A deep dive into the ancestral chromosome number of flowering plants

Angelino Carta*, Gianni Bedini, Lorenzo Peruzzi

Dipartimento di Biologia, Botany Unit, University of Pisa, via Derna 1, I-56126 Pisa, Italy

* Corresponding author: angelino.carta@unipi.it

Abstract

Chromosome rearrangements are a well-known evolutionary feature in eukaryotic organisms[1], especially plants. The remarkable diversity of flowering plants (angiosperms) has been attributed, in part, to the tremendous variation in their chromosome number[2]. This variation has stimulated a blossoming number of speculations about the ancestral chromosome number of angiosperms[2-7], but estimates so far remain equivocal and relied on algebraic approaches lacking an explicit phylogenetic framework. Here we used a probabilistic approach to model haploid chromosome number (*n*) changes[8] along a phylogeny embracing more than 10 thousands taxa, to reconstruct the ancestral chromosome number of the common ancestor of extant angiosperms and the most recent common ancestor for single angiosperm families.

Bayesian inference revealed an ancestral haploid chromosome number for angiosperms $n = 7$, reinforcing previous hypotheses[2-7] that suggested a low ancestral basic number. Inferred *n* for single families, more than half of which are provided here for the first time, are mostly congruent with previous evaluations. Chromosome fusion (loss) and duplication (polyploidy) are the predominant transition types inferred along the phylogenetic tree, emphasising the importance of both dysploidy[6,9,10] and genome duplication[2,7,11-13] in chromosome number evolution. Significantly, while dysploidy is equally distributed early and late across the whole phylogeny, polyploidy is detected mainly towards the tips of the tree. Therefore, little evidence exists for a link between ancestral chromosome numbers and putative ancient polyploidization events[14], suggesting that further insights are needed to elucidate the organization of genome packaging into chromosomes.

Introduction

Each eukaryotic organism has a characteristic chromosome complement, its karyotype, which represents the highest level of structural and functional organization of the nuclear genome[15]. Karyotype constancy ensures the transfer of the same genetic material to the next generation, while karyotype variation provides genetic support to ecological differentiation and adaptation[2,15]. Cytogenetic studies have shown that the tremendous inter- and intra-taxonomic variation of chromosome number documented in flowering plants[2,5,16] is mostly driven by two major mechanisms: a) increases through polyploidy (which may entail a Whole Genome Duplication [WGD] or an increase by half of the genome, demi-duplication[8]); b) decreases or increases through structural chromosomal rearrangements like chromosome fusion, i.e. descending dysploidy, and chromosome fission, i.e. ascending dysploidy.

Polyploidy is a common and ongoing phenomenon, especially in plants[13], that has played an important role in many lineages, with evidence of several rounds of both ancient and recent polyploidization[11,17,18], albeit its distribution in time remains contested[14]. Indeed, although the crucial role of polyploidy in plant diversification on small timescales is widely accepted[2,6], the evolutionary significance of polyplodization for the long-term diversity of angiosperms is still controversial[12]. On the other hand, while dysploidy is more frequent than polyploidy in angiosperms[6], its adaptive consequences have been mostly unexamined[19], until recent studies demonstrated its high evolutionary impact[9].

Chromosome number variation across angiosperm lineages spans two orders of magnitude[15], from $2n = 4$ to $2n = 640$. Previous hypotheses[2-7] of the ancestral basic (monoploid)[20] chromosome number p in angiosperms suggest low numbers, between $p = 6$ and $p = 9$. These hypotheses placed particular attention to 'primitive' extant angiosperms[5] to estimate putative ancestral basic chromosome numbers. More recently, an ancestral chromosome number has been reconstructed using a maximum parsimony approach[7]. However, although parsimony has been widely used to infer ancestral chromosome numbers, it carries significant shortcomings[8],

and more rigorous and complex models to infer chromosome number evolution are currently available [8,21,22,23].

Here we use probabilistic models, accounting for various types of chromosome number transitions, to reconstruct the ancestral haploid chromosome number and the occurrence of chromosome change events across the most massive data set ever assembled linking chromosome numbers to a phylogeny, sampling 10,766 taxa from 59 orders (92%) and 318 families (73%) of angiosperms.

Chromosome numbers were extracted from the Chromosome Counts DataBase[24], and the analyses were conducted using pruned versions of two recently published, dated mega-trees for seed plants[25], the first one (GBMB) constructed with a backbone based on Magallón et al.[26], and the second one (GBOTB) grounded on Open Tree of Life version 9.1. In addition, we conducted all analyses again using a different ultrametric tree of 1,559 taxa extracted from a recently published plastid phylogenomic angiosperm (PPA) tree[27].

Results and Discussion

Regardless of the three alternative phylogenies, n = 7 was inferred as the ancestral haploid chromosome number with the highest posterior probability (Table 1) and likelihood (Table 2). The ancestral haploid chromosome number n = 7 was remarkably stable in the deepest part of the phylogeny (Fig. 1), while slight variations (± 1) in n were inferred at the base of some lineages. Greater variations were shown in the ancestral haploid chromosome number of many angiosperm families (see Supplementary Table 1). Monocots exhibited the largest variation of inferred n among Most Recent Common Ancestors (MRCAs) of plant families (Fig. 2b), paralleled by a considerable variation in current haploid chromosome numbers (Fig. 2a). Over 70% of inferred n in the 158 families for which previous inferences were available are in line with previous proposals. For the remaining 160 families (50.3%) the first inferences are presented here. Discrepancies at the family level among inferences obtained in this study and

those from previous literature are possibly due to the use of routine algebraic approaches, instead of phylogenetic models, to infer chromosome number changes[28]. For example, the inferred *n* of MRCAs of Brassicaceae, Lamiaceae, and Rosaceae are respectively 7, 8, and 7 in our study, but were previously inferred as 12, 14, and 9, respectively[5]. Indeed, even in the presence of a strong phylogenetic signal (e.g., closely related species sharing similar chromosome numbers)[29,30], algebraic inferences of chromosome numbers become increasingly difficult with increasing phylogenetic depth, as identical chromosome numbers will occur in unrelated lineages[19]. The dataset analysed here is the most extensive ever used for inferring ancestral haploid number in angiosperms, but it still poses challenges concerning incomplete taxon sampling and phylogenetic resolution at the family level.

For both GBMB and GBOTB phylogenies, the best model considers up to six parameters (Table 3), i.e. chromosome gain, loss, duplication, demi-duplication rates and rates of gain and loss linearly dependent on the current chromosome number. Our results support the conclusion that genome duplication and dysploidy were critical events in the evolution of angiosperms. Specifically, descending dysploidy, most likely through chromosome fusion, was the most common cytogenetic mechanism of chromosome number change during the evolution of flowering plants, and this is inferred both on branches leading to major clades and on terminal branches (Supplementary Figs. 1-3). Our results emphasize the importance of dysploidy in the evolution of chromosome numbers in angiosperms[9]. Interestingly, demi-duplication events, associated with hybridisation between different ploidy levels and with allopolyploidization processes, were also inferred in a significant number of events, albeit mainly in terminal branches.

Whilst polyploidization is the second most frequent transition type, ancient polyploidy events are underrepresented. The absence of polyploidization events at the base of the tree is in agreement with the maintenance of the ancestral haploid chromosome number $n = 7$ inferred in the deepest part of the phylogeny. Polyploidization events were instead inferred mainly toward

the tips of the tree (Supplementary Fig. 1,2,3), partially supporting previous evidence revealing independent genome duplications near the base of several families[9,18,28], leading to high haploid chromosome numbers. Indeed, this may be the case of many families in the Magnoliid clade. Our results provide no direct insight for a link among some of the most extensive plant radiations and ancient polyploidization rounds[11]. However, our analyses do not contradict evidence of WGD events at the origin of angiosperms or before it[14], but rather highlight that genome size may vary independently of chromosome number[7]. Inferring ancient polyploidy events from cytological data is indeed a challenging task, because genome rearrangements[31] following polyploidisation gradually can hide signals of genome duplication over time[9]. Phylogenomic analyses interpret gene duplications as the result of a shared duplication event occurring in a common ancestor, while models of chromosome number evolution consider WGDs as separately occurring in different lineages[31].

The main results presented here were drawn using the largest dated mega-tree currently available for seed plants. We explored the sensitivity of our results by conducting all analyses again using a different ultrametric tree[26], including a lower number of sampled taxa but allowing to consider intraspecific chromosome number variation for each taxon. We found only minor differences at the root (Tables 1,2) and at some internal nodes (Fig. 1).

Reconstructing the ancestral chromosome number is difficult, because there are no suitable outgroups for direct comparison[32], and because extant early branching angiosperms (e.g., *Amborella* and Nymphaeales) are not necessarily holding plesiomorphic character-states. With these limitations in mind, we made inferences based on the distribution of *n* in extant angiosperms, and using probabilistic models accounting for various types of chromosome number transitions. Our study is not able to address the origin of chromosome number of the first angiosperms. Instead, it provides novel, detailed, and well-supported inference of ancestral haploid chromosome number of the common ancestor of all extant angiosperms, as well as the earliest steps of the subsequent chromosome number transitions, including *n* inferences for 318

angiosperm families. Interestingly, our inferred ancestral state for the haploid number *n* coincides with the ancestral basic chromosome number *p* previously proposed for angiosperms based on empirical counts[2-7] or paleogenomic approaches[33].

Our study allowed to clarify a long-standing question[2], but such reconstruction necessarily comes with limitations. Nevertheless, this is a major step forward in understanding the ancestral chromosome number for angiosperms, and we believe that this issue should be added to the angiosperm macroevolutionary agenda[34]. Progress in reconstructing the ancestral chromosome number may require the development of models that include heterogeneity in the patterns of chromosome evolution across a phylogenetic tree[23], along with a deeper insight into genome and karyotype evolution.

Methods

Phylogenetic reconstruction

We used two recently published[25] dated megaphylogenies for seed plants, GBMB and GBOTB, as backbones to generate two alternative phylogenies for angiosperms included in the dataset. GBMB and GBOTB were constructed using 79,874 and 79,881 taxa, respectively, available in GenBank and in a backbone provided either by Magallón et al.[26] (GBMB) or by Open Tree of Life, version 9.1 (GBOTB). In addition, we also used a different ultrametric tree provided by a recently published plastid phylogenomic angiosperm (PPA) tree[27].

Chromosome numbers collection

The haploid chromosome numbers (*n*) of the species were obtained from the Chromosome Counts Database (CCDB[24]; http://ccdb.tau.ac.il/; last accessed May 2019) using the R package chromer[35]. CCDB contains records from original sources that have irregularities of chromosome counts, so that the ca. 150,000 records were curated semiautomatically using the CCDBcurator package[36] and custom R scripts. After a first round of automatic cleaning, we examined results by hand and corrected records where needed.

Species with unknown chromosome counts were pruned from the trees, thus we collected chromosome numbers for 10,766 taxa included in the GBMB and GBOTB phylogenetic trees, and for 1,559 taxa included in the PPA tree. In cases where multiple chromosome numbers were reported for a given taxon, the modal number was used[8,37]. For taxa with numbers suggesting different ploidy levels, we used the lowest haploid chromosome number available[38]. This coding scheme allowed us to deal with the problem of the existence of different ploidy levels in a taxon and also with the low-density sampling conducted in most taxa[38]. Analyses conducted using the PPA tree encountered less computation limitations, so that we were able to perform them by explicitly considering intraspecific polymorphism, allowing several chromosome numbers, together with their respective frequencies, to be set for each taxon[21].

Analyses

The evolution of haploid chromosome numbers of angiosperms was inferred using chromEvol[21] software v.2.0 (http://www.tau.ac.il/~itaymay/cp/chromEvol/index.html). This software determines the likelihood of a model to explain the given data along the phylogeny, based on the combination of two or more of the following parameters: dysploidisation (ascending, chromosome gain rate λ; descending, chromosome loss rate δ), polyploidisation (chromosome number duplication with rate ρ, demi-polyploidisation or triploidisation with rate μ) and incremental changes to the basic number with regard to a rate of multiplication that is different from a regular duplication[8]. Two additional parameters (λ_1, δ_1) detect linear dependency between the current haploid number and the rate of gain and loss of chromosomes. We tested 10 models based on a different combination of the parameters above. Four of these models consider only constant rates (Mc1, Mc2, Mc3, and Mc0), whereas the other four include two linear rate parameters (Ml1, Ml2, Ml3, and Ml0; Table 3). Both sets have a null model (Mc0 and Ml0) that assumes no polyploidisation events. Finally, two models (Mb1 and Mb2) consider that the evolution of chromosome number can also be influenced by the basic number (β) and by its transition rates (ν).

The minimum chromosome number allowed in the analyses was set to 2, whereas the maximum number was set to 5 units higher than the highest chromosome number found in the empirical data. We removed all counts $n > 43$ from the analysis, because for many lineages the sampling was inadequate to reconstruct such a drastic change in chromosome number[38-39] and because of computation limitations[23]. The branch lengths were scaled according to the software author's instructions. The null hypothesis (no polyploidy) was tested with likelihood ratio tests using the Akaike information criterion (AIC)[40]. To compute the expected number of changes along each branch, as well as the ancestral haploid chromosome numbers at internal nodes, the best fitted model for both data sets was rerun using 1,000 simulations. The best model was plotted on the trees using the ChromEvol functions v0.9-1 elaborated by N. Cusimano (http://www.sysbot.biologie.uni-muenchen.de/en/people/cusimano/use_r.html) in R.

To test which ancestral haploid chromosome number is most likely fort the root of angiosperms, the following haploid chromosome numbers were fixed at the root and the likelihood of the resulting models was compared: n = 4,5,6,7,8,9. These numbers were tested either because considered putative ancestral character-states[2-7], or because they were identified as the chromosome numbers showing the highest PP under our Bayesian analysis. All analyses were performed in the high-performance computing cluster at the University of Pisa.

References

1. Coghlan, A., Eichler, E.E., Oliver, S.G., Paterson, A.H. & Stein, L. Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet*. **21**, 673–682 (2005).
2. Stebbins, G.L. *Chromosomal Evolution in Higher Plants*. (Edward Arnold, London, 1971).
3. Ehrendorfer, F., Krendl, F., Habeler, E., & Sauer, W. Chromosome numbers and evolution in primitive angiosperms. *Taxon* **17**, 337–353 (1968).

4. Walker, J.W. Chromosome numbers, phylogeny, phytogeography of the Annonaceae and their bearing on the (original) basic chromosome number of angiosperms. Taxon **21**, 57–65 (1972).
5. Raven, P.H. The bases of angiosperm phylogeny: cytology. *Ann. Missouri Bot. Gard*. **62**, 724–764 (1975).
6. Grant, V. *Plant Speciation* (ed. 2) (Columbia University Press, New York, 1981).
7. Soltis, D.E., Soltis, P.S., Endress, P.K. & Chase, M.W. in *Phylogeny and Evolution of Angiosperms* (ed. Soltis, D.E., Soltis, P.S., Endress, P.K. & Chase, M.W.) 287–302 (Sinauer Associates, Sunderland, 2005).
8. Mayrose, I., Barker, M.S. & Otto, S.P. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Syst. Biol*. **59**, 132–144 (2009).
9. Escudero, M. et al. Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PLOS One* **9**, e85266 (2014).
10. Guerra, M. Chromosome numbers in plant cytotaxonomy: concepts and implications. *Cytogenet. Genome Res*. **120**, 339–350 (2008).
11. Jiao, Y. et al. Ancestral polyploidy in seed plants and angiosperms. Nature **473**, 97 (2011).
12. Mayrose, I. et al. Recently formed polyploid plants diversify at lower rates. *Science* **333**, 1257–1257 (2011).
13. Wood, T.E. et al. The frequency of polyploid speciation in vascular plants. *Proc. Nat. Acad. Sci. USA* 106, 13875–13879 (2009).
14. Ruprecht, C. et al. (2017) Revisiting ancestral polyploidy in plants. *Science Adv*. **3**, e1603195 (2017).
15. Stace, C.A. Cytology and cytogenetics as a fundamental taxonomic resource for the 20th and 21st centuries. *Taxon* **49**: 451–477 (2000).
16. Levitzky, G.A. The karyotype in systematics. *Bull. Appl. Bot. Genet. Plant Breed*. **27**, 220–240 (1931).

17. Li, Z. et al. Early genome duplications in conifers and other seed plants. *Sci. Adv.* **1**, e1501084 (2015).
18. Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J. et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
19. Weiss-Schneeweiss, H., & Schneeweiss, G.M. in *Plant Genome Diversity* Vol. 2 (eds. Leitch, I.J., Greilhuber, J., Dolezel, J.W.J) 209–230 (Springer, Vienna, 2013).
20. Peruzzi, L. "x" is not a bias, but a number with real biological significance. Plant Biosyst. 147, 1238–1241 (2013).
21. Glick, L. & Mayrose, I. ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.* **31**, 1914–1922 (2014).
22. Freyman, W.A. & Höhna, S. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. *Syst. Biol.* **67**, 195–215 (2018).
23. Zenil-Ferguson, R., Burleigh, J.G. & Ponciano, J.L. Chromploid: an R package for chromosome number evolution across the plant tree of life. *Appl. Plant Sci.* **6**, e1037 10.1002/aps3.1037 (2018).
24. Rice, A. et al. The Chromosome Counts Database (CCDB)–a community resource of plant chromosome numbers. *New Phytol.* **206**, 19–26 (2015).
25. Smith, S.A. & Brown, J.W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 1–13 (2018).
26. Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**, 437–453 (2015).
27. Li, H.T. et al. Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants*, **5**, 461 (2019).

28. Cusimano, N., Sousa, A., & Renner, S.S. Maximum likelihood inference implies a high, not a low, ancestral haploid chromosome number in Araceae, with a critique of the bias introduced by ‘x’. *Ann. Bot.* **109**, 681–692 (2012).
29. Escudero, M. et al. Selection and inertia in the evolution of holocentric chromosomes in sedges (*Carex*, Cyperaceae). *New Phytol*. **195**, 237–247 (2012).
30. Schubert, I., & Lysak, M.A. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Gen*. **27**, 207–216 (2011).
31. Mandakova, T. & Lysak, M.A. Post-polyploid diploidization and diversification through dysploid changes. *Curr. Opinion Plant Biol*. **42**, 55–65 (2018).
32. Doyle, J.A. Molecular and fossil evidence on the origin of angiosperms. *Ann. Rev. Earth Planet. Sci*. **40**, 301–326 (2012).
33. Salse, J. In silico archeogenomics unveils modern plant genome organization, regulation and evolution. *Curr. Opinion Plant Biol*. **15**, 122–130 (2012).
34. Sauquet, H. & Magallón, S. Key questions and challenges in angiosperm macroevolution. *New Phytol*. **219**, 1170–1187 (2018).
35. Pennell, M.W. Chromer: Interface to Chromosome Counts Database API. R package version 0.1.2.9000 (2016).
36. Rivero, R., Sessa, E. B., & Zenil-Ferguson, R. EyeChrom and CCDB curator: Visualizing chromosome count data from plants. *Applications in Plant Sciences*, **7**, e01207 (2019).
37. Salman-Minkov, A., Sabath, N. & Mayrose, I. Whole-genome duplication as a key factor in crop domestication. *Nature Plants* **2**, 16115 (2016).
38. Márquez-Corro, J.I., Martín-Bravo, S., Spalink, D., Luceño, M. & Escudero, M. Inferring hypothesis-based transitions in clade-specific models of chromosome number evolution in sedges (Cyperaceae). *Mol. Phylogenet. Evol*. **135**, 203–209 (2019).
39. Barrett, C.F. et al. Ancient polyploidy and genome evolution in palms. *Genome Biol. Evol*. **11**, 1501–1511 (2019).

40. Burnham, K.P. & Anderson, D.R. Model inference. Understanding AIC and BIC in model selection. *Socio. Meth. Res. Int. J. Bot. Res*. **33**, 261–304 (2004).

Table 1. Summary of chromosome number evolutionary models and inferred ancestral haploid chromosome number (*n*) in Angiosperms under the best-fitting model.

				Rates						Events inferred with PP > 0.5				Chromosome no. at root node		
tree	Best model	LogLik	AIC	λ	δ	ρ	μ	λ_1	δ_1	Gain	Losses	Duplications	Demi	Bayes (PP): Best *p*	Bayes(PP): 2nd best *p*	ML
GBMB (10,766 taxa)	Ml3	-18120.0	36250.0	0.0081	0.0113	0.0131	0.0051	0.0051	0.0007	509.7	1627.5	1438.1	376.1	7 (0.97)	8 (0.02)	5
GBOTB (10,766 taxa)	Ml3	-18110.0	36240.0	0.0106	0.0096	0.0130	0.0049	0.0001	0.0008	589.5	1625.3	1435.3	363.3	7 (0.98)	8 (0.01)	5
PPA (1559 taxa)	Ml2	-3788.0	7586.0	0.0101	0.0044	0.0078	-	-0.0002	0.0012	145.0	528.1	244.6	191.9	7 (0.24)	5 (0.21)	2

Only the best-fitting models are shown. Tree refers to the three alternative phylogenies used. Best model, Ml3 (linear rate model with duplication rate ρ and demi-duplication rate μ) Ml2 (linear rate model with equal duplication and demi-duplication rates); Logarithmic likelihood (LogLik) and AIC scores; rate parameters (λ = chromosome gain rate, δ = chromosome loss rate, ρ = duplication rate, μ = demi-duplication rate, λ_1= linear chromosome gain, δ_1 = linear chromosome loss); frequency of the four possible event types with a posterior probability (PP) > 0.5; haploid chromosome number inferred at the root node under Bayesian optimization with the respective PP, and under maximum likelihood (ML).

Table 2. Testing hypotheses about root ancestral haploid chromosome number (*n*).

Tree	GBMB		GBOTB		PPA	
	LogLik	AIC	LogLik	AIC	LogLik	AIC
Root fixed at *n* = 4	-18127.1	36266.2	-18129.9	36271.7	-3788.8	7587.6
Root fixed at *n* = 5	-18128.1	36268.2	-18142.6	36297.2	-3787.2	7584.5
Root fixed at *n* = 6	-18123.6	36259.2	-18124.2	36260.4	-3786.7	7583.5
Root fixed at *n* = 7	**-18117.1**	**36246.3**	**-18115.5**	**36243.0**	**-3786.2**	**7582.4**
Root fixed at *n* = 8	-18119.4	36250.9	-18127.8	36267.6	-3787.6	7585.1
Root fixed at *n* = 9	-18122.2	36256.4	-18121.3	36254.7	-3788.4	7586.9

AIC and LogLik values obtained by fixing the root with given ancestral haploid chromosome number. The lowest AIC and LogLik values are shown in bold.

Table 3. Goodness of fit of the 10 different models of chromosome number evolution applied to the three alternative phylogenies used.

Model	GBMB			GBOTB			PPA			Parameters
	LogLik	AIC	AIC_w	LogLik	AIC	AIC_w	LogLik	AIC	AIC_w	
Mc1	-19870	39750	0.0000000000	-19900	39800	0.0000000000	-4032	8070	0.0000000000	$\lambda; \delta; \rho$
Mc2	-18280	36560	0.0000000000	-18270	36550	0.0000000000	-3804	7613	0.0000013710	$\lambda; \delta; \rho=\mu$
Mc3	-18140	36290	0.0000000021	-18140	36280	0.0000000021	-3804	7616	0.0000003059	$\lambda; \delta; \rho; \mu$
Mc0	-47870	95740	0.0000000000	-49060	98120	0.0000000000	-5135	10270	0.0000000000	$\lambda; \delta$
Ml1	-19670	39350	0.0000000000	-19660	39320	0.0000000000	-3999	8008	0.0000000000	$\lambda; \delta; \rho; \lambda_1; \delta_1$
Ml2	-18260	36520	0.0000000000	-18250	36510	0.0000000000	**-3788**	**7586**	**0.9999999979**	$\lambda; \delta; \rho=\mu; \lambda_1; \delta_1$
Ml3	**-18120**	**36250**	**0.9999999979**	**-18110**	**36240**	**0.9999999979**	-3787	7587	0.6065306585	$\lambda; \delta; \rho; \mu; \lambda_1; \delta_1$
Ml0	-45260	90530	0.0000000000	-44910	89820	0.0000000000	-4918	9843	0.0000000000	$\lambda; \delta; \lambda_1; \delta_1$
Mb1	-18650	37310	0.0000000000	-18470	36950	0.0000000000	-4001	8011	0.0000000000	λ; δ; β; ν
Mb2	-21790	43580	0.0000000000	-21830	43670	0.0000000000	-4287	8581	0.0000000000	λ; δ; ρ; β; ν

Mc indicate models with constant rates, Ml models that include linear rate parameters and Mb models that include base number (not the chromosome number at the root of the phylogeny) parameters[8-21]. Logarithmic likelihood (LogLik), AIC and relative weights scores (AICw). In bold, the lowest AIC value for each phylogeny indicates the best model. The last column indicates the parameter estimates included in each model (see Methods for details).

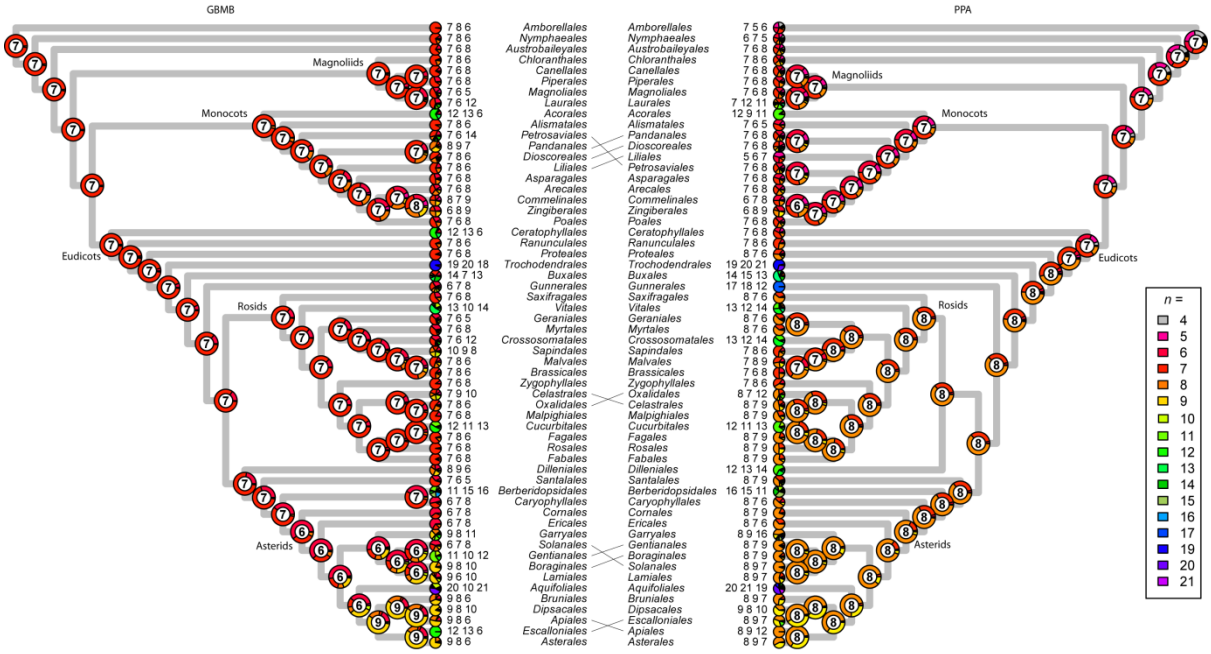

Figure 1 Reconstruction of ancestral haploid chromosome number (*n*) of angiosperms with the best-fitting model on the different types of trees used (GBMB and PPA). Please note that the plotted trees depict ordinal phylogenetic relationships (sub-ordinal topologies were collapsed to build the figure), and are shown without branch length information. Pie charts at nodes represent the probability of the ancestral haploid chromosome numbers inferred under Bayesian estimation; the numbers at nodes are those with the highest probability. Pie charts and numbers at the tips are the three best inferred ancestral haploid chromosome numbers per each angiosperm order. Black lines, difference in phylogenetic position between GBMB and PPA trees.

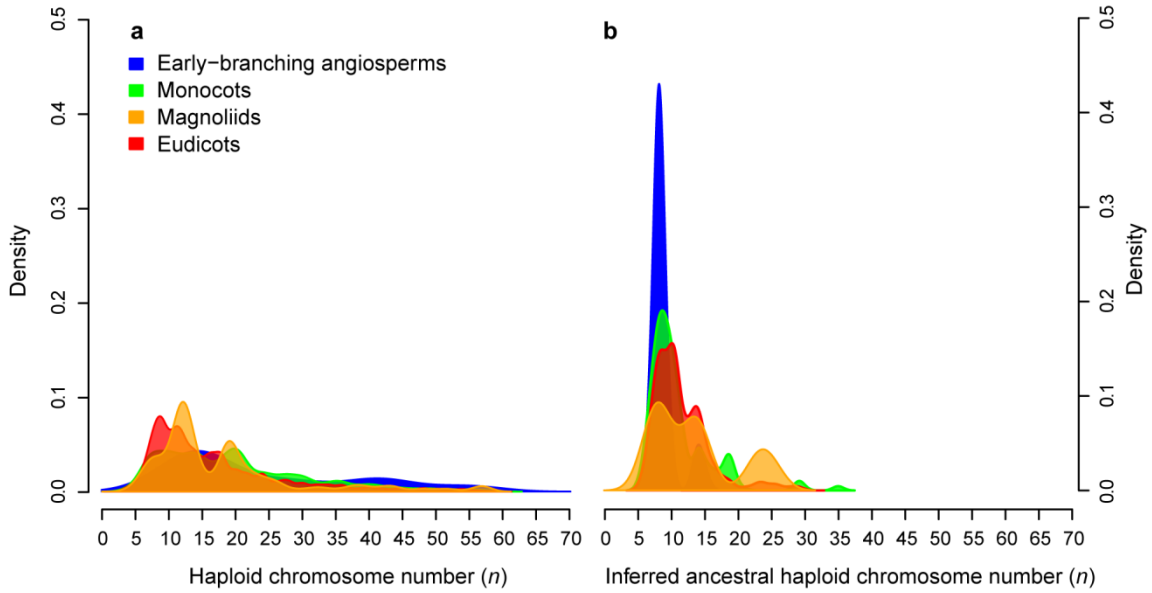

Figure 2 Density plots of haploid chromosome numbers (*n*) and of inferred ancestral haploid chromosome number (*n*) for each angiosperm family in four major angiosperm clades (APG IV). a, we identified the number of unique chromosome counts per taxon, i.e. cytotypes, from the original dataset, after excluding counts with $n > 60$, to focus on the most frequent *n* and their putative relation with inferred *p*. **b**, density plots were scaled by the Bayesian posterior probability (PP) estimated for each inferred *p*.

Acknowledgements

The authors thank Marcial Escudero and Itay Mayrose for their help with ChromEvol analyses.

Author contributions

A.C. planned and designed the research, analysed the data and wrote the manuscript. G.B. assisted in chromosome numbers acquisition. L.P. and G.B. contributed to successive versions of the manuscript and in solving theoretical and nomenclatural issues. All authors read and approved the final manuscript.