Karyotypic diversity: a neglected trait to explain angiosperm diversification?

Angelino Carta1, and Marcial Escudero2,

1Department of Biology, Botany Unit, University of Pisa, Pisa, Italy

2Department of Plant Biology and Ecology, University of Seville, Seville, Spain

Corresponding author: Department of Biology, Botany Unit, University of Pisa, Pisa, Italy. Email: angelino.carta@unipi.it

Abstract

Evolutionary changes in karyotype provides genetic support to organisms' differentiation and adaptation; however, the association between karyotype diversity and species diversification in flowering plants (angiosperms) remains to be fully elucidated. We sought evidence for this association within a phylogenetic framework using a dataset comprising > 413,000 world-wide chromosome counts of 66,000 angiosperms species. Karyotypic diversity (KD; e.g., number of distinct chromosome numbers) explains species richness and diversification rates at both family and genus levels highlighting that chromosome evolution has probably played, at least, an important role in reinforcing speciation that was already initiated or completed by other geographical or ecological drivers. Thus, research programmes investigating chromosome variation as direct or indirect driver of diversification should be encouraged.

Keywords: angiosperms, chromosome evolution, chromosome number, diversification, species richness

Introduction

The plant nuclear genome is organized into discrete chromosomes, whose count (chromosome number) is the primary information for the karyotype description of an organism (Heslop-Harrison & Schwarzacher, 2011). Whilst karyotype constancy ensures gene transfer to the next generation, karyotype variation provides genetic support to organisms' differentiation and adaptation (Stebbins, 1971). Chromosome numbers are extraordinarily variable in flowering plants (angiosperms), ranging from $n = 2$ to 320 (Rice et al., 2015; Carta et al., 2020) and with many genera and families exhibiting an order of magnitude or more of variation in chromosome number (Rice et al., 2015; Fig. 1). As such, chromosome number is a dynamic feature of plant evolution, especially among the angiosperms (Stebbins, 1971). Changes in chromosome number occur via dysploidy (single chromosome number changes; Escudero et al., 2014) or polyploidy (whole genome duplication, WGD; Soltis et al., 2015). These mechanisms of chromosomal evolution have contributed to the rich karyotypic diversity in

angiosperms (Escudero et al., 2014) and may produce strong reproductive barriers (through polyploidy or dysploidy changes) or changes in recombination patterns (through dysploidy changes), potentially leading to speciation and lineage diversification (Levin & Wilson, 1976; Soltis et al., 2009; Wood et al., 2009; Zhan et al., 2021). Additionally, although chromosomal changes may not directly give rise to speciation, they might still be crucial in the speciation process by reinforcing speciation that has been triggered by other geographical or ecological factors (Rieseberg, 2001; Rieseberg & Willis, 2007).

The effects of chromosome mutations and karyotypic rearrangements in the diversification of plants are controversial (Kellogg, 2016). Particularly, chromosome number transitions through polyploidy and dysploidy, and their impact in diversification have been widely disputed (Soltis et al., 2009; Mayrose et al., 2011; Arrigo & Barker, 2012, but see also Soltis et al., 2014; Escudero et al., 2014, 2018; Pimentel et al., 2017; Freyman & Höhna 2018). Given the potential interplay between karyotypic and other phenotypic features (Zenil-Ferguson et al., 2017, 2019; Fumia et al., 2022), these contrasting evidences hamper plant science and are in deep contrast with the central role that karyological research still has in new species descriptions, floras, and management plans for threatened species (Pires & Hertweck, 2008).

Here, we explored variation in karyotypic diversity (KD) across angiosperms to examine the association between KD with lineage diversification at both family and genus levels. To this end, we combined chromosome numbers from the Chromosome Count Database (CCDB) to assemble data sets for 328 families and 7817 genera of angiosperms (78% extant families and 60% genera). For each taxonomic rank, we calculated KD (number of distinct chromosome numbers; Fig. 1) and estimated net diversification rates using the Magallón and Sanderson method (2001) or provided by Smith and Brown (2018) based on the most comprehensive time-calibrated molecular phylogeny for vascular plants published hitherto (but see also Jin & Qian, 2019). We then used phylogenetic regressions and phylogenetic path analyses to assess the extent to which KD is associated with diversification rates and whether it is also related with clade-wise age (Myr) or species richness (SR). All analyses were conducted at the family and genus levels in a phylogenetic comparative framework using a recently published time-calibrated phylogenetic tree as reference (Ramírez-Barahona et al., 2020) but analyses were also complemented with other recently published angiosperm trees (Smith & Brown, 2018; Li et al., 2019).

Methods

Chromosome counts, plant nomenclature and species richness

The haploid chromosome numbers (n) of the species were obtained from the CCDB (Rice et al., 2015; http://ccdb.tau.ac.il/) using the R package chromer (Pennell, 2016). The CCDB contains records from original sources that have multiple complex symbol patterns denoting multivalence, irregularities of chromosome counts, so that the > 413.000 records were curated semi-automatically using the CCDBcurator package which was designed to provide reproducible means of cleaning and curating chromosome count data (Rivero et al., 2019). Nevertheless, to further account for uncertainty in the curated chromosome count data, we also adopted different alternative analyses considering intraspecific (all available counts per species) vs non-intraspecific variation (see below).

All plant names were standardized against The Plant List (TPL). Trials analyses using Plants of the World Online (POWO, 2021) returned congruent results with independence of the chosen taxonomic database. We are aware that World Flora Online (WFO) is available but this framework is still not yet fully implemented. Indeed, using the taxonlookup package (Pennell et al., 2016), we also gathered the number of accepted species for each family and genus as an estimate of species richness (SR) and to calculate the proportion of sampled species against the number of extant species for each family and genus (see below).

For each family and genus matching the phylogenetic trees (see below), we calculated two alternative measures of karyotypic diversity (KD): (A) selecting all available counts per species and then counting the number of unique distinct chromosome numbers found in each taxonomic rank (either families or genera), (B) selecting for each species the lowest chromosome number available and then counting the number of distinct chromosome numbers found in each taxonomic rank (either families or genera). The first measure accounts for intraspecific variation (e.g., intraspecific dysploidy or polyploidy) as part of KD; the second approach instead explicitly remove the effect of intraspecific variation. An alternative would be to consider the modal chromosome number per species (Glick & Mayrose, 2014) but our trial analyses do not detected any difference from using the lowest number available.

Phylogenetic information, clade ages and diversification rates

The recently published time-calibrated phylogenetic tree RC_complete_MCCv_2 was used as reference (Ramírez-Barahona et al., 2020). This tree was prepared using 238 fossils relaxed calibration (RC) strategy (Ramírez-Barahona et al., 2020). We then either pruned this tree at the family or genus level, matching 328 families or 639 genera respectively with chromosome counts data and including more than 1 extant species (e.g., we excluded all clades with diversification rates = 0).

We also explored the sensitivity of our results by conducting all analyses at the genus level again using two additional time-calibrated trees extracted from published alternative angiosperm trees: the PPA tree providing a substantially older crown age for angiosperms (Li et al., 2019) and the GBOTB.extended phylogenetic tree (Smith & Brown; 2018; the genus-level of which was provided by Molina-Venegas and Lima, 2021). When using the former, 1059 genera were represented, while for the latter, 3223 genera matched the chromosome data. The genus level analyses allowed us also to explore patterns at a smaller evolutionary scale, closer to species level.

Extending our analyses at the genus level could introduce some noise due to the non-monophyly in some genera. Nevertheless, this is a minor, probably non-significant pattern because the comparative approach used here accounts that genera are not independent data points in statistical analysis due to their shared evolutionary history. We did not use the phylogeny to reconstruct chromosome numbers nor their evolutionary changes (see Zhan et al. 2021) which would be probably more affected by the non-monophyly of genera. Moreover, in the most comprehensive phylogeny in terms of genus sampling in our analyses (Smith and Brown, 2018), the non-monophyly of the genera has been documented as very significant only in a few lineages of the phylogeny (see comments in Smith and Brown, 2018). Finally, to further keep minimal its effect, Molina-Venegas and Lima (2021) kept a representative species at random from the largest monophyletic cluster of each genus in GBOTB.extended (the tree we also used), and in the few cases where multiple monophyletic clusters of equal size were eligible, they first selected one of the clusters at random and then picked one representative species.

Clade-wise ages were extracted by considering the stem age because the crown is not always well identifiable, especially for genera/families with few species. Ages were used either as predictor in the regression models and phylogenetic path analyses or used to estimate net diversification rates following Magallón and Sanderson method (2001) assuming two extremes of the relative extinction rate (epsilon = 0, no extinction; and epsilon = 0.9, high rate of extinction; where epsilon = extinction rate/speciation rate) as implemented in the R package geiger (Pennell et al., 2014). We also used diversification rate estimates provided by Smith & Brown (2018) who applied the diversification rate-shifts MEDUSA algorithm.

Statistical analyses

The relationships between KD and the response variables were tested using phylogenetic generalized least-squares (pgls) in the ape and caper packages (Paradis et al., 2004; Orme et al., 2013) of R (R

Development Core Team, 2022). Weighted regressions accounting for different sampling effort in each clade (~1/sampling effort) were fitted by maximum likelihood and using Pagel's λ (Pagel, 1999; Freckleton et al., 2002) as a measure of the phylogenetic signal. All variables were $\log_{10}$ transformed before the analyses.

The complex relationship between KD, diversification rates, species richness and clade age was further explored with phylogenetic path analyses (Hardenberg and González-Voyer, 2013) using the R package phylopath (van der Bijl, 2018). We compared the null model (diversification rates are explained by species richness and clade age; KD is not included in the model) with other more complex models in which (i) diversification rates are also explained by KD, (ii) species richness is explained by KD, and (iii) KD explains both diversification rates and species richness. We would expect that the null model is the best model only if KD does not play a role at all as a direct driver of speciation (or indirectly reinforcing speciation process). We might expect that KD mainly predicts diversification rates if KD is a direct driver of speciation/diversification (in this case, KD would only predict species richness if species richness also predicts diversification rates, and we would expect approximately the same strength of relation between KD and species richness and between species richness and diversification). Alternatively, we might expect that KD mainly predicts species richness if KD does not drive speciation directly and only reinforces the speciation process (in this case, KD would only predict diversification rates if species richness also predicts diversification rates, and we expect approximately the same strength of relation between KD and species richness and between species richness and diversification). Finally, we might expect that KD explains both species richness and diversification rates, if KD drives and reinforces speciation. We acknowledged that multiple causal drivers, not included in our models, could be also potentially responsible for the observed patterns in our tested models. All variables were scaled into a common measurement scale so that their effects in the equations can be compared. Data were visualised on the phylogenetic tree using the package ggtree (Yu et al., 2017).

Results and Discussion

KD is randomly distributed across angiosperm families (Fig. 1; Table S1), with phylogenetic signal (λ=0) not significantly different from zero (λ=0, P=1) but significantly different from one (λ=1, P = 0.000), suggesting weak phylogenetic constraint on the number of distinct chromosome numbers within each family (and genus; Table S2). On the contrary, basic (monoploid chromosome number) and modal

chromosome numbers (the most frequent number in a taxon) are probably strongly phylogenetically clustered (Escudero et al., 2012; Carta et al., 2018).

Phylogenetic least square regressions suggest that KD is positively related to diversification rates across all angiosperm families ($R_2 = 0.56$; $\beta = 0.44$, $P < 0.001$) and this relation was consistent within major angiosperm clades (Fig. 2; Table 1). KD was also positively related with SR ($R_2 = 0.72$; $\beta = 1.45$, $P < 0.001$) but not with clade age ($R_2 = 0.59$; $\beta = 0.0$, $P = 1$). Our results from phylopath analyses suggest that the best model is the most complex in which KD is the best predictor of both, diversification rates (0.33-0.64) and species richness (0.66) (Fig. 3).

When analyzing the data at the genus level, we found very similar results; albeit model fit accounted for slightly less variation (see Table S3). In fact, rates of increase in species diversity and rates of chromosomal evolution were already reported to be strongly correlated in plants (Levin & Wilson, 1976; these rates were calculated dividing species richness and karyotype diversity per unit of time estimated from fossil ages and using a very reduced sampling) and in mammals (Martinez et al., 2017).

We explored the sensitivity of our results by conducting all analyses at the genus level again using two additional time-calibrated trees extracted from published alternative angiosperm trees: the PPA tree providing a substantially older crown age for angiosperms (Li et al., 2019) and the GBOTB.extended phylogenetic tree (Smith & Brown; 2018; the genus-level of which provided by Molina-Venegas & Lima, 2021). Overall, the findings obtained using these two alternative trees are consistent in highlighting common KD relationships with diversification rates and species richness (Table S4-5), in agreement with the results obtained using our reference tree. However, our reference tree (RC_complete_MCCv_2; Ramírez-Barahona et al., 2020) was compiled using the largest number of fossil calibrations compared to other available trees; the different and more accurate time-calibration of our reference tree is likely the reason why it is best in explaining diversification compared to the other two, alternative phylogenetic trees.

Overall, our results presented above suggest that KD is positively associated to diversification rates regardless of the considered diversification estimate. However, the strength of this association varies depending on the diversification estimate used in our analyses (Table 1, Table S3-S5), with stronger association when using diversification rate estimated applying the Magallón & Sanderson method (2001) and weaker association when using diversification estimates obtained from diversification rate-shift reconstructions provided by Smith & Brown (2018).

Karyotypic variation can be measured as the number of distinct numbers per taxon, or considering the variation in basic chromosome number (Pimentel et al., 2017; Carta et al., 2018). As such, the

calculation of KD is only apparently straightforward because different approaches can be used. In this study we first considered counting the number of unique distinct chromosome numbers found in each taxonomic rank (either families or genera; see also Levin & Wilson, 1976; Martinez et al., 2017; Herrick & Sclavi, 2019). This approach accounts for polymorphic chromosome numbers in a given species (e.g., intraspecific variation due to polyploidy or dysploidy events). Alternatively, before calculating KD we selected for each species the lowest chromosome number available. These two alternative coding schemes allowed us to deal for uncertainty in the curated chromosome count data and with the problem of the existence of diffeent ploidy levels or dysploidy events in a species. Since the results are congruent (Table S6-S9), taking into account intraspecific variation does not affect the direct relationships between KD and SR or diversification rates. We thus conclude that diversification is probably associated with genomic rearrangements independent from genome duplications or dysploidy events within species.

Here we modelled diversification rates and species richness as dependent on chromosome variation, and our results confirm our expectation. The patterns found here are very similar across major clades, suggesting that, whilst at a finer scale (below genus rank) the patterns may change drastically, at higher rank level as those explored here, chromosome number-based diversification seems driven in a similar way across different lineages.

Whilst our results demonstrate a positive and direct relation between KD, SR and diversification, multiple causal drivers could be responsible for the observed patterns (e.g., Zenil-Ferguson et al., 2017, 2019; Fumia et al., 2022). Our results suggest that chromosome evolution has probably played, at least, an important role in reinforcing speciation that was already initiated or completed by other geographical or ecological drivers. In conclusion, we argue that the positive relationship between KD and diversification rates might suggest that reproductive barriers and / or changes in recombination patterns as results of chromosome rearrangement lead to speciation and lineage diversification in angiosperms, or at least they reinforce the speciation process. Thus, as angiosperms show large diversity in their reproductive systems (Bennet et al., 2022), research programmes investigating chromosome variation as direct or indirect driver of diversification should be encouraged.

Data Accessibility

Code and data supporting the results can be accessed at the GitHub repository https://github.com/angelinocarta/CNvar. A version of record of the recorded repository can be found at https://zenodo.org/record/7573459#.Y9KXY3bMI2w.

References

Arrigo N, Barker MS. 2012. Rarely successful polyploids and their legacy in plant genomes. Current opinion in plant biology 15: 140-146.

Benton M.J., Wilf P., Sauquet H. 2022. The Angiosperm Terrestrial Revolution and the origins of modern biodiversity. New Phytologist, 233: 2017-2035.

Carta A, Bedini G, Peruzzi L. 2018. Unscrambling phylogenetic effects and ecological determinants of chromosome number in major angiosperm clades. Scientific reports 8: 1-14.

Carta A, Bedini G, Peruzzi L. 2020. A deep dive into the ancestral chromosome number and genome size of flowering plants. New Phytologist, 228: 1097-1106.

Escudero M, Hipp AL, Hansen TF, Voje KL, Luceño M. 2012. Selection and inertia in the evolution of holocentric chromosomes in sedges (Carex, Cyperaceae). New Phytologist 195: 237-247.

Escudero M, Martín-Bravo S, Mayrose I, Fernández-Mazuecos M, Fiz-Palacios O, Hipp AL, ... Luceño M. 2014. Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. PloS one 9: e85266.

Escudero, M., Balao, F., Martín-Bravo, S., Valente, L., & Valcárcel, V. (2018). Is the diversification of Mediterranean Basin plant lineages coupled to karyotypic changes?. Plant Biology, 20, 166-175.

Freckleton RP. 2002. On the misuse of residuals in ecology: regression of residuals vs. multiple regression. Journal of Animal Ecology 542-545.

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

Fumia N, Rubinoff D, Zenil-Ferguson R, Khoury CK, Pironon S, Gore MA, Kantar MB. 2022. Interactions between breeding system and ploidy affect niche breadth in Solanum. Royal Society Open Science 9: 211862.

Glick L, Mayrose I. 2014. ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Molecular biology and evolution 31: 1914-1922.

Hardenberg, A. V., & Gonzalez-Voyer, A. (2013). Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. Evolution: International Journal of Organic Evolution, 67(2), 378-387.

Herrick, J, Sclavi B. 2019. Genome diversity and species richness in mammals. bioRxiv, 709311.

Heslop-Harrison JS, Schwarzacher T. 2011. Organisation of the plant genome in chromosomes. The Plant Journal 66: 18-33.

Jin Y, Qian H. 2019. V. PhyloMaker: an R package that can generate very large phylogenies for vascular plants. Ecography 42: 1353-1359.

Kellogg EA. 2016. Has the connection between polyploidy and diversification actually been tested?. Current opinion in plant biology 30: 25-32.

Levin DA, Wilson AC. 1976. Rates of evolution in seed plants: net increase in diversity of chromosome numbers and species numbers through time. Proceedings of the National Academy of sciences 73: 2086-2090.

Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, ... & Li DZ. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. Nature plants 5: 461-470.

Magallon S, & Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. Evolution 55: 1762-1780.

Martinez PA, Jacobina UP, Fernandes, RV, Brito C, Penone C, Amado TF, ... & Bidau CJ. 2017. A comparative study on karyotypic diversification rate in mammals. Heredity 118: 366-373.

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011. Recently formed polyploid plants diversify at lower rates. Science 333: 1257-1257.

Molina-Venegas R, Lima H. 2021. Should we be concerned about incomplete taxon sampling when assessing the evolutionary history of regional biotas?. Journal of Biogeography 48: 2387-2390.

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W. 2013. The caper package: comparative analysis of phylogenetics and evolution in R. R package version 5: 1-36.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature 401: 877-884.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.

Pennell M. 2016. chromer: Interface to Chromosome Counts Database API. R package version 0.1. 2.9000.

Pennell MW, FitzJohn RG, Cornwell WK. 2016. A simple approach for maximizing the overlap of phylogenetic and comparative data. Methods in Ecology and Evolution 7: 751-758.

Pimentel M, Escudero M, Sahuquillo E, Minaya MÁ, Catalán P. 2017. Are diversification rates and chromosome evolution in the temperate grasses (Pooideae) associated with major environmental changes in the Oligocene-Miocene?. PeerJ 5: e3815.

Pires JC, Hertweck KL. 2008. A Renaissance of Cytogenetics: Studies in Polyploidy and Chromosomal Evolution1. Annals of the Missouri Botanical Garden 95: 275-281.

POWO. 2021. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew., Published on the Internet.

R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramirez-Barahona S, Sauquet H, Magallon S. 2020. The delayed and geographically heterogeneous diversification of flowering plant families. Nature Ecology & Evolution 4: 1232-1238.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. Trends in ecology & evolution, 16(7), 351-358.

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation. science, 317(5840), 910-914.

Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, ... & Mayrose I. 2015. The Chromosome Counts Database (CCDB)–a community resource of plant chromosome numbers. New Phytologist 206: 19-26.

Rivero R, Sessa EB, Zenil-Ferguson R. 2019. EyeChrom and CCDB curator: visualizing chromosome count data from plants. Applications in Plant Sciences 7: e01207.

Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. American journal of botany 105: 302-314.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, ... & Soltis PS. 2009. Polyploidy and angiosperm diversification. American journal of botany 96: 336-348.

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, Majure L, Miles NM, Mavrodiev EV, ... Gitzendanner MA. 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. 2011. New Phytologist 202: 1105-1117.

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. Current opinion in genetics & development 35: 119-125.

Stebbins GL. 1971. Chromosomal evolution in higher plants. London, UK: Edward Arnold.

van der Bijl, W. (2018). phylopath: Easy phylogenetic path analysis in R. PeerJ, 6, e4718.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the national Academy of sciences 106: 13875-13879.

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8: 28-36.

Zenil-Ferguson R, Burleigh JG, Freyman WA, Igić B, Mayrose I, Goldberg EE. 2019. Interaction among ploidy, breeding system and lineage diversification. New Phytologist 224: 1252-1265.

Zenil-Ferguson R, Ponciano JM, Burleigh JG. 2017. Testing the association of phenotypes with polyploidy: An example using herbaceous and woody eudicots. Evolution 71: 1138-1148.

Zhan SH, Otto SP, Barker MS. 2021. Broad variation in rates of polyploidy and dysploidy across flowering plants is correlated with lineage diversification. bioRxiv: 10.1101/2021.03.30.436382

Supplementary Information

Supplementary Tables 1–9.

Figure legends

Fig. 1. Karyotypic diversity (KD) across angiosperms families. For each family the number of distinct chromosome numbers found is shown.

Fig. 2. Phylogenetic least square weighted regressions modelling karyotypic diversity (KD) effects on three alternative estimates of diversification rates (div0, div0.9, MEDUSA), species richness (SR) and clade age, across the full angiosperm dataset and major clades (as indicated), using a family-level tree (RC_complete_MCCv_2 from Ramírez-Barahona et al., 2020). KD was calculated by counting the number of unique distinct chromosome numbers found in each family (thus accounting for polymorphic chromosome numbers in a given species). Each dot represents a family, and its size is proportional to the sampling effort (proportion of species with chromosome count data with respect to the number of extant species richness). Please note that three alternative estimates of diversification rates were considered: one assuming no extinction (div0; epsilon=0) or high rate of extinction (div0.9; epsilon=0.9) and one (MEDUSA) provided by Smith & Brown (2018) who applied the diversification rate-shifts MEDUSA algorithm.

Fig. 3. Visualization of the causal models using phylopath analysis and the standardised path coefficients. Please note that three alternative estimates of diversification rates (DIV) were considered (A, B, C): one assuming no extinction (epsilon=0) or high rate of extinction (epsilon=0.9) and one (MEDUSA) provided by Smith & Brown (2018) who applied the diversification rate-shifts MEDUSA algorithm. KD = Karyotypic diversity, SR = species richness.

Table 1. Phylogenetic least square weighted regressions modelling karyotypic diversity (KD) effects on three alternative estimates of diversification rates (div0, div0.9, MEDUSA), species richness (SR) and clade age (age), across the full angiosperm dataset and major clades (as indicated), using a family-level tree (RC_complete_MCCv_2 from Ramírez-Barahona et al., 2020). KD was calculated by counting the number of unique distinct chromosome numbers found in each family (thus accounting for polymorphic chromosomenumbers in a given species). Statistical significance, variance explained and phylogenetic signal (95% confidence intervals in brackets) are reported. The positive or negative relationship between variables is indicated by the sign of the regression slope coefficient. Significant variables ($P < 0.01$) are reported in bold. The number of families accounted for in the analyses (n) is also reported. Please note that three alternative estimates of diversification rates were considered: one assuming no extinction (div0; epsilon=0) or high rate of extinction (div0.9; epsilon=0.9) and one (MEDUSA) provided by Smith & Brown (2018) who applied the diversification rate-shifts MEDUSA algorithm.

| clade	n (sampled families)	predictor	intercept	slope	P	Rsq	lambda
angiosperms	328	div0	-2.137 (-2.205, -2.069)	0.438 (0.397, 0.479)	**0**	**0.557**	0.866 (0.682, 1.05)
angiosperms	328	div0.9	-3.063 (-3.167, -2.959)	0.756 (0.693, 0.819)	**0**	**0.555**	0.803 (0.589, 1.017)
angiosperms	328	MEDUSA	-1.543 (-1.612, -1.473)	0.259 (0.216, 0.302)	**0**	**0.368**	0.594 (0.328, 0.859)
angiosperms	328	SR	-0.335 (-0.463, -0.207)	1.449 (1.371, 1.527)	**0**	**0.718**	0.772 (0.538, 1.005)
angiosperms	328	age	1.913 (1.885, 1.942)	0 (0, 0)	1	0.596	1.17 (1.17, 1.17)
monocots	69	div0	-2.134 (-2.281, -1.986)	0.41 (0.31, 0.511)	**0**	**0.577**	1.23 (0.758, 1.703)
monocots	69	div0.9	-3.03 (-3.265, -2.795)	0.723 (0.56, 0.886)	**0**	**0.559**	1.13 (0.599, 1.661)
monocots	69	MEDUSA	-1.562 (-1.663, -1.461)	0.223 (0.154, 0.293)	**0**	**0.383**	1.255 (0.829, 1.681)
monocots	69	SR	-0.209 (-0.513, 0.096)	1.4 (1.185, 1.615)	**0**	**0.69**	0.967 (0.408, 1.525)
monocots	69	age	1.96 (1.909, 2.012)	0.011 (-0.023, 0.046)	0.52	0.555	1.463 (1.139, 1.787)
rosids	103	div0	-1.895 (-1.982, -1.807)	0.454 (0.39, 0.518)	**0**	**0.468**	0.705 (0.293, 1.117)
rosids	103	div0.9	-2.677 (-2.811, -2.544)	0.779 (0.682, 0.876)	**0**	**0.504**	0.745 (0.363, 1.127)
rosids	103	MEDUSA	-1.373 (-1.447, -1.3)	0.16 (0.1, 0.22)	**0**	**0.291**	0.339 (-0.208, 0.886)
rosids	103	SR	0.202 (0.02, 0.384)	1.411 (1.288, 1.534)	**0**	**0.718**	1.068 (0.781, 1.354)
rosids	103	age	1.88 (1.827, 1.933)	-0.026 (-0.054, 0.003)	0.074	0.000	1.907 (1.907, 1.907)
asterids	76	div0	-1.861 (-1.969, -1.753)	0.416 (0.349, 0.483)	**0**	**0.621**	1.035 (0.232, 1.838)
asterids	76	div0.9	-2.616 (-2.789, -2.443)	0.72 (0.611, 0.829)	**0**	**0.611**	0.953 (0.156, 1.75)
asterids	76	MEDUSA	-1.528 (-1.677, -1.378)	0.34 (0.251, 0.429)	**0**	**0.355**	1.207 (0.35, 2.064)
asterids	76	SR	-0.073 (-0.325, 0.178)	1.628 (1.481, 1.775)	**0**	**0.79**	1.278 (0.711, 1.845)
asterids	76	age	1.9 (1.86, 1.94)	-0.008 (-0.028, 0.013)	0.451	0.506	1.602 (1.595, 1.61)

1

Fig. 1. Karyotypic diversity (KD) across angiosperms families. For each family the number of distinct chromosome numbers found is shown.

Fig. 2. Phylogenetic least square weighted regressions modeling karyotypic diversity (KD) effects on diversification rates, species richness and clade age, across the full angiosperm dataset and major clades (as indicated), using a family-level tree (RC_complete_MCCv_2 from Ramírez-Barahona et al., 2020). Each dot represents a family, and its size is proportional to the sampling effort (proportion of species with chromosome count data with respect to the number of extant species richness). Please note that three alternative estimates of diversification rates were considered: one obtained from diversification rate-shift analyses using MEDUSA (provided by Smith and Brown, 2018) and two assuming either no extinction (epsilon=0) or high rate of extinction (epsilon=0.9).

Fig. 3. Visualization of the causal models using phylopath analysis and the standardised path coefficients. Please note that three alternative estimates of diversification rates (DIV) were considered
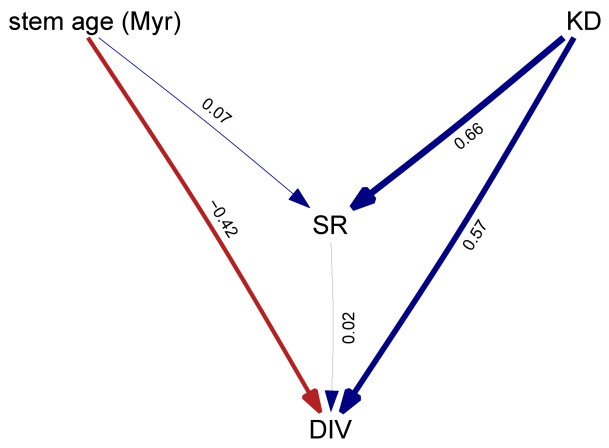
(A, B, C): one obtained from diversification rate-shift analyses using MEDUSA (provided by Smith and Brown, 2018) and two assuming either no extinction (epsilon=0) or high rate of extinction (epsilon=0.9). KD = Karyotypic diversity, SR = species richness.

asterids
monocots
other angiosperms
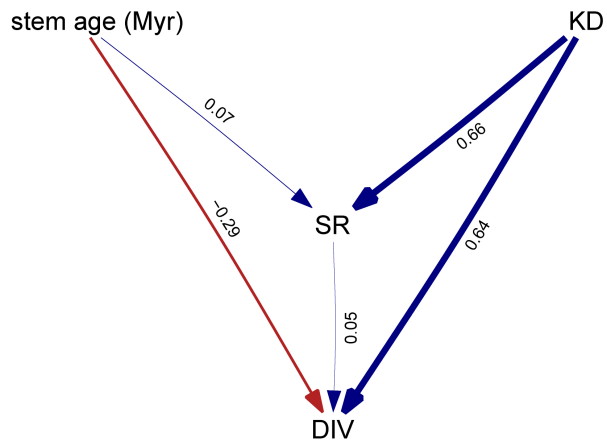rosids

Number of cytotypes
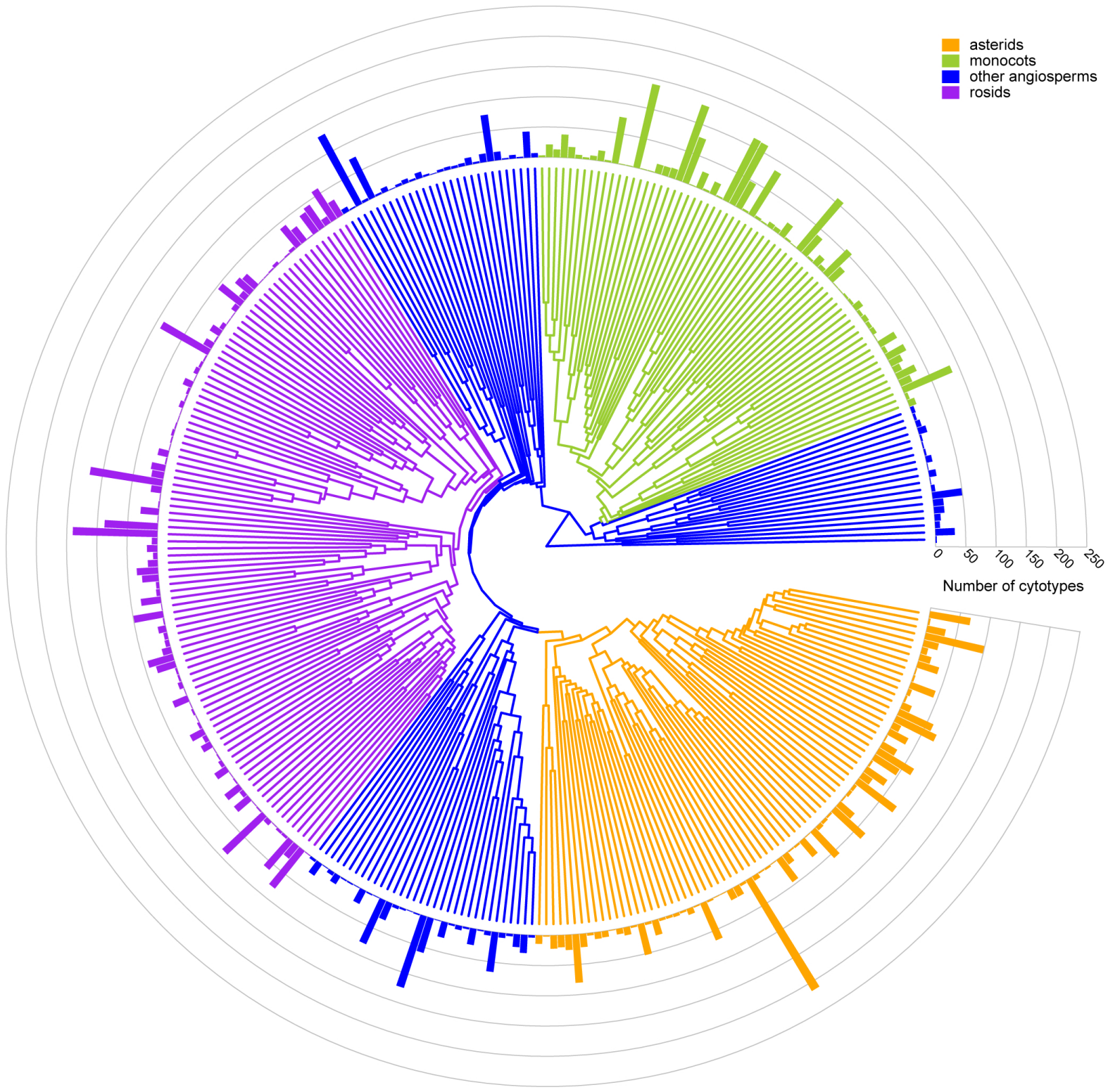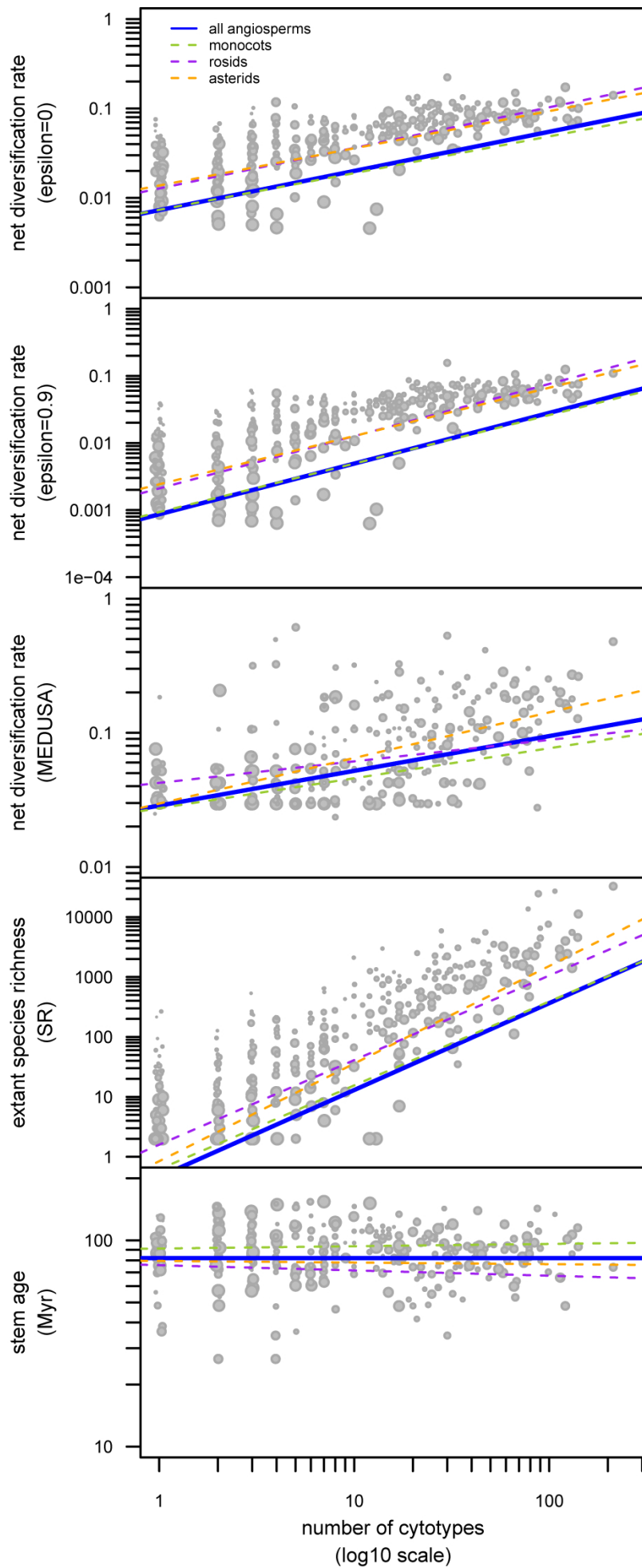
A (net diversification rate [DIV], epsilon=0)

B (net diversification rate [DIV], epsilon=0.9)

C (net diversification rate [DIV], MEDUSA)