses that polyploidy is particularly evident in recently diverged angiosperms (hypothesis 1), especially for recent endemic species (hypothesis 2). While this phenomenon was evident for most of the families investigated, it was not always observed, and the hypothesis of a mechanism working consistently across the angiosperms (hypothesis 3) was only partially supported. The distribution of chromosome counts with time (i.e. towards recent ages) highlighted the pattern of progressive multiplication of the chromosome set, with high concentrations of records corresponding to diploid, tetraploid and hexaploid counts ( $2n=2x$ ,  $2n=4x$  and  $2n=6x$ , respectively): this is particularly evident in Apiaceae, Caryophyllaceae, Ranunculaceae and Rosaceae. Thus, polyploidization appears to be an important mechanism for the emergence of new species on a general level.

As confirmed by the steeper decay for the regression curves of endemic species, the relationship between high number of chromosomes and recent origin is especially strong for species with a limited distribution, and in particular Campanulaceae, Compositae, Leguminosae, Poaceae and Caryophyllaceae, which represent some of the largest and ecologically most important families. However, the considerable number of recent diploids, which necessitated the use of upper boundary regression to highlight the effects of polyploid counts, demonstrate that polyploidy is not an exclusive driver of the emergence of new species, either for endemics or non-endemics, and suggest that additional mechanisms are involved (e.g. geographic isolation, reproductive isolation, hybridization). This could also explain the weaker regression and the gentler slope for endemic, compared to non-endemic, Apiaceae and the general analysis.

However, the contrasting results for different families are evidence that phylogenetic effects operating within each family (revealed by the direction, range and variability of patterns) affect the tendency towards polyploidy. This is also supported by the different tendency for polyploidy evident for the families included in the analysis (Fig S5, Online Resource 4). In Ericaceae the regression analyses were not significant: it is likely that other prevailing mechanisms drive the emergence of new ericaceous species, for both endemics and non-endemics. Despite showing an overall negative tendency between the number of chromosomes and taxon age, a similar interpretation could explain the weak significance for Orchidaceae, also revealed by the highly variable 95% confidence intervals.

In Saxifragaceae, Magnoliaceae, Nymphaeaceae, Violaceae and perhaps Ericaceae, results were most likely affected by the limited availability of data. This could also explain the lower significance of the analyses for endemics and non-endemics with respect to the general family-level analysis in Primulaceae, and the atypical trend in endemic Euphorbiaceae and Brassicaceae, characterised by wide and irregular 95% confidence intervals. In particular, the estimated age for endemic Brassicaceae did not exceed 9 Ma, with only seven species older than 3 Ma (i.e. *Cochlearia aragonensis*, *Draba hederifolia*, *Streptanthus glandulosus*, *Vella asperum*, *V. bourgaeana*, *V. pseudocytisus*, *V. spinosa*). In these situations, the nature of the data requires careful consideration: in this example, four out of seven species belonging to the

same genus, *Vella*, are endemics of Spain and were dated relying essentially on a single study (i.e. Simón-Porcar et al. 2015), and thus on a single dating approach. The independence of observations could not be assured, since *V. asperum*, *V. bourgaeana* and *V. pseudocytisus* are closely related (Simón-Porcar et al. 2015; see also Siljak-Yakovlev and Peruzzi 2012). This is likely to have effects on species distribution: sharing the same habit and certain morphological traits, the three species were also found in the same habitat (xerophytic shrublands on gypsum substrate, characterised by anthropic disturbance) (Gómez Campo 1993; Simón-Porcar et al. 2015). With regard to karyotypes, the three species share the same basic chromosome number ( $x=17$ ), but while *V. bourgaeana* is diploid ( $2n=2x=34$ ), *V. pseudocytisus* is mainly tetraploid ( $2n=4x=68$ ), and *V. asperum* is hexaploid ( $2n=6x=102$ ) (Simón-Porcar et al. 2015), further confirming the role of polyploidy in speciation events. For these reasons, detailed here for a single prominent case, all analyses based on a restricted number of records should be interpreted with caution.

Additionally, the analyses that showed weaker regressions or positive trends were generally for older families, according to the age estimates provided by TTOF (Table S3, Online Resource 3): Orchidaceae (105 Ma), Magnoliaceae (104 Ma), Nymphaeaceae (89 Ma), Saxifragaceae (88 Ma), Ericaceae (82 Ma). This suggests that for certain ancient clades polyploidization may not be the main driver of speciation, although Rosaceae, the oldest clade (106 Ma), agreed with the hypothesis of higher polyploidy occurrence in recently diverged species. It is noticeable that some of these ancient families (i.e. Nymphaeaceae, Saxifragaceae, Rosaceae) showed high percentages of polyploid taxa, *sensu* Wood et al. 2009 (higher than 45%, Fig S5, Online Resource 4). In Saxifragaceae, a lack of relationship over time and high number of chromosomes in older taxa could indicate an initial burst of speciation *via* polyploidy followed by a lesser involvement of polyploidy in speciation. Therefore, polyploidy could be relatively widespread even across ancient families, but it is evidently not the main driver of speciation trends through time for these families.

Why recent polyploids are relegated to limited ranges is not immediately evident from our dataset. As confirmed by a body of research, polyploids are often adaptable species able to survive in harsh environmental contexts, and thus advantaged when colonizing new habitats; this tendency has been observed across the main angiosperm clades (Flovik 1940; Brochmann et al. 2004; Manzaneda et al. 2012; te Beest et al. 2012; Mas de Xaxars et al. 2016; Paule et al. 2018; Stevens et al. 2020). Intriguingly, the alteration of phenotype by polyploidy suggests that plant functioning and fitness may be fundamentally changed, but a preliminary classification of the species in our dataset according to Grime's CSR ecological strategies (method of Pierce et al. 2017) showed that no particular strategy class was associated with polyploidy: species adapted to survive competition, stress or disturbance all exhibited an extremely wide range of chromosome numbers (data not shown). Rather than reflecting fitness and adaptation, the high incidence of endemic polyploids could depend mainly on limited time for dispersal and colonizing new areas, as has been suggested for certain species (Behroozian et al. 2020). Indeed, it has been determined that polyploid plant species rely mainly on vegetative reproduction (Herben et al. 2017), which is usually ineffective for wide or rapid dispersal (Winkler and Fischer 2001; Herben et al. 2016). Indeed, the rearrangements of the chromosome set and the encumbrance created by multiple chromosomes in polyploid cells hinder the meiotic process and thus impose disadvantages for sexual

reproduction (Herben et al. 2017). On this premise, it is likely that neo-polyploids tend to remain limited to a narrow range (and, thus, apo-endemic) with respect to recently diverged diploids.

Increased availability of data regarding species molecular dating, chromosome counts and distributions will provide further support to this analysis. This particularly applies to families with less complete records (e.g. Ericaceae, Magnoliaceae, Nymphaeaceae), in order to determine whether contrasting results for these families are indicative of truly different patterns or are data artifacts. This is of particular concern for ancient clades, highlighting possible changes in prevailing speciation mechanisms throughout evolutionary time.

## Conclusion

The hypotheses were supported: chromosome duplication is more prevalent in recent angiosperms, in particular young endemics, confirming the role of polyploidy as a key driver of neo-endemism throughout the flowering plants. This pattern was generally, but not always, evident across flowering plant families, but some cases in which patterns were lacking may reflect insufficient data availability. Vegetative reproduction is a typical feature of polyploid taxa, suggesting that, despite a potentially greater adaptive capability, neo-polyploids tend to remain relegated to small ranges (compared to diploids) due to lower dispersal potential. However, the majority of young species (both endemics and non-endemics) are diploid, and thus polyploidy is not an exclusive driver of neo-endemism. Ancient clades (Orchidaceae, Magnoliaceae, Nymphaeaceae, Saxifragaceae and Ericaceae) exhibit weaker or contrasting trends, suggesting that genome doubling has become established as a key speciation mechanism mainly in more recent clades.

## Declarations

**Funding:** SV was supported by a Ph.D. scholarship from the Department of Agricultural and Environmental Sciences, University of Milan, as part of the 35<sup>th</sup> round of funding in the Doctoral School of Agriculture, Environment and Bioenergy (*Dottorato di Ricerca in Agricoltura, Ambiente e Bioenergia*).

**Conflicts of interest/Competing interests:** The authors declare that they have no conflict of interest.

**Availability of data and material:** The dataset is available in Microsoft Excel format as Table S1 (Online Resource 1).

**Code availability:** N/A

**Authors' contribution:** SV and SP conceived the study, SV collated data and performed the analyses with support from SP and MM. SV wrote the manuscript and all authors were involved in manuscript editing.

## Acknowledgments

We thank Dr. Diego Fontaneto (CNR, Verbania Pallanza, Italy) for helpful comments on the statistical analysis. The authors also thank Prof. Daniele Bassi and the Department of Agricultural and Environmental Sciences for administrative help and organisation of the postgraduate programme, which made this study possible.

## References

- Beck J, Böller M, Erhardt A, Schwanghart W (2014) Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol Inform* 19:10–15. <https://doi.org/10.1016/j.ecoinf.2013.11.002>
- Behroozian M, Ejtehadi H, Memariani F, Pierce S, Mesdaghi M (2020) Are endemic species necessarily ecological specialists? Functional variability and niche differentiation of two threatened *Dianthus* species in the montane steppes of northeastern Iran. *Sci Rep* 10:11774. <https://doi.org/10.1038/s41598-020-68618-7>
- Boakes EB, McGowan PJK, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol* 8(6):e1000385. <https://doi.org/10.1371/journal.pbio.1000385>
- Brochmann C, Rustan ØH, Lobin W, Kilian N (1997) The endemic vascular plants of the Cape Verde Islands, W Africa. *Sommerfeltia* 24:1-356
- Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen A-C, Elven R (2004) Polyploidy in arctic plants. *Biol J Linn Soc Lond* 82:521–536. <https://doi.org/10.1111/j.1095-8312.2004.00337.x>
- Brysting AK, Oxelman B, Huber KT, Moulton V, Brochmann C (2007) Untangling Complex Histories of Genome Mergings in High Polyploids. *Syst Biol* 56(3):467–476. <https://doi.org/10.1080/10635150701424553>
- Cañadas EM, Fenu G, Peñas J, Lorite J, Mattana E, Bacchetta G (2014) Hotspots within hotspots: Endemic plant richness, environmental drivers, and implications for conservation. *Biol Conserv* 170:282–291. <https://doi.org/10.1016/j.biocon.2013.12.007>
- Cellinese N, Smith SA, Edwards EJ, Kim S-T, Haberle RC, Avramakis M, Donoghue MJ (2009) Historical biogeography of the endemic Campanulaceae of Crete. *J Biogeogr* 36:1253–1269. <https://doi.org/10.1111/j.1365-2699.2008.02077.x>
- Coelho N, Gonçalves S, Romano A (2020) Endemic Plant Species Conservation: Biotechnological Approaches. *Plants* 9(3):345. <https://doi.org/10.3390/plants9030345>
- Crowl AA, Visger CJ, Mansion G, Hand R, Wu H-H, Kamari G, Phitos D, Cellinese N (2015). Evolution and biogeography of the endemic *Roucela* complex (Campanulaceae: *Campanula*) in the Eastern

Mediterranean. *Ecol Evol*, 5(22):5329–5343. <https://doi.org/10.1002/ece3.1791>

Doyle JJ, Egan AN (2010) Dating the origins of polyploidy events. *New Phytol* 186:73–85. <https://doi.org/10.1111/j.1469-8137.2009.03118.x>

Favarger C, Contandriopoulos J (1961) Essai sur l'endemisme. *Bulletin De La Société Botanique Suisse* 71:384-408 (in the French language)

Ferreira PMA, Boldrini II (2011) Potential reflection of distinct ecological units in plant endemism categories. *Conserv Biol* 25(4):672–679. <https://doi.org/10.1111/j.1523-1739.2011.01675.x>

Fithian W, Elith J, Hastie T, Keith DA (2014) Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods Ecol Evol* 6(4):424-438. <https://doi.org/10.1111/2041-210X.12242>

Flovik K (1940) Chromosome numbers and polyploidy within the flora of Spitzbergen. *Hereditas* 26(3-4):430-440. <https://doi.org/10.1111/j.1601-5223.1940.tb03246.x>

Gómez Campo C (1993) *Vella* L. In Castroviejo S et al. (eds.). *Flora Iberica* 4: 414-417. Real Jardín Botánico, CSIC, Madrid.

Grünig S, Fischer M, Parisod C (2021) Recent hybrid speciation at the origin of the narrow endemic *Pulmonaria helvetica*. *Ann Bot* 127:21–31. <https://doi.org/10.1093/aob/mcaa145>

Hedges SB, Dudley J, Kumar S (2006) TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971-2972. <https://doi.org/10.1093/bioinformatics/btl505>

Hedges SB, Marin J, Suleski M, Paymer M, Kumar S (2015) Tree of Life Reveals Clock-Like Speciation and Diversification. *Mol Biol Evol* 32(4):835–845. <https://doi.org/10.1093/molbev/msv037>

Herben T, Tackenberg O, Klimešová J (2016) Reproduction by seed and clonality in plants: correlated syndromes or independent strategies? *J Ecol* 104:1696–1706. <https://doi.org/10.1111/1365-2745.12646>

Herben T, Suda J, Klimešová J (2017) Polyploid species rely on vegetative reproduction more than diploids: a re-examination of the old hypothesis. *Ann Bot* 120(2):341–349. <https://doi.org/10.1093/aob/mcx009>

Husband BC, Sabara HA (2004) Reproductive isolation between autotetraploids and their diploid progenitors in fireweed, *Chamerion angustifolium* (Onagraceae). *New Phytol* 161(3):703-713. <https://www.jstor.org/stable/1514534>

Husband BC, Baldwin SJ, Suda J (2013) The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: Leitch IJ et al. (eds.) *Plant Genome Diversity Volume 2*, Springer-Verlag, Wien, pp 255-276

Janssens SB, Couvreur TLP, Mertens A, Dauby G, Dagallier L-PMJ, Abeele SM, Vandeloock F, Mascarello M, Beeckman H, Sosef M, Droissart V, van der Bank M, Maurin O, Hawthorne W, Marshall C, Réjou-Méchain M, Beina D, Baya F, Merckx V, Verstraete B, Hardy O (2020) A large-scale species level dated angiosperm phylogeny for evolutionary and ecological analyses. *Biodivers Data J* 8:e39677. <https://doi.org/10.3897/BDJ.8.e39677>

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and Angiosperms. *Nature* 473:97–100. <https://doi.org/10.1038/nature09916>  
Kandemir A (2007) A new *Campanula* (Campanulaceae) from east Anatolia, Turkey. *Nord J Bot* 25:53-57. [https://doi.org/10.1111/j.0107-055X.2007.00091\\_14.x](https://doi.org/10.1111/j.0107-055X.2007.00091_14.x)

Kluyver TA, Osborne CP (2013) Taxonome: A software package for linking biological species data. *Ecol Evol* 3(5):1262-1265. <https://doi.org/10.1002/ece3.529>

Kraft NJB, Baldwin BG, Ackerly DD (2010) Range size, taxon age and hotspots of neoendemism in the California flora. *Divers Distrib* 16:403–413. <https://doi.org/10.1111/j.1472-4642.2010.00640.x>

Kumar S, Hedges SB (2011) TimeTree2: species divergence times on the iPhone. *Bioinformatics* 27(14): 2023–2024. <https://doi.org/10.1093/bioinformatics/btr315>

Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* 34(7):1812–1819. <https://doi.org/10.1093/molbev/msx116>

Laport RG, Minckley RL, Ramsey J (2016) Ecological distributions, phenological isolation, and genetic structure in sympatric and parapatric populations of the *Larrea tridentata* polyploid complex. *Am J Bot* 103(7):1358-1374. <https://doi.org/10.3732/ajb.1600105>

Lavania UC (2013) Polyploidy, body size, and opportunities for genetic enhancement and fixation of heterozygosity in plants. *Nucleus* (April 2013) 56(1):1–6. <https://doi.org/10.1007/s13237-013-0075-7>

Lavania UC (2020) Plant speciation and polyploidy: in habitat divergence and environmental perspective. *Nucleus* 63:1–5. <https://doi.org/10.1007/s13237-020-00311-6>

Lazarina M, Kallimanis AS, Dimopoulos P, Psaralexi M, Michailidou D-E, Sgardelis SP (2019) Patterns and drivers of species richness and turnover of neo-endemic and palaeo-endemic vascular plants in a Mediterranean hotspot: the case of Crete, Greece. *Journal of Biological Research-Thessaloniki* 26,12. <https://doi.org/10.1186/s40709-019-0106-x>

Leebens-Mack JH, Barker MS et al (2019) One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679–685. <https://doi.org/10.1038/s41586-019-1693-2>

Levin DA (1983) Polyploidy and novelty in flowering plants. *Am Nat* 122(1):1-25. <https://www.jstor.org/stable/2461003>

- Maers AA, Giller P (2013) Analytical biogeography: an integrated approach to the study of animal and plant distributions. Springer, Dordrecht.
- Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman C, Raskin E, Mitchell-Olds T (2012) Environmental aridity is associated with cytotype segregation and polyploidy occurrence in *Brachypodium distachyon* (Poaceae). *New Phytol* 193(3):797–805. <https://doi.org/10.1111/j.1469-8137.2011.03988.x>.
- Mas de Xaxars G, Garnatje T, Pellicer J, Siljak-Yakovlev S, Vallès J, Garcia S (2016) Impact of dysploidy and polyploidy on the diversification of high mountain *Artemisia* (Asteraceae) and allies. *Alp Bot* 126:35–48. <https://doi.org/10.1007/s00035-015-0159-x>
- Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge.
- Miller WH (1961) Root tip cell squashes for the study of cell and chromosome morphology. *The American Biology Teacher*, 23(8):513-515. <https://doi.org/10.2307/4439736>
- Mishler BD, Knerr N, Gonzáles-Orozco CE, Thornhill AH, Laffan SW, Miller JT (2014) Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian *Acacia*. *Nat Commun* 5,4473. <https://doi.org/10.1038/ncomms5473>
- Noroozi J, Talebi A, Doostmohammadi M, Rumpf SB, Linder HP, Schneeweiss GM (2018) Hotspots within a global biodiversity hotspot - areas of endemism are associated with high mountain ranges. *Sci Rep* 8, 10345. <https://doi.org/10.1038/s41598-018-28504-9>
- Noss RF, Platt WJ, Sorrie BA, Weakley AS, Means DB, Costanza J, Peet RK (2015) How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. *Diversity Distrib.* 21: 236–244. <https://doi.org/10.1111/ddi.12278>
- Olivieri I, Tonnabel J, Ronce O, Mignot A (2015) Why evolution matters for species conservation: perspectives from three case studies of plant metapopulations. *Evol Appl* 9(1):196-211. <https://doi.org/10.1111/eva.12336>
- Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annu. Rev. Genet.* 34:401–437. <https://doi.org/10.1146/annurev.genet.34.1.401>
- Parisod C, Holderegger R, Brochmann C (2010) Evolutionary consequences of autopolyploidy. *New Phytol* 186:5–17. <https://doi.org/10.1111/j.1469-8137.2009.03142.x>
- Paule J, Dunkel FG, Schmidt M, Gregor T (2018) Climatic differentiation in polyploidy apomictic *Ranunculus auricomus* complex in Europe. *BMC Ecol* 18:16. <https://doi.org/10.1186/s12898-018-0172-1>
- Petrova G, Moyankova D, Nishii K, Forrest L, Tsiripidis I, Drouzas AD, Djilianov D, Möller M (2015) The European Paleoendemic *Haberlea rhodopensis* (Gesneriaceae) has an Oligocene origin and a Pleistocene

diversification and occurs in a long-persisting refugial area in Southeastern Europe. *Int J Plant Sci* 176(6):499-514. <https://doi.org/10.1086/681990>

Pierce S (2014) Implications for biodiversity conservation of the lack of consensus regarding the humped-back model of species richness and biomass production. *Funct Ecol* 28:253–257. <https://doi.org/10.1111/1365-2435.12147>

Pierce S, Negreiros D, Cerabolini BEL, Kattge J, Díaz S, Kleyer M, Shipley B, Wright SJ, Soudzilovskaia NA, Onipchenko VG, van Bodegom PM, Frenette-Dussault C, Weiher E, Pinho BX, Cornelissen JHC, Grime JP, Thompson K, Hunt R, Wilson PJ, Buffa G, Nyakunga OC, Reich PB, Caccianiga M, Mangili F, Ceriani RM, Luzzaro A, Brusa G, Siefert A, Barbosa NPU, Chapin III FS, Cornwell WK, Fang J, Fernandes GW, Garnier E, Le Stradic S, Peñuelas J, Melo FPL, Slaviero A, Tabarelli M, Tampucci D (2017) A global method for calculating plant CSR ecological strategies applied across biomes worldwide. *Funct Ecol* 31(2):444-457. <https://doi.org/10.1111/1365-2435.12722>

Pulquério MJF, Nichols RA (2007) Dates from the molecular clock: how wrong can we be? *Trends Ecol Evol* 22(4):180-184. <https://doi.org/10.1016/j.tree.2006.11.013>

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol* 20,38. <https://doi.org/10.1186/s13059-019-1650-2>

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploidy formation in flowering plants. *Annu Rev Ecol Syst* 29:467-501. <https://doi.org/10.1146/annurev.ecolsys.29.1.467>

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>

Rezende L, Suzigan J, Amorim FW, Moraes AP (2020) Can plant hybridization and polyploidy lead to pollinator shift? *Acta Botanica Brasilica* 34(2): 229-242. <https://doi:10.1590/0102-33062020abb0025>

Schubert I, Lysak MA (2011) Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends in Genetics* 27(6):207-216. <https://doi.org/10.1016/j.tig.2011.03.004>

Siljak-Yakovlev S, Peruzzi L (2012) Cytogenetic characterization of endemics: past and future. *Plant Biosystems* 146(3): 694–702.

Simón-Porcar VI, Pérez-Collazos E, Catalán P (2015) Phylogeny and systematics of the western Mediterranean *Vella pseudocytisus*-*V. aspera* complex (Brassicaceae). *Turk J Bot* 39:472-486. <https://doi.org/10.3906/bot-1406-87>

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and Angiosperm diversification. *Am J Bot* 96(1):336–348. <https://doi.org/10.3732/ajb.0800079>

- Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then...and now: Stebbins revisited. *Am J Bot* 101(7):1057–1078. <https://doi.org/10.3732/ajb.1400178>
- Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in Angiosperms. *Current Opinion in Plant Biology* 30:159–165. <https://doi.org/10.1016/j.pbi.2016.03.015>
- Stebbins GL, Major J (1965) Endemism and Speciation in the California Flora. *Ecol Monogr* 35(1):1-35. <https://doi.org/10.2307/1942216>
- Stevens AV, Nicotra AB, Godfree RC, Guja LK (2020) Polyploidy affects the seed, dormancy and seedling characteristics of a perennial grass, conferring an advantage in stressful climates. *Plant Biol* 22:500–513. <https://doi.org/10.1111/plb.13094>
- Tamayo-Ordóñez MC, Espinosa-Barrera LA, Tamayo-Ordóñez YJ, Ayil-Gutiérrez B, Sánchez-Teyer LF (2016) Advances and perspectives in the generation of polyploid plant species. *Euphytica* 209:1–22. <https://doi.org/10.1007/s10681-016-1646-x>
- te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P (2012) The more the better? The role of polyploidy in facilitating plant invasions. *Ann Bot* 109:19–45. <https://doi.org/10.1093/aob/mcr277>
- Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of polyploidy. *Nat Rev Genet* 18:411–424. <https://doi.org/10.1038/nrg.2017.26>
- Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D, RStudio (2020) Create elegant data visualisations using the grammar of graphics. R package version 3.3.3
- Williams KJ, Ford A, Rosauer DF, De Silva N, Mittermeier R, Bruce C, Larsen FW, Margules C (2011) Forests of East Australia: The 35th Biodiversity Hotspot. In: Zachos FE, Habel JC (Eds.) *Biodiversity Hotspots. Distribution and protection of conservation priority areas*. Springer-Verlag, Berlin, Heidelberg, pp 295-310
- Williams SE, Williams YM, Van Der Wal J, Isaac JL, Shoo LP, Johnson CN (2009) Ecological specialization and population size in a biodiversity hotspot: how rare species avoid extinction. *PNAS* 106(2):19737–19741. <https://doi.org/10.1073/pnas.0901640106>
- Windham MD, Pryer KM, Poindexter DB, Li F-W, Rothfels CJ, Beck JB. (2020) A step-by-step protocol for meiotic chromosome counts in flowering plants: A powerful and economical technique revisited. *Applications in Plant Sciences*, 8(4): e11342. <https://doi.org/10.1002/aps3.11342>
- Winkler E, Fischer M (2001) The role of vegetative spread and seed dispersal for optimal life histories of clonal plants: a simulation study. *Evol Ecol* 15:281–301. <https://doi.org/10.1023/A:1016080714200>

Wood TE, Takebayashic N, Barker, MS, Mayrose I, Greenspoond PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. PNAS 106(33):13875–13879. <https://doi.org/10.1073/pnas.0811575106>.

Zhang J-W, Nie Z-L, Wen J & Sun H (2011). Molecular phylogeny and biogeography of three closely related genera, *Soro-seris*, *Stebbinsia*, and *Syncalathium* (Asteraceae, Cichorieae), endemic to the Tibetan Plateau, SW China. Taxon 60(1):15–26. <https://doi.org/10.1002/tax.601003>

Zozomová-Lihová J, Krak K, Mandáková T, Shimizu KK, Španiel S, Vít P, Lysak MA (2014) Multiple hybridization events in *Cardamine* (Brassicaceae) during the last 150 years: revisiting a textbook example of neoallopolyploidy. Ann Bot 113(5):817–830. <https://doi.org/10.1093/aob/mcu012>

## Figures

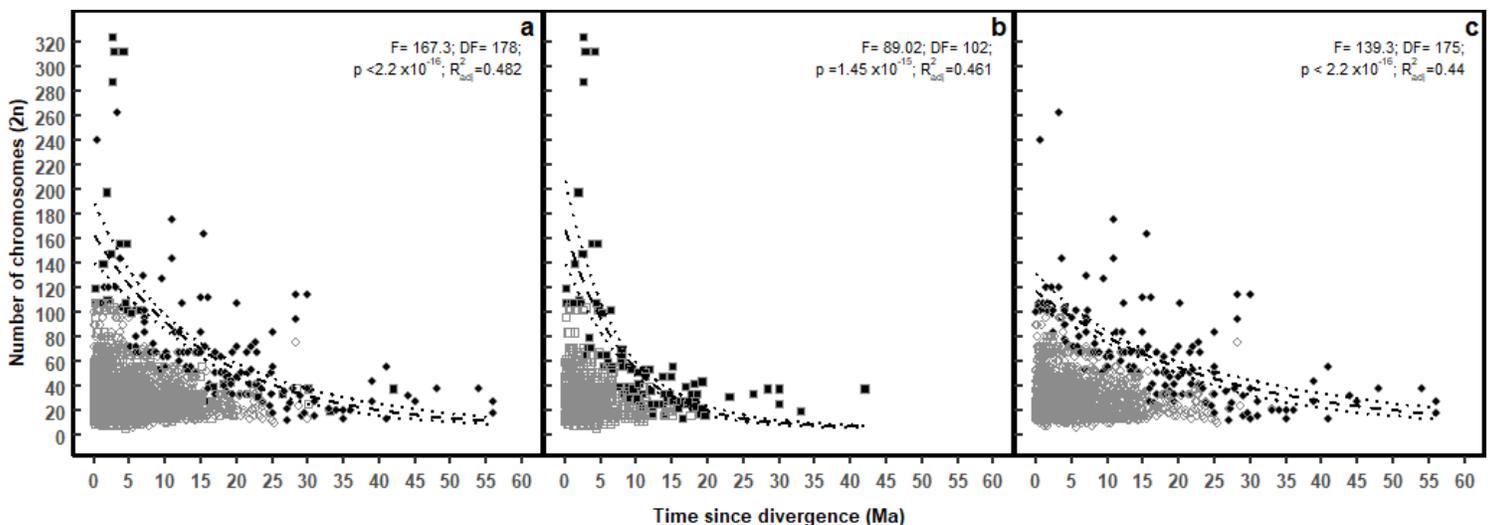
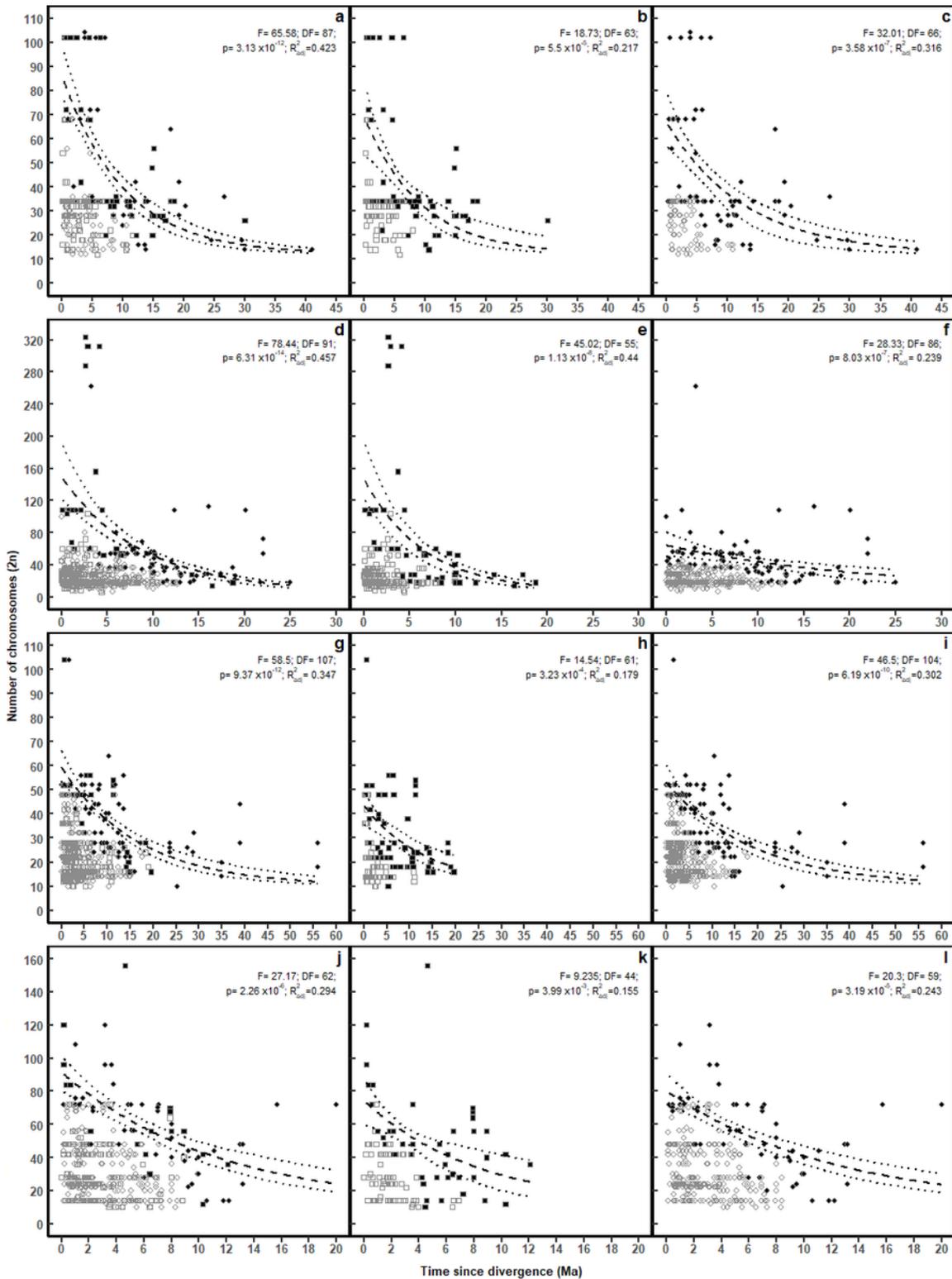


Figure 1

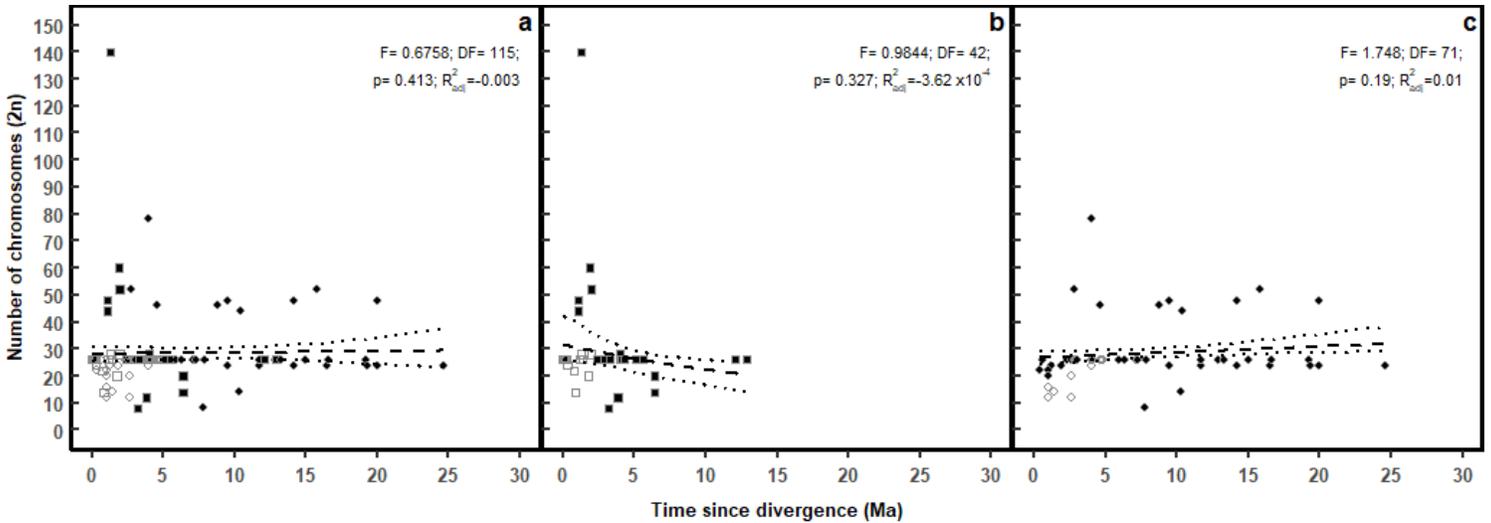
The relationship between time since taxon divergence and number of chromosomes (as a proxy for ploidy level) across all major clades of the Angiosperms, applied to: a) the entire dataset; b) endemic species; c) non-endemic species. Squares represent endemic species, circles represent non-endemic species. Broken lines represent ‘upper boundary regressions’, or 3-parameter Lorentzian regressions fitted to the five highest values in each ‘bin’ or 1 million year interval (bin value data points are filled in dark grey; points under the upper boundary curve are unfilled). Dotted lines represent the  $\pm 95\%$  confidence interval



**Figure 2**

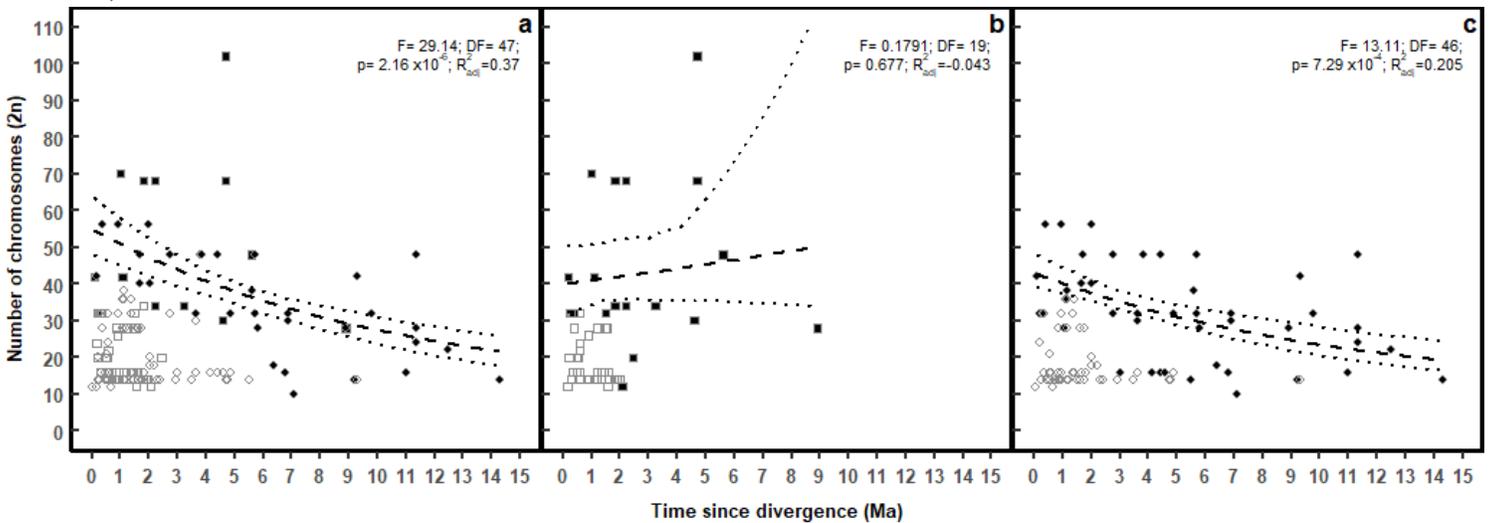
The relationship between time since taxon divergence and number of chromosomes for examples of angiosperm families exhibiting a declining upper boundary relationship: a-c) Campanulaceae, all spp., endemic spp. and non-endemic spp., respectively; d-f) Compositae, g-i) Leguminosae; j-l) Poaceae. Squares represent endemic species, circles represent non-endemic species. Broken lines represent 'upper boundary regressions', or 3-parameter Lorentzian regressions fitted to the five highest values in each 'bin'

or 1 million year interval (bin value data points are filled in dark grey; points under the upper boundary curve are unfilled). Dotted lines =  $\pm 95\%$  C.I. Note that x and y data ranges are different for each family



**Figure 3**

Non significant trends in the family of Ericaceae, with a positive slope in a) the analysis of the whole family and c) non-endemic species only, and a negative slope in b) endemics. Squares represent endemic species, circles represent non-endemic species. Broken lines represent 'upper boundary regressions' fitted to the five highest values in each 'bin' (points filled in dark grey; points under the upper boundary curve are unfilled). Dotted lines =  $\pm 95\%$  C.I.



**Figure 4**

Significant negative relationship between the number of chromosomes and estimated taxon age in a) the entire family of Brassicaceae and c) non-endemic Brassicaceae; b) non significant positive trend in endemic Brassicaceae. Squares represent endemic species, circles represent non-endemic species. Broken lines represent 'upper boundary regressions' fitted to the five highest values in each 'bin' (points filled in dark grey; points under the upper boundary curve are unfilled). Dotted lines =  $\pm 95\%$  C.I.

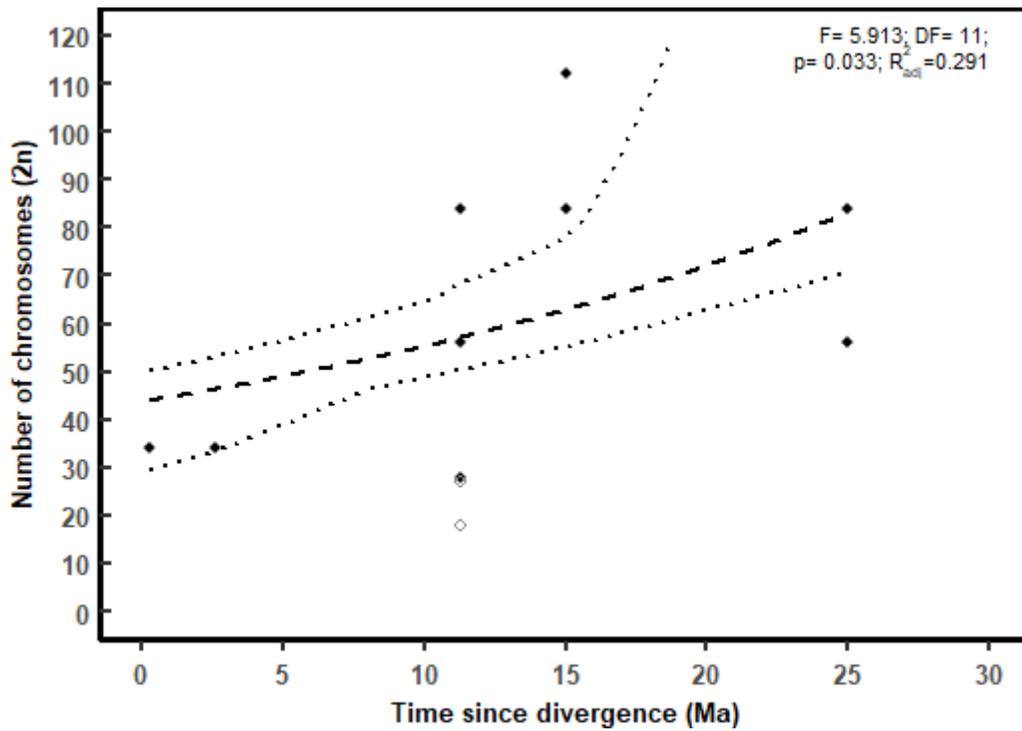


Figure 5

The weakly significant positive trend between taxon age and chromosome number in Nymphaeaceae. Broken lines represent 'upper boundary regressions' fitted to the five highest values in each 'bin' (points filled in dark grey; points under the upper boundary curve are unfilled). Dotted lines =  $\pm 95\%$  C.I.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfigures.pdf](#)
- [TableS1.xlsx](#)
- [TableS2.pdf](#)
- [TableS3.pdf](#)