



OPEN

Centromere size scales with genome size across Eukaryotes

Klára Plačková, Petr Bureš & František Zedek

Previous studies on grass species suggested that the total centromere size (sum of all centromere sizes in a cell) may be determined by the genome size, possibly because stable scaling is important for proper cell division. However, it is unclear whether this relationship is universal. Here we analyze the total centromere size using the CenH3-immunofluorescence area as a proxy in 130 taxa including plants, animals, fungi, and protists. We verified the reliability of our methodological approach by comparing our measurements with available ChIP-seq-based measurements of the size of CenH3-binding domains. Data based on these two independent methods showed the same positive relationship between the total centromere size and genome size. Our results demonstrate that the genome size is a strong predictor (R -squared = 0.964) of the total centromere size universally across Eukaryotes. We also show that this relationship is independent of phylogenetic relatedness and centromere type (monocentric, metapolycentric, and holocentric), implying a common mechanism maintaining stable total centromere size in Eukaryotes.

The centromere is the chromosomal region where the kinetochore, a protein complex that mediates the chromosome's attachment to spindle microtubules, assembles¹. Thus, the centromere plays a vital role in the cell division of Eukaryotes, as it mediates the proper segregation of chromosomes into daughter cells. In most Eukaryotes, the centromeric function is defined epigenetically by the presence of centromeric histone H3 (CenH3 or CENP-A), which recruits other kinetochore proteins², although CenH3-independent systems also exist, e.g., holocentric insects³, some holocentric plants⁴, kinetoplastids⁵, and some fungi⁶. Centromeric DNA is usually AT-rich¹, but specific DNA sequences are not necessary nor sufficient for the kinetochore assembly².

The size of CenH3-containing domains determines the kinetochore size and thus the functional centromere size^{7–10}. Forty years ago, Bennett et al.¹¹ analyzed nine species of grasses (Poaceae) with electron microscopy and found that the total centromere volume (the sum of all centromere volumes in a cell) linearly scales with the nuclear DNA content (genome size) across these species. Zhang and Dawe⁷ confirmed this relationship by analyzing a partially overlapping set of ten grass species and showing that the total size of the CenH3-immunostained area (a proxy for total kinetochore size) strongly positively correlates with genome size. A positive, although much weaker, relationship has also been observed within species across 26 maize lines differing in genome size¹⁰. These results indicate that there may be a mechanism maintaining the stable proportion of total centromere size to the genome size that is based on general intracellular scaling principles⁷. This notion was supported by observation in maize/oat and maize/maize hybrids that centromeres may expand when chromosomes are introduced into larger genomes^{9,10}. However, it is unclear whether the scaling of total centromere size to genome size is a universal phenomenon because it was observed on grasses only.

Therefore, in the present study, we have measured the total centromere size (using immunostained-CenH3 areas as a proxy) in 130 eukaryotic species, including plants, animals, fungi, and protists (Fig. 1; Supplementary Table S1), and tested whether the relationship between the centromere and genome size observed in grasses is valid in general. To validate our approach, we compared the CenH3-immunofluorescence measurements with the measurements of CenH3 domains based on ChIP-seq analyses. We also considered the analyzed taxa's phylogenetic relatedness and also differences in centromere organization due to possession of monocentric chromosomes (with a single regional centromere), metapolycentric chromosomes (having multiple separated kinetochore regions in the primary constriction¹²), or holocentric chromosomes (centromere function along the chromosome¹³).

Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, 61137 Brno, Czech Republic.
 email: fzedek@gmail.com

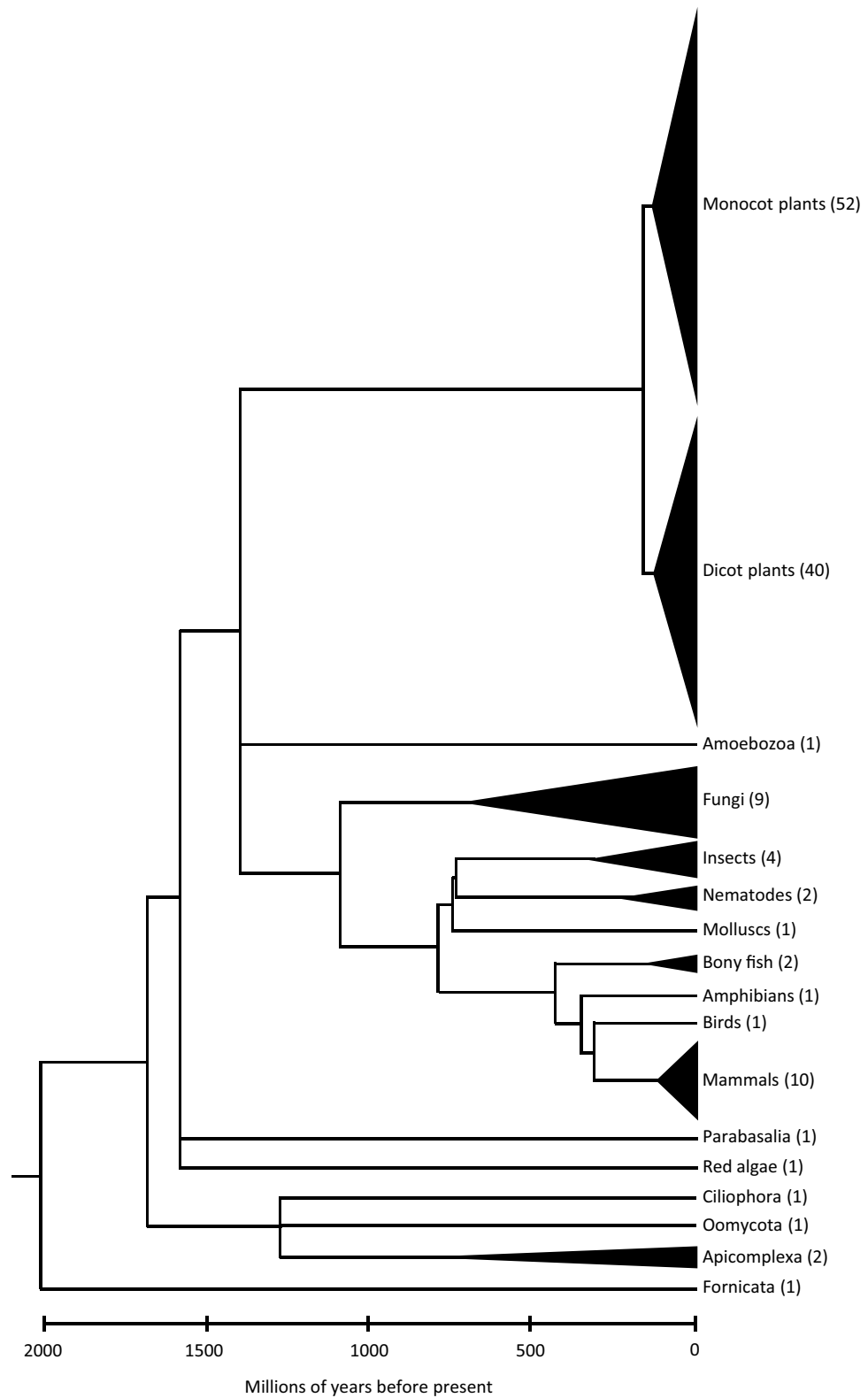


Figure 1. A simplified phylogenetic tree shows the distribution of analyzed taxa across Eukaryotes. Numbers in brackets correspond to the number of analyzed species from the respective clade.

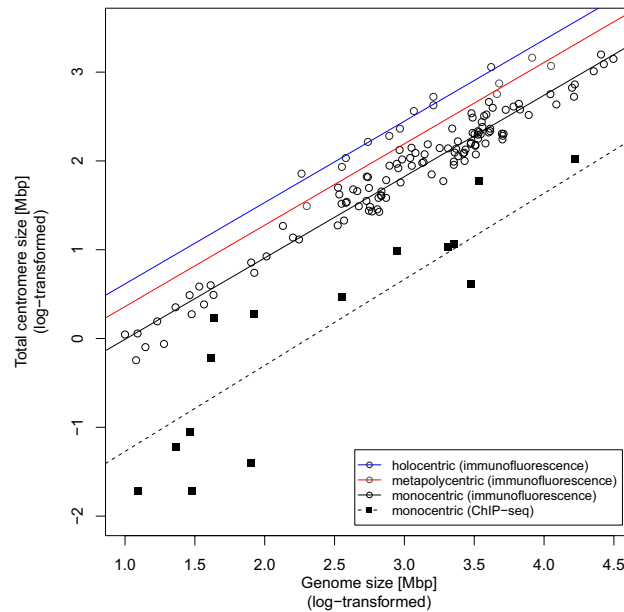


Figure 2. Relationship between the genome size (log-transformed) and the total centromere size (log-transformed) and the effects of chromosome type on this relationship as estimated by the phylogenetically corrected regression model (see details in Table 1). The figure was generated using basic plot functions in R v4.0.2³⁵.

Model term	b_i	$se(b_i)$	t	P
Holocentric chromosomes (Intercept)	-0.302	0.065	-4.632	<0.0001
Genome size	0.916	0.016	56.478	<0.0001
Metapolycentric chromosomes	-0.255	0.078	-3.255	0.0015
Monocentric chromosomes	-0.626	0.047	-13.361	<0.0001

Table 1. Outcome of the additive regression model: Total centromere size ~ Genome size + Chromosomes. The outcome of multiple linear regression model of the relationship of Total centromere size (log-transformed) to Genome size (log-transformed) for different chromosome types. The slope of the regression line ($b = 0.916$) is the same for all chromosome types. The intercept of the regression line is highest for holocentric chromosomes ($b = -0.302$), lower for metapolycentric chromosomes ($b = -0.302 - 0.255 = -0.557$) and lowest for monocentric chromosomes ($b = -0.302 - 0.626 = -0.928$). b_i —coefficient estimate, $se(b_i)$ —standard error of the coefficient estimate, t — t -statistics, P —significance. Total Centromere Size = $-0.302 + 0.916 \times$ Genome size $- 0.255 \times$ Metapolycentric chromosomes (0 or 1) $- 0.626 \times$ Monocentric chromosomes (0 or 1).

Results

As we measured most of the total centromere sizes from figures presented in the literature (see Materials and Methods for details), we wanted to verify the validity of our approach. First, we compared the total centromere size for 10 grass species that we measured from Fig. 1B from Zhang and Dawe (2012)⁷ with the authors' own values obtained as averages of measurements of multiple nuclei from high-resolution images⁷. We observed a very good agreement (Pearson's $r = 0.995$, $p < 0.0001$) (Supplementary Fig. S1).

We found that the centromere size strongly positively correlated with the genome size (Fig. 2). The detailed outcome from the regression model is presented in Table 1. The observed relationship was independent of phylogenetic relatedness (Pagel's lambda did not differ from zero, $p = 1$). Nor did chromosome type affect this relationship because the slope of the regression line ($b = 0.916$, $p < 0.0001$) was the same for monocentric, metapolycentric, and holocentric taxa (the interaction term allowing different slopes was not significant, see Supplementary Table S2). However, as expected, monocentric, metapolycentric, and holocentric taxa differed in their genomes' proportion occupied by the functional centromeres, with the lowest proportion in monocentrics and the highest in holocentrics (Fig. 3). The relatively narrow range of the genomes' proportion occupied by the centromeres (Fig. 3) agreed well with the high total variance explained by the regression model (adjusted $R^2 = 0.964$), suggesting that the genome size is a strong predictor of the total centromere size.

To further verify our results, for 15 monocentric species of our dataset, we collected data reported in previous studies on total centromere sizes derived independently from ChIP-seq measurements of CenH3-binding

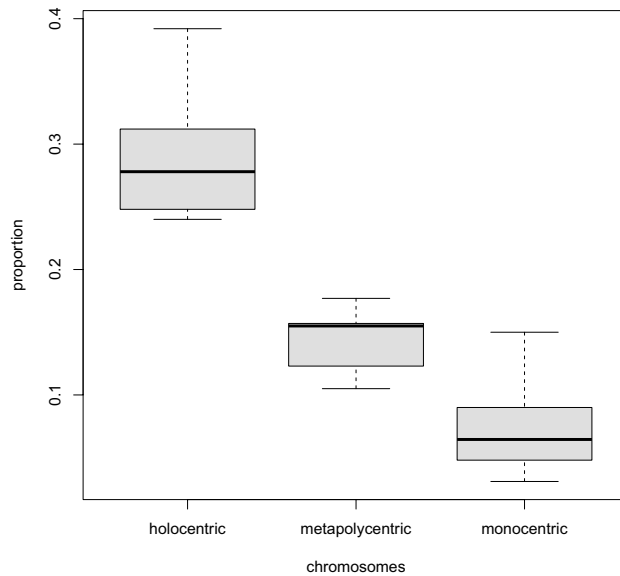


Figure 3. Proportion of the genome area occupied by the functional centromere in taxa with different chromosome types. The figure was generated using basic plot functions in R v4.0.2³⁵.

Model term	b_i	se(b_i)	t	P
Intercept	-2.241	0.595	-3.766	0.0024
Genome size	0.969	0.192	5.056	0.0002

Table 2. Outcome of the regression model: Total centromere size based on ChIP-seq ~ Genome size. The outcome of the linear regression model of the relationship of Total centromere size (log-transformed) to Genome size (log-transformed). b_i —coefficient estimate, se(b_i)—standard error of the coefficient estimate, t—t-statistics, P—significance.

areas (Supplementary Table S1) and tested their relationship to the genome size. The detailed outcome from the regression model is presented in Table 2. On average, the absolute values of total centromere size based on CenH3-immunofluorescence are 20 times higher than estimates based on ChIP-seq analyses (this follows from the difference of intercept from Table 2 and the intercept for monocentric chromosomes in Table 1). However, the slope of the regression line ($b = 0.969$, $p = 0.0002$) from the ChIP-seq based data (Table 2) is very similar to the slope obtained from the analysis of the CenH3-staining-based data (Fig. 2, Table 1) which means that analyses based on these two independent methods show a very similar relationship between the total centromere size and genome size across Eukaryotes.

Discussion

Our results provide evidence that the scaling relationship between the total centromere and genome size, initially observed in grasses^{7,11}, is universal for all Eukaryotes (Fig. 2). Although the estimates of total centromere sizes using the immunofluorescence method are likely to be overestimated, this overestimation is consistent across the species analyzed (Fig. 2), and more importantly, the data obtained by immunofluorescence show the same relationship between the total centromere size and genome size as the data obtained by ChIP-seq (Fig. 2). Thus, we can conclude that immunofluorescence measurement of the total centromere size is a reliable method for the type of comparative analysis used in this study. The larger variation of ChIP-seq based data (Fig. 2) likely stems from the fact that the source genomes are not completely sequenced, so the centromere size estimates are less precise. For the same reason, ChIP-seq-based data are probably underestimated, because incomplete genomes will naturally have smaller centromeres since centromeres are the most difficult regions to assemble.

The independence of the relationship between the total centromere size and genome size on phylogenetic relatedness (see Results) means that it remains the same whether we look at closely or distantly related species, implying that Eukaryotes share a mechanism maintaining a stable proportion of functional centromere to genome size. The potential mechanism also appears the same for taxa with monocentric, metapolycentric, or holocentric chromosomes because the relationship between total centromere size and genome size does not change with chromosome type (Fig. 2), and the proportion of genome occupied by centromeres is stable within each chromosome type (Fig. 3). The mechanism responsible for the strong dependence of total centromere size on genome size may stem from intracellular scaling principles¹⁰ that maintain the size ratio of intracellular

components to ensure their proper function¹⁴, perhaps via regulation of the amount of available CenH3, directly or indirectly through chaperones or licensing factors. The larger centromere proportions in metapolycentrics and holocentrics (Fig. 3) could imply a higher concentration of available CenH3 in these organisms. Zhang and Dawe⁷ hypothesized that a species' genome size determines its total centromere area required to stabilize the spindle. They also surmised that the total centromere area is equally distributed to individual chromosomes⁷. Individual centromeres of uniform sizes could contribute to proper congression and segregation because chromosomes with too large or too small functional centromeres tend to missegregate and get lost^{15–19}. This would also mean that, within a karyotype, functional centromere size does not vary with chromosome size, a notion that has been supported by showing that small chromosomes of maize introduced into oat equalized their centromere sizes with large chromosomes of oat⁹. However, reports from studies on human^{16,20,21}, fescue hybrid^{11,22}, *Arabidopsis*²³, and recently from maize¹⁰ suggest that a moderate within-karyotype correlation between the size of chromosomes and their centromeres may exist.

The reason for such a within-karyotype correlation is unclear, but it appears that for a chromosome of a specific size, there is a lower limit of kinetochore size reflecting the minimum number of kinetochore microtubules required for proper chromosomal segregation^{24–26}. Chromosomes whose kinetochore size falls below this limit are more likely to be lost during repeated rounds of cell division^{16,24–26}. Thus, a sufficiently significant increase in chromosome size could require a corresponding increase in kinetochore size, and/or an increase in kinetochore size could allow an increase in chromosome size. Changes in the size of individual kinetochores could occur either by drift or as a result of deterministic processes such as centromeric or holokinetic drive^{27,28}. It is, therefore, possible there are two antagonistic processes affecting centromere size. The first equalizes the size of individual centromeres (and possibly even entire chromosomes) to ensure proper chromosome behavior during cell division on a cellular level. By contrast, the second process operates on the level of individual chromosomes and may cause centromere and chromosome size divergence within a single karyotype. As the correlation between the total centromere size and genome size (Fig. 2 here;^{7,11}) is much stronger than the within-karyotype correlation between sizes of individual chromosomes and their centromeres^{10,16}, it seems likely that the mechanism keeping centromeres of similar sizes prevails.

Methods

Obtaining the total centromere size and genome size. We reviewed the available literature and collected studies containing microscopic photographs of immunolabeled CenH3 (Supplementary Table S1). We also performed new CenH3 immunostaining in eight grass (Supplementary Fig. S2) and nine agavoid species (Supplementary Fig. S3). For the immunostaining protocols, see Supplementary Text S1. We then processed each microscopic photograph in the ImageJ program²⁹ as follows: (i) We applied split channels and adjust threshold functions to obtain and separate the CenH3-staining area (a proxy for total centromere/kinetochore size) from the DAPI-stained DNA area (a proxy for genome size). First, we used the auto-adjust threshold option, and then we fine-tuned the threshold manually to properly circumscribe the DAPI or CenH3 areas. (ii) We measured the size of CenH3-staining and DAPI-stained DNA areas. (iii) To enable the comparison of area measurements between photographs from different studies/species, we standardized all the measurements using the known 1C genome size for each species as follows: we equalized the measured DAPI area with the known 1C genome size (in Mbp) and calculated the total centromere size in Mbp as [(total CenH3 staining area × 1C genome size)/DAPI staining area]. We had one measurement of the DNA and CenH3 area for each species whose figure was obtained from the literature. For each of the eight grass and nine agavoid species that we newly measured, we measured the DNA and CenH3 area in 6 to 27 nuclei per species (Supplementary Table S3) and used the average value for the final analysis. The values of 1C genome sizes in Mbp were obtained either from the same studies as the figures for the DNA and centromere area measurements, from genome sequencing projects, from the Animal Genome Size Database³⁰, from the Plant C-values Database³¹, from Genome Size of the Czech Vascular Flora³², or from the Fungal Genome Size Database³³. In the case of the nine Agavoideae species, we measured their genome size using flow cytometry (see Supplementary Text S1).

The values for the total centromere size based on ChIP-seq measurements were obtained from the published papers where they were reported either in the text, tables, or supplementary data. The values in Mbp as well as the references to the respective papers are listed in Supplementary Table S1.

Statistical analyses. To test the relationship between the total centromere size and genome size, while accounting for a potential non-independence of species due to their shared ancestry, we used a phylogenetically corrected linear regression using the *pgls* function implemented in the package *caper*³⁴ in R v4.0.2³⁵. We set the total centromere size as a response variable and genome size and chromosome type as explanatory variables. Both the total centromere size and genome size were log-transformed before the regression analysis to increase the homogeneity of variances in the response variable. First, we explored the model with the interaction between explanatory variables to check whether the effect of genome size on the total centromere size depends on chromosome type. The interaction was not significant (Supplementary Table S2). We have, therefore, continued with the additive model (Table 1). Because phylogenetically corrected regression assumes the tree is ultrametric (i.e., all the tips are equidistant from the root)³⁶, we used a dated phylogeny, as it fulfills this assumption. The dated phylogenetic tree (Supplementary Text S2) for most analyzed taxa was obtained by combining the TimeTree³⁷ and the comprehensive dated phylogeny of Angiosperms³⁸. If a species was not present in the trees, we replaced their tip with the closest relative or added it manually based on the published phylogenies as in the case of *Allium*³⁹ and *Cuscuta*⁴⁰.

Methods statement. All methods were performed in accordance with the relevant guidelines and regulations.

Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

Received: 16 July 2021; Accepted: 24 September 2021

Published online: 06 October 2021

References

- Talbert, P. B. & Henikoff, S. What makes a centromere?. *Exp. Cell Res.* **389**, 111895 (2020).
- Murillo-Pineda, M. & Jansen, L. E. T. Genetics, epigenetics and back again: Lessons learned from neocentromeres. *Exp. Cell Res.* **389**, 111909 (2020).
- Drinnenberg, I. A., deYoung, D., Henikoff, S. & Malik, H. S. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife* **3**, 3676 (2014).
- Krátká, M., Šmerda, J., Lojdová, K., Bureš, P. & Zedek, F. Holocentric Chromosomes Probably Do Not Prevent Centromere Drive in Cyperaceae. *Front. Plant Sci.* **12**, 642661 (2021).
- Akiyoshi, B. & Gull, K. Discovery of unconventional kinetochores in kinetoplastids. *Cell* **156**, 1247–1258 (2014).
- Navarro-Mendoza, M. I. *et al.* Early diverging fungus mucor circinelloides lacks centromeric histone CENP-A and displays a mosaic of point and regional centromeres. *Curr. Biol.* **29**, 3791–3802 (2019).
- Zhang, H. & Dawe, R. K. Total centromere size and genome size are strongly correlated in ten grass species. *Chromosome Res.* **20**, 403–412 (2012).
- Bodor, D.L., Mata, J.F., Sergeev, M., David, A.F., Salimian, K.J., Panchenko, T. *et al.* The quantitative architecture of centromeric chromatin. *Elife* **3**, e02137 (2014).
- Wang, K., Wu, Y., Zhang, W., Dawe, R. K. & Jiang, J. Maize centromeres expand and adopt a uniform size in the genetic background of oat. *Genome Res.* **24**, 107–116 (2014).
- Wang, N., Liu, J., Ricci, W.A., Gent, J.I., Dawe, R.K. Maize centromeric chromatin scales with changes in genome size. *Genetics* **217**, iyab020 (2021).
- Bennett, M. D., Smith, J. B., Ward, J. & Jenkins, G. The relationship between nuclear DNA content and centromere volume in higher plants. *J. Cell Sci.* **47**, 91–115 (1981).
- Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., *et al.* Stretching the Rules: Monocentric Chromosomes with Multiple Centromere Domains. *PLoS Genet* **8**, e1002777 (2012).
- Zedek, F. & Bureš, P. Holocentric chromosomes: From tolerance to fragmentation to colonization of the land. *Ann. Bot.* **121**, 9–16 (2018).
- Levy, D. L. & Heald, R. Mechanisms of intracellular scaling. *Annu. Rev. Cell Dev. Biol.* **28**, 113–135 (2012).
- Heslop-Harrison, J., Chapman, V. & Bennett, M. D. Heteromorphic bivalent association at meiosis in bread wheat. *Heredity* **55**, 93–103 (1985).
- Irvine, D. V. *et al.* Chromosome size and origin as determinants of the level of CENP-A incorporation into human centromeres. *Chromosome Res.* **12**, 805–815 (2004).
- Drpic, D. *et al.* Chromosome segregation is biased by Kinetochore size. *Curr. Biol.* **28**, 1344–1356.e5 (2018).
- Wang, N. & Dawe, R. K. Centromere size and its relationship to haploid formation in plants. *Mol. Plant.* **11**, 398–406 (2018).
- Worrall, J. T. *et al.* Non-random Mis-segregation of Human Chromosomes. *Cell Rep.* **23**, 3366–3380 (2018).
- Sánchez, L., Martínez, P. & Goyanes, V. Analysis of centromere size in human chromosomes 1, 9, 15, and 16 by electron microscopy. *Genome* **34**, 710–713 (1991).
- Martorell, M. R., Benet, J., Márquez, C., Egozcue, J. & Navarro, J. Correlation between centromere and chromosome length in human male pronuclear chromosomes: ultrastructural analysis. *Zygote* **8**, 79–85 (2000).
- Jenkins, G. & Bennett, M. D. The intranuclear relationship between centromere volume and chromosome size in *Festuca scariosa* X *drymeja*. *J. Cell Sci.* **47**, 117–125 (1981).
- Koornneef, M., Fransz, P. & de Jong, H. Cytogenetic tools for *Arabidopsis thaliana*. *Chromosome Res.* **11**, 183–194 (2003).
- Moen, P. B. Kinetochore microtubule numbers of different sized chromosomes. *J. Cell Biol.* **83**, 556–561 (1979).
- Cherry, L. M., Faulkner, A. J., Grossberg, L. A. & Balczon, R. Kinetochore size variation in mammalian chromosomes: an image analysis study with evolutionary implications. *J. Cell Sci.* **92**, 281–289 (1989).
- McEwen, B. F., Ding, Y. & Heagle, A. B. Relevance of kinetochore size and microtubule-binding capacity for stable chromosome attachment during mitosis in PtK1 cells. *Chromosome Res.* **6**, 123–132 (1998).
- Bureš, P. & Zedek, F. Holokinetic drive: centromere drive in chromosomes without centromeres. *Evolution* **68**, 2412–2420 (2014).
- Kursel, L. E. & Malik, H. S. The cellular mechanisms and consequences of centromere drive. *Curr. Opin. Cell Biol.* **52**, 58–65 (2018).
- Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
- Gregory, T.R. Animal Genome Size Database <http://www.genomesize.com> (2021).
- Leitch, I.J., Johnston, E., Pellicer, J., Hidalgo, O., Bennett, M.D. Plant DNA C-values database (release 7.1, Apr 2019) <https://cvalues.science.kew.org/> (2019).
- Šmarda, P. *et al.* Genome sizes and genomic guanine+cytosine (GC) contents of the Czech vascular flora with new estimates for 1700 species. *Preslia* **91**, 117–142 (2019).
- Kullman, B., Tamm, H., Kullman, K. Fungal Genome Size Database <http://www.zbi.ee/fungal-genomesize/> (2005).
- Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S. The caper package: comparative analysis of phylogenetics and evolution in R. R Package Version 5. <https://cran.r-project.org/web/packages/caper/> (2013).
- R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (2020).
- Garamszegi, L. Z. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology* (Springer, 2014).
- Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
- Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).
- Xie, D. F. *et al.* Insights into phylogeny, age and evolution of *Allium* (Amaryllidaceae) based on the whole plastome sequences. *Ann. Bot.* **125**, 1039–1055 (2020).
- Neumann, P. *et al.* Impact of parasitic lifestyle and different types of centromere organization on chromosome and genome evolution in the plant genus *Cuscuta*. *New Phytol.* **229**, 2365–2377 (2021).

Acknowledgements

We thank Paul Talbert for his critical comments on the manuscript. We are indebted to Paul Talbert and Terri Bryson for generously providing vials with the grass CenH3 antibody. We highly appreciate the kind help of Andreas Houben and Veit Schubert with CenH3 immunolabeling in Agavoideae species. We also thank the staff from The Botanical Garden of the Faculty of Science, Masaryk University, for providing seeds from grass species. We acknowledge the core facility CELLIM of CEITEC supported by the Czech-BioImaging large RI project (LM2018129 funded by MEYS CR) for their support with obtaining scientific data presented in this paper. This work was supported by the Czech Science Foundation, grant number GA20-15989S.

Author contributions

K.P. performed karyological analyses and participated in the measurements. F.Z. and P.B. conceived of the study, participated in the measurements, and analyzed the data. K.P., P.B., and F.Z. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-99386-7>.

Correspondence and requests for materials should be addressed to F.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021