The global biogeography of angiosperm genome size is shaped by climate and range size

Petr Bureš[1], Tammy L. Elliott[1], Pavel Veselý[1], Petr Šmarda[1], Félix Forest[2], Ilia J. Leitch[2], Eimear Nic Lughadha[2], Marybel Soto Gomez[2], Samuel Pironon[2,3], Matilda J. M. Brown[2], Jakub Šmerda[1], František Zedek[1]

[1]Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlarska 2, 611 37 Brno, Czech Republic
[2]Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, United Kingdom
[3]UN Environment Programme World Conservation Monitoring Centre (UNEP-WCMC), Cambridge, Cambridgeshire, CB3 0DL, United Kingdom

correspondence: bures@sci.muni.cz, tammy.elliott@mail.mcgill.ca, fzedek@gmail.com

Total word count (Introduction+MM+Results+Discussion): 6207

Summary: 200

Introduction: 1307

Materials and Methods: 2094

Results: 967

Discussion: 1839

No. of Figures: 5 (4 in color)

No. of Tables: 2

No. of Supporting Information files: 3 (Supporting Information file containing Figs S1-S8 and Tables S1-S7; Dataset S1; Dataset S2)

Summary

(1) Evolution of genome size is shaped by various intrinsic and extrinsic factors. We explored three long-standing hypotheses concerning factors shaping the global distribution of genome size: the mutational hazard hypothesis, the polyploidy-mediated hypothesis and the climate-mediated hypothesis.

(2) We compiled the largest genome size dataset to date, encompassing >5% of known angiosperm species and analysed genome size distribution using a comprehensive geographic distribution dataset for all angiosperms across all continents.

(3) Angiosperm species with large range sizes only have small genomes, consistent with the mutational hazard hypothesis. We also uncovered a unique latitudinal pattern in the distribution of genome size diversity. Climate had a strong effect on the latitudinal pattern in genome size, while the effect of polyploidy was small. Contrary to the unimodal latitudinal patterns found for plant size traits and polyploidy, the increase in angiosperm genome sizes from the equator to 40-50°N/S is probably mediated by different (mostly climatic)

mechanisms from those underpinning the decrease in genome sizes observed from 40–50°N northwards.

(4) Overall, our global analyses highlight that species range size and climate factors are the main drivers shaping the global distribution of angiosperm genome sizes, with their relative importance varying across the latitudinal gradient.

Keywords: C-value, chromosome size, geographic range size, flowering plants, glaciation, latitudinal gradient, mutational-hazard hypothesis, nuclear DNA content, polyploid proportion, Rapoport's rule, UV-B radiation, temperature

Introduction

The most essential structure of any organism is its genome, of which the size is a relatively stable species-specific property. Geographic ranges are other inherent features of species that relate to a network of biotic and abiotic interactions shaping the evolution of species traits, including genome size. Compared to other multicellular organisms, the tremendous variation in genome sizes across angiosperms (more than 2400-fold; Pellicer *et al.,* 2018) makes them a powerful model system for studying the mechanisms that shape genome size evolution. Similarly, angiosperms are found across all continents, spanning from a majority of narrow endemics to a minority of widespread cosmopolitan species (Enquist *et al.*, 2019). The recent increase in the use of flow cytometry in botanical studies has led to a substantial accumulation of standardized genome size data across wide phylogenetic and geographic scales (Garcia *et al.*, 2014; Leitch *et al.*, 2019; Šmarda *et al.*, 2019; Zonneveld, 2019). Given consistent geographic data has recently been made available across most known species through the World Checklist of Vascular Plants (WCVP; Govaerts *et al.*, 2021), it is now possible to examine hypotheses on the causal links between angiosperms' genome size, distribution, and environment at a global scale.

A source of debate spanning decades has been whether genome size evolution is caused solely by molecular and karyological mechanisms that generate fluctuations in DNA content or if it is additionally driven by genetic drift or natural selection (Cavalier-Smith, 1982; Bennetzen & Kellogg, 1997; Petrov, 2001; Lynch & Conery, 2003; Soltis *et al.*, 2003; Bennett & Leitch, 2005; Knight *et al.*, 2005; Whitney & Garland, 2010; Whitney *et al.*, 2010; Hidalgo *et al.*, 2017; Blommaert, 2020; Faizullah *et al.*, 2021). A key proximal mechanism generating changes in genome size is the accumulation and removal of repetitive DNA (Levin, 2002; Wendel *et al.*, 2016; Lwin *et al.*, 2017), especially transposable elements (TEs), which constitute the main component of most plant genomes (Bennetzen *et al.*, 2005; Tenaillon *et al.*, 2010; Lisch, 2013; Bennetzen & Wang, 2014). Although TE insertions can occasionally have adaptive effects (Casacuberta & González, 2013), they are mostly neutral or deleterious (Deniz *et al.*, 2019). Thus, TE insertions mostly become fixed via genetic drift rather than by natural selection (Deniz *et al.*, 2019). As the relative importance of natural selection versus random genetic drift depends on population size, the *mutational hazard hypothesis* (MHH) predicts that genome growth via TEs should occur more readily in smaller populations, where genetic drift is more prominent than natural selection (i.e., species with smaller effective population sizes will have larger genomes; Lynch & Conery, 2003; Lynch, 2007). While the MHH has not been universally supported (Whitney *et al.*, 2010; Ai *et al.*, 2012; Mohlhenrich & Mueller, 2016), several studies support it (Lockton *et al.*, 2008; Bonchev & Parisod, 2013; Szitenberg *et al.*, 2016; Chen *et al.*, 2017; Lefébure *et al.*, 2017;

Elliott *et al.*, 2022). Nevertheless, many of these studies have been limited by the reliability of the parameters chosen to estimate effective population size (i.e., N_e: the number of individuals in a population that can effectively reproduce), such as mutation rate or heterozygosity (e.g., Lynch & Conery, 2003; Whitney *et al.*, 2010). In general, it is difficult to estimate effective population size reliably, and there is no single metric that has been used to do so (Wang *et al.*, 2016); this has limited our ability to test this hypothesis at a global scale. Given that population size often correlates with global geographic range size (Brown, 1984; Johnson, 1998; Gaston, 2003), and information is more widely available for species range size than for population size, coarse estimates of range size may provide us with the unique opportunity to test the MHH in influencing the global distribution of genome sizes. At a global scale, the MHH would be supported if we observed a decrease in angiosperm genome sizes with increasing range size due to the more effective elimination of deleterious TEs in species with larger effective population sizes (Fig. 1a). Moreover, if Rapoport's rule (increasing range size from south to north; Veter *et al.*, 2013) holds for angiosperms, then when combined with MHH, we would expect smaller genomes towards high latitudes.

Polyploidization is another contributor to plant genome size evolution (Wendel, 2000; Leitch & Leitch, 2008; Soltis *et al.,* 2015; Guignard *et al.*, 2016; Šmarda *et al.*, 2019), which is in turn tightly linked with chromosome number. The proportion of polyploid species at different latitudes across the globe shows a U-shape pattern (Fig. 1b), being low in the tropics and increasing polewards (Rice *et al.*, 2019). However, over time, the immediate multiplicative effect of polyploidy on both genome size and chromosome number is gradually blurred by post-polyploid diploidization processes (Mandáková & Lysák, 2018), to the point that their tight relationship disappears at deeper phylogenetic scales (i.e., over longer geological time; Roddy *et al.*, 2020). The latitudinal U-shape of polyploid proportion is likely a consequence of the similarly U-shaped distribution of the mechanisms underlying polyploid origin, for example, through the increased rate of formation of unreduced gametes in low temperatures (Ramsey & Schemske, 1998; Mason & Pires, 2015). As the relative positions of continents have remained similar over millions of years, latitudinal gradients in the mechanisms of polyploid origin should persist over geological time scales, as already documented for biodiversity (Mannion *et al*., 2014). In such a case, it might be predicted that variation in the rate of the repeated polyploidization-diploidization cycles (Wendel, 2015; Wendel *et al.*, 2016; Clark & Donoghue, 2017) would gradually lead to a U-shape latitudinal distribution of genome size, provided that genome size at least slightly increases after each cycle (Fig. 1b; *polyploid-mediated hypothesis*: PMH).

Latitudinal gradients encompass gradients of climatic and other environmental variables that could be important factors contributing to genome size variation. These factors include temperature, precipitation, aridity, seasonality, ultraviolet-B radiation (UVB), or length of the growing season (e.g., Bennett, 1976; Bennett *et al.*, 1982; Grime & Mowforth, 1982; Rayburn & Auger, 1990; MacGillivray & Grime, 1995; Bottini *et al.*, 2000; Knight & Ackerly, 2002; Grotkopp *et al.*, 2004; Dušková *et al.*, 2010; Díez *et al.,* 2013; Kang *et al.*, 2014; Du *et al.*, 2017; Bilinski *et al.*, 2018; Souza *et al.*, 2019; Becher *et al.*, 2021; Cacho *et al.*, 2021; Greimler *et al.*, 2022; Sklenář *et al.*, 2022). Studies of climatically-mediated (latitudinal or altitudinal) genome size distribution have found positive, negative, mixed, or quadratic responses of genome size to climatic gradients (reviewed in Cacho *et al.*, 2021), which may be explained by their narrow geographic and taxonomic scopes (Knight & Ackerly, 2002; Greilhuber & Leitch, 2013). Nevertheless, one pattern that often emerges from these studies is the exclusion of the largest genomes from both ends of the climatic spectrum. This may arise from the complex ways in which the biophysical constraints imposed by genome size (e.g., setting the minimum cell size and duration of mitosis and meiosis) may impact many aspects of a plant's biology, such as the timing of growth and

physiological factors such as water and nutrient use efficiency, and hence influencing where plants grow. For example, in the case of temperature, large genomes would be excluded from areas with both lowest and highest mean temperature (underpinned, in part, due to the impact of genome size on the rate of cell division) (Fig. 1c; *climate-mediated hypothesis*: CMH).

Here, we test the following hypotheses: 1) the *mutational hazard hypothesis* (MHH), which predicts that genome size decreases with increasing geographic range size; 2) the *polyploid-mediated hypothesis* (PMH), which predicts an increase in genome size from the equator to the poles; and 3) the *climate-mediated hypothesis* (CMH), which predicts the exclusion of large genomes from both ends of the climatic spectrum (Fig. 1c). We achieve this by combining the largest dataset compiled to date for genome sizes of angiosperms (16,017 species) with newly-available data on the global distribution of angiosperms from the WCVP, and mapping the global distribution of angiosperms' genome size.

Material and Methods

Taxonomic framework and geographic distribution

The angiosperm species nomenclature considered in this study follows the World Checklist of Vascular Plants (Govaerts *et al.*, 2021). We provide details of the accepted names, pertinent synonyms, and authorities for sampled taxa, as well as their WCVP 'plant_name_id' and distribution ranges based on Level‑3 Continental and Regional Codes (i.e., botanical countries) established by the International Working Group on Taxonomic Databases for Plant Sciences (TDWGs hereafter; Brummitt *et al.*, 2001) in Supporting Information Dataset S1. In addition, distribution range sizes were calculated based on the number of TDWG Level-3 regions (Dataset S1). This dataset also includes new validly-published species yet to be included in the WCVP database (marked as "NA" in the column "POWO ID" in Dataset S1), their distribution ranges converted to TDWGs, and corresponding sources. In special cases when the WCVP taxonomic framework differed from Catalogue of Life (Roskov *et al.,* 2019), World Plants (Hassler, 2022), or other sources and this difference was supported by different genome sizes, we adopted the framework congruent with genome size data (Dataset S1). We discarded taxa that were inaccurately identified (e.g., those only determined at the generic level), cultivated species with unknown native distributions, and hybrids (with the exception of a few cases where hybrid taxa have been accepted as species in some floras).

Genome size compilation

We extracted genome size estimates from several sources, including (1) research papers published between 2012 and 2022 (or older that were absent from Release 8.0 of Angiosperm DNA C-values Database) retrieved using 'Web of Science', 'ResearchGate' and 'Google Scholar' (9,515 taxa, 59.4 %); (2) the Angiosperm DNA C-values Database (5,973 taxa, 37.3 %; Release 8.0: December 2012, Bennett & Leitch, 2012; Release 9.0: April 2019, Leitch *et al.*, 2019), and (3) unpublished genome size measurements from the Plant Biosystematics Research Group of Masaryk University and the Royal Botanic Gardens, Kew (529 taxa, 3.3 %). Three different criteria were applied in cases where genome sizes for the same species were reported independently by different authors. These comprised (i) selecting values measured by flow cytometry over those estimated with Feulgen densitometry, (ii) choosing estimates from more recent reports over older ones, and (iii) assessing the taxonomic expertise of the authors for the species studied (i.e., we preferentially selected estimates from authors with taxonomic expertise in the group of interest when possible). We chose the smaller genome size (and thus the smaller DNA ploidy level) in cases where genome size

varied within a species, corresponding to different DNA ploidy levels. If there were multiple estimates presented for a species in the same publication, the genome size values were averaged. Finally, in cases where publications used a nomenclature that conflicts with the WCVP and genome size values reflected this difference, we chose an alternative taxonomic framework (predominantly the Catalogue of Life) and listed the source in Dataset S1. Genome size estimations reported in pg were converted to Mbp using the equation 1 pg = 0.978 Gbp (Doležel *et al.,* 2003).

The genome size per TDWG was calculated as the average of reported genome sizes for all taxa occurring in each region, which were $\log_{10}$-transformed (Dataset S2).

Chromosome number compilation

Chromosome numbers were extracted (in order of preference) from: (1) the same publications as genome size when both estimates were reported together; (2) the Chromosome Counts Database (CCDB: Rice *et al.,* 2015); and (3) publications reporting only chromosome number (Dataset S1). We first ensured the estimations were not pseudo-replicated, and then we selected the prevalent value for a species. We report the median value for a species when it is impossible to discern the prevailing chromosome number (e.g., in cases of aneuploidy). When chromosome numbers varied based on differing ploidy levels within a species, we compared the ploidy levels and chromosome numbers of other congeners to aid in selecting the chromosome number corresponding to the reported genome size of that species. Finally, we calculated the mean chromosome size of a species as 2C genome size (in Mbp) divided by the diploid (2n) chromosome number. This metric for mean chromosome size was used throughout the study as a correction for neopolyploidy (i.e., polyploids still recognizable cytologically rather than those with polyploidy in their ancestry recognizable only through DNA sequence analysis).

Polyploid distributions

We extracted inferred ploidy-level data from Rice *et al.* (2019: https://figshare.com/collections/The_Global_Biogeography_of_Polyploid_Plants/4306004*)*. Duplicate records and species that are not accepted in the WCVP were omitted from the dataset. We linked the remaining species to their geographic distribution based on TDWGs, as specified by the WCVP. We used the ploidy-level inferences to calculate the proportion of polyploids per TDWG (Dataset S2).

Phylogenetic tree construction

We obtained species-level phylogenetic trees for all angiosperm species from Forest *et al*. (in prep). These trees were reconstructed using the GBMB phylogenetic tree of Smith & Brown (2018) as a backbone tree. The names of the terminals in this tree were reconciled to the complete species list of the World Checklist of Vascular Plants (Govaerts *et al.*, 2021; version 6, Sept 2021). Those for which it was not possible to assign to an accepted name recognized by the WCVP were removed. Missing families were manually added to the tree based on the information available in the literature, and missing genera and species were added using the R package V.PHYLOMAKER (Jin & Qian, 2019) under its scenario 2 option. The imputation of missing species was repeated 100 times to account for the uncertainty surrounding the placement of missing species.

Growth form classification

A relationship between genome size and growth form has been suggested by many authors (e.g., Bennett, 1971; 1987; Beaulieu *et al.,* 2008; Francis *et al.,* 2008, Veselý *et al.* 2012; 2013). To control for this effect, all taxa were classified according to four plant growth forms (Dataset S1): (1) annuals (= therophytes), (2) geophytes, (3) non-geophytes (perennial herbs = hemicryptophytes + parasites + hydrophytes + epiphytes), (4) woody plants (= chamaephyte + phanerophytes), using standard floras or The World Checklist of Selected Plant Families (WCSP, 2017).

Latitude estimations

We assigned a latitude to each TDWG (Dataset S2) using their geographic centroids, determined using ArcGIS v.10 (Environmental Systems Research Institute, 2014). The latitude associated with each species (Dataset S1) was then calculated as a mean of latitudinal centroids of all the TDWGs occupied by a given species.

Climatic variables

We extracted 25 bioclimatic variables from the CHELSA database (Karger *et al.*, 2017; https://chelsa-climate.org/bioclim/; Karger *et al.*, s.a.), three ultraviolet-B-related variables from Beckmann *et al.* (2014; UVB1 = Annual Mean UV-B, UVB3 = Mean UV-B of Highest Month, and UVB5 = Sum of Monthly Mean UV-B during Highest Quarter), and the Global-Aridity Index (Global-Aridity_ET0; Trabucco & Zomer, 2022) at 30 arc-second resolution (~1km). We then calculated the mean of each variable per TDWG region (Dataset S2) with QGIS v.3.14 "pi" (QGIS Development Team, 2022). Collinearity was then assessed by calculating Pearson correlation coefficients among all pairs of the 29 variables. Correlated variables (Pearson correlation coefficient > 0.7) were assembled into six groups (Fig. S1, Table S1). To select a single variable from the six groups for further analyses, we used each variable as a predictor of 2C genome size in a polynomial regression and selected those with the best explanatory power within their groups. To select an appropriate order of the polynomials for the regression, we used the cost function combined with a visual inspection of the bivariate plots of each variable and 2C genome size. We omitted GDD0 (Growing degree days heat sum above 0°C) and Aridity index from further consideration because both explained very little variation in the regression models (R^2_{adj} = -0.002 and 0.001, respectively). Thus, the variables selected for further analyses (Table S1) were GST (Growing Season mean Temperature), BIO2 (mean diurnal air temperature range), BIO13 (precipitation of the wettest month), and BIO15 (precipitation seasonality).
Even if variables are collinear, the essence of their influence on genome size may differ (e.g., UVB-caused deletion bias vs. temperature-affected cell size). Therefore, we performed additional analyzes with selected variables that did not pass the above-mentioned filtering steps (GSL – length of the growing season, UVB1 – mean annual UVB, BIO11 – Daily mean air temperatures of the coldest quarter), if they had biological relevance or their effect on genome size had already been hypothesized.

For the purpose of the Discussion, we assessed the glaciation status of each TDWG region at the Last Glacial Maximum (LGM; ~18,000 years BP) using past climatic reconstructions from Ehlers (2015). We considered TDWG regions to be 'Glaciated' if their centroids were covered by the ice sheets during the LGM (Dataset S2).

Statistical analyses

We applied a series of linear regressions to test our three hypotheses (Fig. 1). We tested the *mutational-hazard hypothesis* (MHH) by modeling genome size as a function of range size, with both variables log-transformed (base 10) to account for the skew towards low values. We first performed ordinary least squares regression (OLS) using the function 'lm' implemented in base R, followed by phylogenetic generalized least square (PGLS) regression (Freckleton *et al.*, 2002) with the R package PHYLOLM v.2.6.2 (Ho & Ané, 2014) where we optimized branch lengths based on the data (model = 'lambda') using maximum likelihood estimation. To examine whether the association between range and genome size is dependent upon differences in genome size, we applied quantile regression analysis with nineteen different quantiles (from 0.05 to 0.95 at 0.05 intervals) using function 'rq' in the R package QUANTREG v.5.93 (Koenker *et al.,* 2022). To our knowledge, a tool is yet to be developed that is capable of performing quantile regression while correcting for evolutionary relationships among taxa. To circumvent this problem, we followed the multistep approach of Jovani *et al.* (2016). In short, we rescaled the branches of the phylogenetic tree by the phylogenetic signal value (i.e., Pagel's lambda; Pagel, 1999) that was obtained from the PGLS analysis. Then, for both genome size and range size, we calculated phylogenetic independent contrasts (Felsenstein, 1985) using the function 'pic' in the R package CAPER v.1.0.1 (Orme, 2013) based on the rescaled phylogenetic tree. When phylogenetic contrasts are supplied to standard OLS regression models and forced through the origin, the slope obtained is equivalent to the PGLS regression results (Symonds & Blomberg, 2014). Therefore, we passed the independent contrasts to the quantile regression analysis forced through the origin, which can be considered a phylogenetically controlled approach (Jovani *et al*., 2016). Because the QUANTREG function does not provide confidence intervals (CI) when the regression is forced through the origin, we derived CIs by repeating these analyses on the 100 phylogenetic trees. The 95% CI, in this case, can be interpreted as the uncertainty in the parameter estimate that is caused by the uncertainty in the phylogenetic hypothesis (Jovani *et al*., 2016).

To examine how genome size is associated with latitude, we specified genome size (log-10 transformed) as the response variable and latitude as the predictor variable in an OLS regression model. We used the cost function and the visual inspection of the bivariate plot of genome size and latitude to select the order of the polynomial fit and found that the best model was the third-degree polynomial ($\log_{10}$(Genome size)~latitude+latitude^2+latitude^3). We also performed a multiple linear regression (MLR) that included the selected bioclimatic variables (i.e., GST, BIO2, BIO13, BIO15 - see above) as predictors to evaluate the potential effects of climatic factors on the distribution of genome size across latitude. In this MLR, we specified interaction terms among all predictor variables and conducted a backward stepwise model selection based on the Akaike information criterion (AIC; Akaike, 1978) using the "step" function in base R. Based on the AICs from the backward selection process, the best model included only GST as a single predictor of 2C genome size ($\log_{10}$(Genome size)~GST+GST^2). In all MLRs with polynomials, we fitted orthogonal polynomials using the "poly" function in base R, but the "raw" parameter was set to "TRUE" to obtain parameter estimates corresponding to response variable units. Each TDWG was weighted in the regression analyses to account for the total number of species reported to occur in the region and the percentage of these species for which we have genome size or polyploid data. The weight was then calculated as "$\log_{10}$(number of species in a TDWG)" divided by the ratio of the number of species for which we have genome size data (or the proportion of polyploids): number of all species in the TDWG (Dataset S2).

Results

Sampling bias

Large datasets of phylogenetic representation and traits, including genome size data, are latitudinally biased, with northern latitudes more thoroughly sampled (Vasconcelos, 2022). To check how this may have affected our data, we compared the across-TDWG latitudinal distribution of range sizes of all angiosperms to that of taxa from our genome size dataset. Both datasets show the increasing range size from south to north (Fig. S2), in agreement with Rapoport's rule. The proportion of the four growth forms in our species genome dataset (annuals : geophytes : non-geophyte perennials : woody plants = 0.12 : 0.11 : 0.47 : 0.30) is not dramatically different from their proportions in all angiosperms (0.06 : 0.08 : 0.39 : 0.46; Govaerts *et al.*, 2021).

Genome size and range size (MHH)

Genome size and range size exhibit a triangular relationship (Fig. 2a), indicating that species with small ranges can have any genome size, while species with large ranges only have small genomes (i.e., species with large genomes do not have large range sizes). The OLS regression model based on log-transformed data (Table 1) revealed a significant decrease in genome size with increasing range size (Fig. 2b). The slope from the PGLS analysis, although still significantly negative (b = -0.021, P < 4e-7), was flatter than that from the OLS (b = -0.104, P < 2e-16), due to a strong phylogenetic signal (Pagel's λ = 0.924) in the genome size/range size relationship (Table 1, Fig. S3a). Both ordinary (Fig. 2c, Table S2) and phylogenetic (Fig. S3b, Table S3) quantile regressions showed more negative slopes for higher quantiles of genome size, indicating that the relationship between genome size and geographical range size is genome size dependent – becoming increasingly negative as genome size increases. Although the slopes started decreasing at the genome size quantile 0.8 for the ordinary quantile regression (Fig. 2c), in the phylogenetic quantile regression, the slope decreased continuously with increasing quantiles (Fig. S3b). When analyzed as 'per-TDWG averages', we also found a significant decrease in genome size with increasing range size (Fig. 2d).

We obtained very similar results when we controlled for the effect of neopolyploidy by performing the across-species analyses using mean chromosome size (2C/2n) instead of 2C genome size (Fig. S4, Tables S4-S6). However, the decrease in mean chromosome size with increasing range size was steeper than that of 2C genome size in both OLS and PGLS (Table 1, Table S4).

Genome size, neopolyploidy (PMH), latitude, and climate (CMH)

Overall, the smallest genomes occur in the tropics, and their size increases towards the poles. However, in the northern hemisphere, genome size decreases again from the temperate to the arctic regions. The global distribution of genome size averaged per TDWG is shown on the map in Fig. 3a. The genome size distribution maps of the two most species-rich eudicot (Asteraceae, Fabaceae) and monocot (Orchidaceae, Poaceae) families are shown in Fig. S5. Their genome size distribution resembles the overall trend in angiosperms. When the 2C genome size is plotted against the latitudinal centroids of TDWGs, the S-shaped pattern becomes evident (Fig. 3a). In the 3rd-order polynomial regression, latitude alone explained 40.12 % of the variation in 2C genome size (Table 2). The proportion of neopolyploid species displayed a U-shaped distribution with the smallest values in the tropics and a continuous increase in the proportion of polyploids towards the poles (Fig. 3b; Table 2).

Genome size and polyploid proportion exhibited very different latitudinal distributions (Fig. 3), and polyploid proportion explained only 1.77 % of the variation in 2C genome size (Table 2).

When we controlled for neopolyploidy by analyzing mean chromosome size across TDWGs, the S-shape latitudinal trend remained broadly unchanged (Fig. S6). The S-shaped latitudinal trend in genome size was robust to longitude, as the same pattern was recovered when the data were separately analyzed for the New and Old Worlds (Fig. S7). Herbaceous growth forms differed in the magnitude of the S-shape (Fig. 4), while woody plants did not display the S-shape, having slightly larger genomes in the tropics than in temperate or arctic regions. The same patterns for growth forms were recovered in analyses across both species (Fig. 4a) and TDWGs (Fig. 4b-e).

To assess which climatic parameters could be associated with the observed latitudinal trend in 2C genome size, we tested 29 climatic variables, but only GST (mean temperature of the growing season) was used in the final regression model based on backward selection (see Methods for details). The best-fitting model was a quadratic polynomial regression of 2C genome size on the GST (Table 2). The quadratic term had a negative coefficient, indicating that genomes are smaller in TDWGs with both high and low temperatures and larger in intermediate temperatures (Table 2; compare with the graph in Fig. 2a). The GST explained 40.75 % of the variance in 2C genome size which is all the variance explained by latitude (40.12 %; Table 2). If BIO11, which falls below 0°C in the northern hemisphere, is added into the model, the explained variance increases to 46.35 % (Table S7), highlighting the importance of freezing temperatures. Furthermore, if ‘Rapoport's rule’ (Fig. S2), i.e., MHH, is combined with CMH by adding the range size to the model with GST, the explained variance increases to 51.58 % (Table S7).

Finally, we tested whether smaller genomes are linked to shorter growing seasons. Our regression analysis showed that as the genome gets larger, the growing season (GSL) gets shorter (P=0.0004; Table S7). When analyzed only for TDWGs with latitudinal centroids of at least 48.93° (the threshold from which genomes start decreasing northward), genome size decreases with a shortening growing season, but the relationship is not significant (P=0.481; Table S7). UVB1 (mean annual UVB) explained 34.6 % of the variation in mean genome size across TDWGs (Table S7).

Discussion

We compiled the largest genome size dataset to date, encompassing >5% of known angiosperm species (Dataset S1), and explored three hypotheses concerning the global distribution of genome size in angiosperms (Fig. 1). We found support for the mutational hazard hypothesis (MHH). We also found that polyploidy (PMH) has only a small impact, whereas climate (CMH) has a strong effect on the latitudinal pattern in genome sizes. Factors shaping the global biogeography of genome size are summarized in Fig. 5.

Support for the MHH: genome size and range size are inversely related

As predicted by the MHH (Lynch & Connery, 2003; Lynch, 2007), we found a negative association between range size and genome size (Fig. 2). Considering range size as a proxy for population size (Brown, 1984; Johnson, 1998), genetic drift is predicted to become more important relative to natural selection in range-restricted species according to the MHH. Thus, the relative importance of purifying selection pressure on keeping genome size small should increase from the south pole to the north pole, along with increasing range size (as per ‘Rapoport's rule’; Fig. S2). Hence, the largest genomes are found in the southern hemisphere

(Fig. 3a), where angiosperms have the smallest ranges and are thus most susceptible to genetic drift (Fig. S2).

Moreover, we found that the negative association between range size and genome size is accentuated as genome sizes increase (Fig. 2c). While genetic drift may be an important neutral process underlying genome growth in range-restricted species, purifying selection may strongly select against large genomes, as these may have disadvantageous nucleotypic effects on a plant's anatomy and physiology. This idea is compatible with the 'large-genome-constraint' hypothesis, which postulates that large genomes might be maladaptive (Vinogradov, 2003; Knight *et al.,* 2005). For example, large-genomed plants are restricted to having large seeds (Knight & Ackerly, 2002; Beaulieu *et al.,* 2007; Carta *et al.,* 2022) and stomata (Beaulieu *et al.,* 2008; Veselý *et al.,* 2012) because a larger genome occupies more volume (Cavalier-Smith, 2005). Large seeds and stomata might limit species' dispersal ability (Sonkoly *et al*., 2017) and narrow their ecological niche, as large stomata are detrimental in drier conditions (Veselý *et al.*, 2020), leading to smaller geographical ranges (Sheth *et al.*, 2020). Notably, the most pronounced S-shape in the latitudinal distribution of genome size is in geophytes (Fig. 4c), whose genomes are the largest among the analyzed growth forms (Fig. 4a). Associations between genome size, seed size, and ecological niche might be reinforced by the fact that species with larger genomes have greater extinction risk (Vinogradov, 2003; LaBar & Adami, 2020; Soto Gomez *et al.,* in prep).

Small genomes in the tropics

The decrease in genome size from temperate to tropical regions across both hemispheres is consistent with previous studies focusing on genome size (or its proxies) in Poaceae (Avdulov, 1931; Bennett, 1976), Fabaceae (Stebbins, 1966; Bennett, 1976; Souza *et al.*, 2019), Orchidaceae (Trávníček *et al.*, 2019), Zygophyllaceae (Vidal-Russell *et al.*, 2022) and across angiosperms more generally (Levin & Funderburg, 1979; Yu *et al.,* 2018). In our study, the environmental variables most correlated with latitude were temperature-related, and peaked in the tropics (Fig. 3a). In higher ambient temperatures, metazoan ectotherms, unicellular eukaryotes and prokaryotes tend to have smaller cells (Atkinson *et al.*, 2003; Hessen *et al.*, 2013; Sabath *et al.*, 2013), likely because maintenance of large cells becomes more difficult with increasing temperatures (Sabath *et al.*, 2013). Our finding of small genomes might reflect this relationship, where it is advantageous to have smaller cells (and thus a smaller genome; Cavalier-Smith, 2005) in the tropics.

Alternatively, small genomes in the tropics could be a result of the deleterious DNA-damaging effects of UV-B radiation (Bennett, 1976), which is generally highest in the tropics (Beckmann *et al*., 2014) and might select for smaller chromosomes that absorb less energy, therefore decreasing radiosensitivity (Sparrow *et al.*, 1967). Recent findings show that plants with holocentric chromosomes, which tolerate fragmentation (Zedek & Bureš, 2019), are less stressed (Zedek *et al.,* 2020; Zedek *et al.,* 2021) and more competitive (Zedek *et al.,* 2022) under higher UV-B doses support this idea. Moreover, homologous recombination used to repair UV-B-induced damage might increase deletion rates, thereby further promoting genome downsizing (Schubert & Vu, 2016). However, as UV-B radiation intensity (explained variation 34.6 %) is strongly correlated with temperature (Fig. S1), the individual effects of these two factors on genome size in the tropics cannot be easily differentiated.

Finally, nutrient limitation might constrain genome size in tropical plants, as many tropical soils are low in nutrients (especially available phosphorus; Vitousek *et al.*, 2010),

possibly limiting the amount of DNA in a plant's genome (Leitch & Leitch, 2008; Šmarda *et al.*, 2013, Guignard *et al.*, 2016; Faizullah *et al*., 2021; Veleba *et al.*, 2020).

Genome size decreases from temperate regions towards the pole, but only in the north

Hemisphere-scale differences in genome size trends across latitudinal gradients may be explained by climate; the northern hemisphere is more continental, with a higher thermal gradient than the more equable maritime southern hemisphere. If genome sizes in plants followed a similar pattern to polyploid proportion, genomes would be expected to be smaller near the equator and increase polewards. We find that this trend holds in the southern hemisphere, it only holds until a certain limiting low temperature in the Northern Hemisphere, beyond which genome sizes decrease towards the high northern latitudes. Such a latitudinal breakpoint was predicted by Bennett *et al.,* (1982). In the southern hemisphere, this low-temperature threshold is probably not reached (Fig. 1a). The main drivers of selection pressure against larger genomes in polar regions are predicted to be: (i) slower cell division mediated by lower temperatures (Francis & Barlow, 1988) and (ii) longer generation time mediated by lower temperature or shorter growing seasons (Bennett *et al.*, 1982; Bennett, 1987). Indeed, temperature variables alone explain a relatively large proportion (up to ~40 %) of genome size variation (Table 2, Table S1).

Several authors hypothesized that the decrease in temperatures toward the poles leads to higher production of unreduced gametes in plants (Belling, 1925; Sakamura & Stow, 1926; de Mol, 1928; Heilborn, 1930; Hagerup, 1932; Bretagnolle & Thompson, 1995; Mason & Pires, 2015; Kreiner *et al.*, 2017). If this phenomenon explains the increase in the proportion of polyploidy from the equator to the poles (Fig. 3b; Rice *et al.*, 2019), then polyploid proportion should be significantly higher in the northern hemisphere, where temperatures reach lower values. However, neither our study nor that by Rice *et al.* (2019) found such a difference in polyploid proportion between the hemispheres (Fig. 3b), suggesting that unreduced gamete production might not be important in explaining latitudinal variation in polyploidy (but see below).

In the northern hemisphere, decreasing genome size from temperate regions to the pole might be mediated by lower frost resistance in species with large genomes as their larger cells (Beaulieu *et al*., 2008) contain greater volumes of cytoplasmic water, which would make them more sensitive to frost (Sakai & Larcher, 1987; Zachariassen & Kristiansen, 2000). The decrease in genome size in the northern hemisphere from temperate regions to the Arctic could also be related to glaciation cycles, as smaller-genomed species tend to occur in previously glaciated TDWGs (Fig. S8). During glacial migrations, species with large genomes might have increased extinction risk because of their smaller range sizes, as suggested by the negative association between geographic range size and genome size (Fig. 2a). Similarly, repeated glaciation cycles could have led to the extinction of some (neo)polyploids whose genome sizes exceeded an upper selection limit, which could further explain why polyploid proportion is not higher in the northern than the southern hemisphere. In this case, the hypothesis relating the increased formation of unreduced gametes to low temperatures and its role in increasing the proportion of polyploidy from tropical to polar regions would still be relevant. A further possibility explaining the decrease in genome size from the northern temperate to polar regions is that shorter growing seasons towards high latitudes might be important in selecting plants with smaller genomes, which have faster growth rates and can complete their growth cycles in less time (Knight *et al.,* 2005). However, our results showed that if the length of the growing season in the Arctic has any effect on genome size, it is likely minor (Table S7).

Plant growth forms differ in genome size; for example, bulb- or tuber-forming geophytes generally have larger genomes (Veselý *et al.*, 2012; 2013; Šmarda *et al.*, 2019), while annual or woody angiosperms have smaller genomes (Bennett, 1972; 1987; Knight & Beaulieu, 2008; Bennett & Leitch, 2012; Šmarda *et al.*, 2019). Thus, the change in average genome size along a latitudinal gradient might simply reflect a shift in the representation of individual growth forms. For example, in the Arctic, the prevalence of permafrost (associated with cryoturbation and solifluction: Ping *et al.*, 2015) limits the occurrence of geophytes (Klimešová *et al.*, 2012; Lubbe *et al.*, 2021), which may partly explain the decrease in genome sizes from northern temperate to polar regions. However, geophytes, annuals, and non-geophyte perennial herbs all display decreases in genome size in the Arctic, similar to that of all flowering plants, and their genome size response curves are similar across the whole latitudinal gradient (Fig. 4). Woody plants show a different pattern, but their genome size still decreases northward (Fig. 4). Woody angiosperms are seldom polyploid (Müntzing, 1936; Stebbins, 1940; Otto & Whitton, 2000; Zenil-Fergusson *et al*., 2017; Rice *et al.*, 2019), which could explain the similarity of their genomes sizes in tropical and temperate regions. Overall, the gradient in genome sizes cannot be explained solely by latitudinally-mediated changes in the proportion of different growth forms. Instead, it is likely to have arisen from a diversity of extrinsic (climatic) and intrinsic factor(s) impacting genome size across the global distribution of angiosperms.

Relatively large genomes in temperate regions

In winter or early spring, many early-flowering temperate perennials grow by enlarging cells that divided during the preceding growing season (Grime & Mowforth, 1982; Schnablová *et al.*, 2021). This strategy relaxes selection against larger genomes (Schnablová *et al.*, 2021) because the expansive growth is not bound to cell division, which is slow for large genomes (Bennett, 1971) and slowed down or halted in low temperatures (Francis & Barlow, 1988). The seasonal climate thus enables the existence of both small and large genomes that could result in increased variation in genome size in temperate zones of both hemispheres (Fig. 4a).

Future directions

Species range size and climate are key to shaping the global distribution of genome sizes in angiosperms. Our global scale study showed that contradictory results from earlier studies make sense when viewed on a global scale. Experimental approaches and detailed studies focusing on the distribution of genome sizes across plant communities along altitudinal gradients might be valuable in determining the proximal causes responsible for filtering smaller genome sizes at certain latitudes, especially in tropical regions where the altitudinal zonation of vegetation is the widest. Further research could also be directed at determining the roles of various factors underpinning the global distribution of genome sizes, such as glaciation, unreduced gametes, growth forms, and UV-B radiation compared to climatic variables.

Author contributions

PB, TLE, and FZ designed the study, performed the analyses and drafted the first version of the manuscript. PB and FZ collected the genome size data. PV and PB prepared the datasets for analyses, PV assigned growth forms and contributed to the analyses. FF prepared phylogenetic trees. JŠ performed flow cytometric measurement of unpublished data used in

this study. PŠ, FF, IJL, ENL, MSG, SP, and MJMB contributed to the analyses, interpretation of the results and the final form of the manuscript.

Acknowledgments

We thank Rafaël Govaerts for providing data from the World Checklist of Vascular Plants.

Funding

This work was financially supported by the Czech Science Foundation, grant no. GA20-15989S.

Conflict of Interest Declaration

The authors have no conflict to declare.

References

Ai B, Wang ZS, Ge S. 2012. Genome size is not correlated with effective population size in the *Oryza* species. *Evolution* **66:** 3302–3310. doi: 10.1111/j.1558-5646.2012.01674.x.

Akaike H. 1978. A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics* **30**: 9–14. doi: 10.1007/BF02480194.

Atkinson D, Ciotti BJ, Montagnes DJ. 2003. Protists decrease in size linearly with temperature: ca. 2.5% degrees C(-1). *Proceedings of the Royal Society B: Biological Sciences* **270**: 2605–2611. doi: 10.1098/rspb.2003.2538.

Avdulov NP. 1931. Karyo-systematische Untersuchungen der Familie Gramineen. *Bulletin of Applied Botany, of Genetics and Plant Breeding, Leningrad* **44**: 1–428. (Russian with German summary).

Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA. 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist* **179**: 975–986. doi: 10.1111/j.1469-8137.2008.02528.x.

Beaulieu JM, Moles AT, Leitch IJ, Bennett MD, Dickie JB, Knight CA. 2007. Correlated evolution of genome size and seed mass. *New Phytologist* **173**: 422–437. doi: 10.1111/j.1469-8137.2006.01919.x.

Beckmann M, Václavík T, Manceur AM, Šprtová L, von Wehrden H, Welk E, Cord AF. 2014. glUV: a global UV-B radiation data set for macroecological studies. *Methods in Ecology and Evolution* **5**: 372–383. doi: 10.1111/2041-210X.12168.

Becher H, Powell RF, Brown MR, Metherell C, Pellicer J, Leitch IJ, Twyford AD. 2021. The nature of intraspecific and interspecific genome size variation in taxonomically complex eyebrights. *Annals of Botany* **128:** 639–651. doi: 10.1093/aob/mcab102.

Belling J. 1925. The origin of chromosomal mutations in *Uvularia*. *Journal of Genetics* **15:** 245–266. doi: 10.1007/BF02983121.

Bennett MD, Leitch IJ. 2005. Genome size evolution in plants. In: TR Gregory, ed. *The evolution of genome.* San Diego: Elsevier, 89–162.
Bennett MD. 1971. The duration of meiosis. *Proceedings of the Royal Society B: Biological Sciences* **178**: 277–299. doi: 10.1098/rspb.1971.0066.
Bennett MD. 1976. DNA amount, latitude, and crop plant distribution. *Environmental and Experimental Botany* **16:** 93–108. doi: 10.1016/0098-8472(76)90001-0.
Bennett MD. 1987. Variation in genomic form in plants and its ecological implications. *New Phytologist* **106**: 177–200. doi: 10.1111/j.1469-8137.1987.tb04689.x.
Bennett MD, Leitch IJ. 2012. Plant DNA C-values Database (Release 8.0) https://cvalues.science.kew.org/
Bennett MD, Smith JB, Lewis-Smith RI. 1982. DNA amounts of angiosperms from the Antarctic and South Georgia. *Environmental and Experimental Botany* **22:** 307–318. doi: 10.1016/0098-8472(82)90023-5.
Bennetzen JL, Kellogg EA. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**: 1509–1514. doi: 10.1105/tpc.9.9.1509.
Bennetzen JL, Ma J., Devos KM. 2005. Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95:** 127–132. doi: 10.1093/aob/mci008.
Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function and evolution of plant genomes. *Annual Review of Plant Biology* **65:** 505–530. doi: 10.1146/annurev-arplant-050213-035811.
Bilinski P, Albert PS, Berg JJ, Birchler JA, Grote MN, Lorant A, Quezada J, Swarts K, Yang J, Ross-Ibarra J. 2018. Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genetics* **14:** e1007162. doi: 10.1371/journal.pgen.1007162.
Blommaert J. 2020. Genome size evolution: towards new model systems for old questions. *Proceedings of the Royal Society B: Biological Sciences* **287**: 20201441. doi:10.1098/rspb.2020.1441.
Bonchev G, Parisod C. 2013. Transposable elements and microevolutionary changes in natural populations. *Molecular Ecology Resources* **13:** 765–775. doi: 10.1111/1755-0998.12133.
Bottini MCJ, Greizerstein EJ, Aulicino MB, Poggio L. 2000. Relationships among genome size, environmental conditions and geographical distribution in natural populations of NW Patagonian species of *Berberis* L. (Berberidaceae). *Annals of Botany* **86:** 565–573. doi: 10.1006/anbo.2000.1218.
Bretagnolle F, Thompson JD. 1995. Gametes with the somatic chromosome number: Mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytologist* **129:** 1–22. doi: 10.1111/j.1469-8137.1995.tb03005.x.
Brown JH. 1984. On the relationship between abundance and distribution of species. *American Naturalist* **124:** 255–279. doi: 10.1086/284267.
Brummitt RK, Pando F, Hollis S, Brummitt N. 2001. *World geographical scheme for recording plant distributions*. Pittsburgh, Pennsylvania: Hunt Institute for Botanical Documentation, Carnegie-Mellon University.
Cacho NI, McIntyre PJ, Kliebenstein DJ, Strauss SY. 2021. Genome size evolution is associated with climate seasonality and glucosinolates, but not life history, soil nutrients or range size, across a clade of mustards. *Annals of Botany* **127:** 887–902. doi: 10.1093/aob/mcab028.
Carta A, Mattana E, Dickie J, Vandelook F. 2022. Correlated evolution of seed mass and genome size varies among life forms in flowering plants. *Seed Science Research* **32**: 46–52. doi: 10.1017/ S0960258522000071.

Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. *Molecular Ecology* **22**: 1503–1517. doi: 10.1111/mec.12170.
Cavalier-Smith T. 1982. Skeletal DNA and the evolution of genome size. *Annual Review of Biophysics and Bioengineering* **11:** 273–302. doi: 10.1146/annurev.bb.11.060182.001421.
Cavalier-Smith T. 2005. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Annals of Botany* **95**: 147–175. doi: 10.1093/aob/mci010.
Chen J, Glémin S, Lascoux M. 2017. Genetic diversity and the efficacy of purifying selection across plant and animal species. *Molecular Biology and Evolution* **34:** 1417–1428. doi: 10.1093/molbev/msx088.
Clark JW, Donoghue PCJ. 2017. Constraining the timing of whole genome duplication in plant evolutionary history. *Proceedings of the Royal Society B: Biological Sciences* **284:** 20170912. doi: 10.1098/rspb.2017.0912.
de Mol, W.E. 1928. The originating of diploid and tetraploid pollen-grains in Duc van Thol-Tulips (*Tulipa suaveolens*) dependent on the method of culture applied. *Genetica* **11:** 119–212. doi: 10.1007/BF01726318.
Deniz Ö, Frost JM, Branco MR. 2019. Regulation of transposable elements by DNA modifications. *Nature Reviews Genetics* **20**: 417–431. doi: 10.1038/s41576-019-0106-6.
Díez CM, Gaut BS, Meca E, Scheinvar E, Montes-Hernandez S, Eguiarte LE, Tenaillon MI. 2013. Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist* **199**: 264–276. doi: 10.1111/nph.12247.
Doležel J, Bartoš J, Voglmayr H, Greilhuber J. 2003. Nuclear DNA content and genome size of trout and human. *Cytometry* **51:** 127–128. doi: 10.1002/cyto.a.10013.
Du YP, Bi Y, Zhang MF, Yang FP, Jia GX, Zhang XH. 2017. Genome size diversity in *Lilium* (Liliaceae) is correlated with karyotype and environmental traits. *Frontiers in Plant Science* **8:** 1303. doi: 10.3389/fpls.2017.01303.
Dušková E, Kolář F, Sklenář P, Rauchová J, Kubešová M, Fér T, Suda J, Marhold K. 2010. Genome size correlates with growth form, habitat and phylogeny in the Andean genus *Lasiocephalus* (Asteraceae). *Preslia* **82:** 127–148.
Elliott TL, Zedek F, Barrett B, Bruhl J, Escudero M, Hroudová Z, Joly S, Larridon I, Luceño M, Márquez-Corro JI, Martin-Bravo S, Muasya AM, Šmarda P, Thomas WW, Wilson K, Bures P. 2022. Chromosome size matters: Genome evolution in the cyperid clade. *Annals of Botany,* in press. doi: 10.1093/aob/mcac136.
Enquist BJ, Feng X, Boyle B, Maitner B, Newman EA, Jørgensen PM, Roehrdanz PR, Thiers BM, Burger JR, Corlett RT, Couvreur TLP, Dauby G, Donoghue JC, Foden W, Lovett JC, Marquet PA, Merow C, Midgley G, Morueta-Holme N, Neves DM, Oliveira-Filho AT, Kraft NJB, Park DS, Peet RK, Pillet M, Serra-Diaz JM, Sandel B, Schildhauer M, Šímová I, Violle C, Wieringa JJ, Wiser SK, Hannah L, Svenning JC, McGill BJ. 2019. The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances* **5**: eaaz0414. doi: 10.1126/sciadv.aaz0414.
Environmental Systems Research Institute (ESRI). 2014. ArcGIS Release 10. Redlands, CA.
Faizullah L, Morton JA, Hersch-Green EI, Walczyk AM, Leitch AR, Leitch IJ. 2021. Exploring environmental selection on genome size in angiosperms. *Trends in Plant Science* **26:** 1039–1049. doi: 10.1016/j.tplants.2021.06.001.
Felsenstein J. 1985. Phylogenies and the comparative method. *American Naturalist* **125**: 1–15. doi: 10.1086/284325.

Francis D, Barlow PW. 1988. Temperature and the cell cycle. *Symposia of the Society for Experimental Biology* **42**: 181–201.
Francis D, Davies MS, Barlow PW. 2008. A strong nucleotypic effect on the cell cycle regardless of ploidy level. *Annals of Botany* **101:** 747–757. doi: 10.1093/aob/mcn038.
Freckleton R, Harvey P, Pagel M. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. *American Naturalist* **160:** 712–726. doi: 10.1086/343873.
Garcia S, Leitch IJ, Anadon-Rosell A, Canela MÁ, Gálvez F, Garnatje T, Gras A, Hidalgo O, Johnston E, Mas de Xaxars G, Pellicer J, Siljak-Yakovlev S, Vallès J, Vitales D, Bennett MD. 2014. Recent updates and developments to plant genome size databases. *Nucleic Acids Research* **42(D1):** D1159-D1166. doi: 10.1093/nar/gkt1195.
Gaston KJ, Blackburn TM. 1996. Global scale macroecology: Interactions between population size, geographic range size and body size in the Anseriformes. *Journal of Animal Ecology* **65:** 701–714. doi: 10.2307/5669.
Gaston KJ, Spicer JI. 2001. The relationship between range size and niche breadth: A test using five species of *Gammarus* (Amphipoda). *Global Ecology and Biogeography* **10:** 179–188. doi: 10.1046/j.1466-822x.2001.00225.x.
Gaston KJ. 2003. *The Structure and Dynamics of Geographic Ranges*. New York, USA: Oxford University Press.
Govaerts R, Nic Lughadha E, Black N, Turner R, Paton A. 2021. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Scientific Data* **8**: 215. doi: 10.1038/s41597-021-00997-6.
Gregory TR. 2022. *Animal Genome Size Database*. http://www.genomesize.com (accessed 2/Oct/2022).
Greilhuber J, Leitch IJ. 2013. Genome size and the phenotype. In: J Greilhuber, J Doležel, JF Wendel, eds. *Plant genome diversity, Vol. 2: Physical structure, behaviour and evolution of plant genomes.* Wien: Springer, 323–344.
Greimler J, Temsch EM, Xue Z, Weiss-Schneeweiss H, Volkova P, Peintinger M, Wasowicz P, Shang H, Schanzer I, Chiapella JO. 2022. Genome size variation in *Deschampsia cespitosa* sensu lato (Poaceae) in Eurasia. *Plant Systematics and Evolution* **308:** 9. doi: 10.1007/s00606-021-01796-7.
Grime JP. 1998. Plant classification for ecological purposes: Is there a role for genome size? *Annals of Botany* **82:** 117–120. doi: 10.1006/anbo.1998.0723.
Grime JP, Mowforth MA. 1982. Variation in genome size—an ecological interpretation. *Nature* **299**: 151–153. doi: 10.1038/299151a0.
Grotkopp E, Rejmánek M, Sanderson MJ, Rost TL. 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. *Evolution* **58:** 1705–1729. doi: 10.1111/j.0014-3820.2004.tb00456.x.
Grzywacz B, Skórka P. 2021. Genome size versus geographic range size in birds. *PeerJ* **9**: e10868. doi: 10.7717/peerj.10868.
Guignard MS, Nichols RA, Knell RJ, Macdonald A, Romila CA, Trimmer M, Leitch IJ, Leitch AR. 2016. Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. *New Phytologist* **210**: 1195–1206. doi: 10.1111/nph.13881.
Gustafsson A. 1948. Polyploidy, life form and vegetative reproduction. *Hereditas* **34:** 1–22. doi: 10.1111/j.1601-5223.1948.tb02824.x.
Hadfield JD. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* **33:** 1–22. doi: 10.18637/jss.v033.i02.
Hagerup O. 1932. Über Polyploidie in Beziehung zu Klima, Ökologie und Phylogenie. *Hereditas* **16:** 19–40. doi: 10.1111/j.1601-5223.1932.tb02560.x.

Hassler M. 2022. World Plants. Synonymic Checklist and Distribution of the World Flora. Version 13.2; last update July 2nd, 2022. - www.worldplants.de.
Heilborn, O. 1930. Temperature und Chromosomenkonjugation. *Svensk Botanisk Tidskrift* **24:** 12–24.
Hessen DO, Daufresne M, Leinaas HP. 2013. Temperature-size relations from the cellular-genomic perspective. *Biological Reviews of the Cambridge Philosophical Society* **88**: 476–489. doi: 10.1111/brv.12006.
Hidalgo O, Pellicer J, Christenhusz M, Schneider H, Leitch AR, Leitch IJ. 2017. Is there an upper limit to genome size? *Trends in Plant Science* **22:** 567–573. doi: 10.1016/j.tplants.2017.04.005.
Ho LST, Ane C .2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology* **63**: 397–408. doi: 10.1093/sysbio/syu005.
Jin Y, Qian H. 2019. V. PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography* **42**: 1353–1359. doi: 10.1111/ecog.04434.
Johnson AW, Packer JG. 1965. Polyploidy and environment in arctic Alaska. *Science* **148:** 237–239. doi: 10.1126/science.148.3667.237.
Johnson C. 1998. Species extinction and the relationship between distribution and abundance. *Nature* **394:** 272–274. doi: 10.1038/28385.
Jovani R, Lascelles B, Garamszegi LZ, Mavor R, Thaxter CB, Oro D. 2016. Colony size and foraging range in seabirds. *Oikos* **125**: 968–974. doi: 10.1111/oik.02781.
Kang M, Tao J, Wang J, Ren C, Qi Q, Xiang QY, Huang H. 2014. Adaptive and nonadaptive genome size evolution in Karst endemic flora of China. *New Phytologist* **202:** 1371–1381. doi: 10.1111/nph.12726.
Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder P, Kessler M. 2017. Climatologies at high resolution for the Earth land surface areas. *Scientific Data* **4**: 170122. doi: 10.1038/sdata.2017.122.
Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder P, Kessler M. s. a. Data from: Climatologies at high resolution for the earth's land surface areas. *Dryad Digital Repository*, doi:10.5061/dryad.kd1d4
Klimešová J, Doležal J, Prach K, Košnar J. 2012. Clonal growth forms in Arctic plants and their habitat preferences: A study from Petuniabukta, Spitsbergen. *Polish Polar Research* **33:** 421–442. doi: 10.2478/v1018 3-012-0019-y.
Knight CA, Ackerly DD. 2002. Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecology Letters* **5:** 66–76. doi: 10.1046/j.1461-0248.2002.00283.x.
Knight CA, Beaulieu JM. 2008. Genome size scaling through phenotype space. *Annals of Botany* **101:** 759–766. doi: 10.1093/aob/mcm321.
Knight CA, Molinari NA, Petrov DA. 2005. The large genome constraint hypothesis: Evolution, ecology and phenotype. *Annals of Botany* **95:** 177–190. doi: 10.1093/aob/mci011.
Koenker R, Portnoy S, Ng PT, Melly B, Zeileis A, Grosjean P, Moler C, Saad Y, Chernozhukov V, Fernandez-Val I, Ripley BD. 2022. Package 'quantreg'. Cran R-project. org.
Kreiner JM, Kron P, Husband BC. 2017. Evolutionary Dynamics of Unreduced Gametes. *Trends in Genetics* **33:** 583–593. doi: 10.1016/j.tig.2017.06.009.
LaBar T, Adami C. 2020. Genome Size and the Extinction of Small Populations. In: Banzhaf W, Cheng BHC, Deb K, Holekamp KE, Lenski RE, Ofria C, Pennock RT, Punch WF, Whittaker DJ, eds. *Evolution in Action: Past, Present and Future. Genetic and*

Evolutionary Computation. Springer, Cham., 167–183. https://doi.org/10.1007/978-3-030-39831-6_14
Lefébure T, Morvan C, Malard F, François C, Konecny-Dupré L, Guéguen L, Weiss-Gayet M, Seguin-Orlando A, Ermini L, Sarkissian C, Charrier NP, Eme D, Mermillod-Blondin F, Duret L, Vieira C, Orlando L, Douady CJ. 2017. Less effective selection leads to larger genomes. *Genome Research* **27:** 1016–1028. doi: 10.1101/gr.212589.116.
Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* **320**: 481–483. doi: 10.1126/science.1153585.
Leitch IJ, Johnston E, Pellicer J, Hidalgo O, Bennett MD. 2019. Angiosperm DNA C-values Database (Release 9.0) https://cvalues.science.kew.org/
Levin DA, Funderburg SW. 1979. Genome size in angiosperms: Temperate versus tropical species. *American Naturalist* **114:** 784–795. doi: 10.1086/283528.
Levin DA. 2002. *The role of chromosomal change in plant evolution.* Oxford: Oxford University Press.
Lisch D. 2013. How important are transposons for plant evolution? *Nature Review Genetics* **14:** 49–61. doi: 10.1038/nrg3374.
Lockton S, Ross-Ibarra J, Gaut BS. 2008. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences USA* **105:** 13965–13970. doi: 10.1073/pnas.0804671105.
Lubbe FC, Klimešová J, Henry HAL. 2021. Winter belowground: Changing winters and the perennating organs of herbaceous plants. *Functional Ecology* **35:** 1627–1639. doi: 10.1111/1365-2435.13858.
Lwin AK, Bertolini E, Pè ME, Zuccolo A. 2017. Genomic skimming for identification of medium/highly abundant transposable elements in *Arundo donax* and *Arundo plinii*. *Molecular Genetics and Genomics* **292:** 157–171. doi: 10.1007/s00438-016-1263-3.
Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* **302**: 1401–1404. doi: 10.1126/science.1089370.
Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences USA* **104**: 8597–8604. doi: 10.1073/pnas.0702207104.
MacGillivray CW & Grime JP. 1995. Genome size predicts frost resistance in British herbaceous plants: Implications for rates of vegetation response to global warming. *Functional Ecology* **9:** 320. doi: 10.2307/2390580.
Mandáková T, Lysak MA. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology* **42:** 55–65. doi: 10.1016/j.pbi.2018.03.001.
Mannion PD, Upchurch P, Benson RB, Goswami A. 2014. The latitudinal biodiversity gradient through deep time. *Trends in Ecology Evolution* **29**: 42–50. doi: 10.1016/j.tree.2013.09.012.
Mason AS, Pires CJ. 2015. Unreduced gametes: meiotic mishap or evolutionary mechanism? *Trends in Genetics* **31:** 5–10. doi: 10.1016/j.tig.2014.09.011.
Mohlhenrich ER, Mueller RL. 2016. Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. *Evolution* **70:** 2865–2878. doi: 10.1111/evo.13084.
Müntzing A. 1936. The evolutionary significance of autopolyploidy. *Hereditas* **21:** 263–378. doi: 10.1111/j.1601-5223.1936.tb03204.x.
Orme D. 2013. The caper package: comparative analysis of phylogenetics and evolution in R. https://cran.r-project.org/web/packages/caper/index.html.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* **34:** 401–437. doi: 10.1146/annurev.genet.34.1.401.
Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884. doi: 10.1038/44766.
Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ. 2018. Genome size diversity and its impact on the evolution of land plants. *Genes* **9:** 88. doi: 10.3390/genes9020088.
Petrov DA. 2001. Evolution of genome size: new approaches to an old problem. *Trends in Genetics* **17:** 23–28. doi: 10.1016/s0168-9525(00)02157-0.
Ping CL, Jastrow JD, Jorgenson MT, Michaelson GJ, Shur YL. 2015. Permafrost soils and carbon cycling. *Soil* **1**: 147–171. doi.org/10.5194/soil-1-147-201.
QGIS Development Team. 2022. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.
Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29:** 467–501. doi: 10.1146/annurev.ecolsys.29.1.467.
Rayburn AL, Auger JA. 1990. Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. *Theoretical and Applied Genetics* **79:** 470–474. doi: 10.1007/BF00226155.
Rice A, Glick G, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. 2015. The chromosome counts database (CCDB) – a community resource of plant chromosome numbers. *New Phytologist* **206:** 16–26. doi: 10.1111/nph.13191.
Rice A, Šmarda P, Novosolov M, Drori M, Glick L, Sabath N, Meiri S, Belmaker J, Mayrose I. 2019. The global biogeography of polyploid plants. *Nature Ecology & Evolution* **3:** 265–273. doi: 10.1038/s41559-018-0787-9.
Roddy AB, Théroux-Rancourt G, Abbo T, Benedetti JW, Brodersen CR, Castro M, Castro S, Gilbride AB, Jensen B, Jiang G-F, Perkins JA, Perkins SD, Loureiro J, Syed Z, Thompson RA, Kuebbing SE, Simonin KA. 2020. The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. *International Journal of Plant Science* **181:** 75–87. doi: 10.1086/706186.
Roskov Y, Ower G, Orrell T, Nicolson D, Bailly N, Kirk PM, Bourgoin T, DeWalt RE, Decock W, Nieukerken E van, Zarucchi J, Penev L, eds. 2019. Species 2000 & ITIS Catalogue of Life, 2019 Annual Checklist. Digital resource at www.catalogueoflife.org/annual-checklist/2019. Species 2000: Naturalis, Leiden, the Netherlands. ISSN 2405-884X.
Sabath N, Ferrada E, Barve A, Wagner A. 2013. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biology and Evolution* **5**: 966–977. doi: 10.1093/gbe/evt050.
Sakai A, Larcher W. 1987 *Frost survival of plant. Responses and adaptations to freezing stress*. Springer, Berlin.
Sakamura T, Stow I. 1926. Über die experimentell veranlasste Entstehung von keimfähigen Pollenkörnern mit abweichenden Chromosomenzahlen. *Japanese Journal of Botany* **3:** 111–137.
Schnablová R, Huang L, Klimešová J, Šmarda P, Herben T. 2021. Inflorescence preformation prior to winter: a surprisingly widespread strategy that drives phenology of temperate perennial herbs. *New Phytologist* **229**: 620–630. doi: 10.1111/nph.16880.
Schubert I, Vu GTH. 2016. Genome stability and evolution: Attempting a holistic view. *Trends in Plant Science* **21:** 749–757. doi: 10.1016/j.tplants.2016.06.003.
Sheth SN, Morueta-Holme N, Angert AL. 2020. Determinants of geographic range size in plants. *New Phytologist* **226**: 650–665. doi: 10.1111/nph.16406.

Sklenář P, Ptáček J, Klimeš A. 2022. Genome size of alpine plants does not predict temperature resistance. *Planta* **256:** 18. doi: 10.1007/s00425-022-03935-x.
Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* **105**: 302–314. doi: 10.1002/ajb2.1019.
Soltis DE, Soltis PS, Bennett MD, Leitch IJ. 2003. Evolution of genome size in the angiosperms. *American Journal of Botany* **90:** 1596–1603. doi: 10.3732/ajb.90.11.1596.
Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics and Development* **35**: 119–25. doi: 10.1016/j.gde.2015.11.003.
Sonkoly J, Deák B, Valkó O, Molnár VA, Tóthmérész B, Török P. 2017. Do large-seeded herbs have a small range size? The seed mass–distribution range trade-off hypothesis. *Ecology and Evolution* **7:** 11204–11212. doi: 10.1002/ece3.3568.
Souza G, Costa L, Guignard MS, Van-Lume B, Pellicer J, Gagnon E, Leitch IJ, Lewis GP. 2019. Do tropical plants have smaller genomes? Correlation between genome size and climatic variables in the *Caesalpinia* Group (Caesalpinioideae, Leguminosae). *Perspectives in Plant Ecology, Evolution and Systematics* **38:** 13–23. doi: 10.1016/j.ppees.2019.03.002.
Sparrow AH, Underbrink AG, Sparrow RC. 1967. Chromosomes and cellular radiosensitivity. I. The relationship of D_0 to chromosome volume and complexity in seventy-nine different organisms. *Radiation Research* **32**: 915–945. doi: 10.2307/3572296.
Stebbins GL. 1940. The significance of polyploidy in plant evolution. *American Naturalist* **74:** 54–66. doi: 10.1086/280872.
Stebbins GL. 1966. Chromosomal variation and evolution: Polyploidy and chromosome size and number shed light on evolutionary processes in higher plants. *Science* **152:** 1463–1469. doi: 10.1126/science.152.3728.1463.
Symonds MRE, Blomberg SP. 2014. A primer on phylogenetic generalised least squares. In: LZ Garamszegi, ed. *Modern phylogenetic comparative methods and their application in evolutionary biology*. Heidelberg, Germany: Springer, 105–130.
Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. 2016. Genetic drift, not life history or RNAi, determine long-term evolution of transposable elements. *Genome Biology and Evolution* **8:** 2964–2978. doi: 10.1093/gbe/evw208.
Šmarda P, Knápek O, Březinová A, Horová L, Grulich V, Danihelka J, Veselý P, Šmerda J, Rotreklová O, Bureš P. 2019. Genome sizes and genomic guanine+cytosine (GC) contents of the Czech vascular flora with new estimates for 1700 species. *Preslia* **91**: 117–142. doi: 10.23855/preslia.2019.117.
Šmarda P, Hejcman M, Březinová A, Horová L, Steigerová H, Zedek F, Bureš P, Hejcmanová P, Schellberg J. 2013. Effect of phosphorus availability on the selection of species with different ploidy levels and genome sizes in a long-term grassland fertilization experiment. *New Phytologist* **200:** 911-921. doi: 10.1111/nph.12399.
Tenaillon MI, Hollister JD, Gaut BS. 2010. A triptych of the evolution of plant transposable elements. *Trends in Plant Science* **15**: 471–478. doi: 10.1016/j.tplants.2010.05.003.
Trabucco A, Zomer RJ. 2018. Global aridity index and potential evapo-transpiration (ET0) Climate Database v2. CGIAR Consortium for Spatial Information (CGIAR-CSI). Published online, available from the CGIAR-CSI GeoPortal at https://cgiarcsi.community
Trávníček P, Čertner M, Ponert J, Chumová Z, Jersáková J, Suda J. 2019. Diversity in genome size and GC content shows adaptive potential in orchids and is closely linked

to partial endoreplication, plant life-history traits and climatic conditions. *New Phytologist* **224:** 1642–1656. doi: 10.1111/nph.15996.
Vasconcelos T. 2022. Discovering the rules of plant biogeography using a trait-based approach, *preprint*, doi: 10.32942/osf.io/azytc.
Veleba A, Zedek F, Horová L, Veselý P, Srba M, Šmarda P, Bureš P. Is the evolution of carnivory connected with genome size reduction? *American Journal of Botany* **107**: 1253–1259. doi: 10.1002/ajb2.1526.
Veselý P, Bureš P, Šmarda P, Pavlíček T. 2012. Genome size and DNA base composition of geophytes: The mirror of phenology and ecology? *Annals of Botany* **109**: 65–75. doi: 10.1093/aob/mcr267.
Veselý P, Bureš P, Šmarda P. 2013. Nutrient reserves may allow for genome size increase: Evidence from comparison of geophytes and their sister non-geophytic relatives. *Annals of Botany* **112:** 1193–1200. doi: 10.1093/aob/mct185.
Veselý P, Šmarda P, Bureš P, Stirton C, Muasya AM, Mucina L, Horová L, Veselá K, Šilerová A, Šmerda J, Knápek O. 2020. Environmental pressures on stomatal size may drive plant genome size evolution: Evidence from a natural experiment with Cape geophytes. *Annals of Botany* **126**: 323–330. doi: 10.1093/aob/mcaa095.
Veter NM, DeSantis LR, Yann LT, Donohue SL, Haupt RJ, Corapi SE, Fathel SL, Gootee EK, Loffredo LF, Romer JL, Velkovsky SM. 2013. Is Rapoport's rule a recent phenomenon? A deep time perspective on potential causal mechanisms. *Biology Letters* **9**: 20130398. doi: 10.1098/rsbl.2013.0398
Vidal-Russell R, Tadey M, Urfusová R, Urfus T, Souto CP. 2022. Evolutionary importance of the relationship between cytogeography and climate: New insights on creosote bushes from North and South America. *Plant Diversity* **44:** 492–498. doi: 10.1016/j.pld.2021.11.006.
Vinogradov AE. 2003. Selfish DNA is maladaptive: evidence from the plant Red list. *Trends in Genetics* **19**: 609–614. doi: 10.1016/j.tig.2003.09.010.
Vitousek PM, Border S, Houlton BZ, Chadwick OA. 2010. Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen–phosphorus interactions. *Ecological Applications* **20**: 5–15. doi: 10.1890/08-0127.1.
Wang J, Santiago E, Caballero A. 2016. Prediction and estimation of effective population size. *Heredity* (Edinb) **117**: 193–206. doi: 10.1038/hdy.2016.43.
WCSP. 2017. WCSP: World Checklist of Selected Plant Families: Royal Botanic Gardens, Kew. URL: http://apps.kew.org/wcsp/home.do, Retrieved 06 February 2017
Wendel JF, Jackson SA, Meyers BC, Wing RA. 2016. Evolution of plant genome architecture. *Genome Biology* **17:** 37. doi: 10.1186/s13059-016-0908-1.
Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* **42:** 225–249. doi: 10.1007/978-94-011-4221-2_12.
Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *American Journal of Botany* **102:** 1753–1756. doi: 10.3732/ajb.1500320.
Whitney KD, Baack EJ, Hamrick JL, Godt MJ, Barringer BC, Bennett MD, Eckert CG, Goodwillie C, Kalisz S, Leitch IJ, Ross-Ibarra J. 2010. A role for nonadaptive processes in plant genome size evolution? *Evolution* **64:** 2097–2109. doi: 10.1111/j.1558-5646.2010.00967.x.
Whitney KD, Garland T Jr. 2010. Did genetic drift drive increases in genome complexity? *PLoS Genetics* **6**: e1001080. doi: 10.1371/journal.pgen.1001080.
Yu J, Li D, Lou Y, Guo S. 2018. Nuclear DNA content variation of herbaceous angiosperm species on 10 global latitudinal transects. *Journal of the Torrey Botanical Society* **145:** 340–352. doi: 10.3159/TORREY-D-16-00062.1.

Zachariassen KE, Kristiansen E. 2000. Ice nucleation and antinucleation in nature (a review). *Cryobiology* **41:** 2710–3279. doi: 10.1006/cryo.2000.2289.
Zedek F, Bureš P. 2019. Pest arthropods with holocentric chromosomes are more resistant to sterilizing ionizing radiation. *Radiation Research* **191**: 255–261. doi: 10.1667/RR15208.1.
Zedek F, Plačková K, Veselý P, Šmerda J, Šmarda P, Horová L, Bureš P. 2020. Endopolyploidy is a common response to UV-B stress in natural plant populations, but its magnitude may be affected by chromosome type. *Annals of Botany* **126**: 883–889. doi: 10.1093/aob/mcaa109.
Zedek F, Šmerda J, Veselý P, Horová L, Kocmanová J, Bureš P. 2021. Elevation-dependent endopolyploid response suggests that plants with holocentric chromosomes are less stressed by UV-B. *Botanical Journal of the Linnean Society* **195**: 106–113. doi: 10.1093/botlinnean/boaa054.
Zedek F, Veselý P, Tichý L, Elliott TL, Garbolino E, de Ruffray P, Bureš P. 2022. Holocentric plants are more competitive under higher UV-B doses. *New Phytologist* **233**: 15–21. doi: 10.1111/nph.17750.
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. doi: 10.1111/evo.13226.
Zonneveld BJM. 2019. The DNA weights per nucleus (genome size) of more than 2350 species of the Flora of The Netherlands, of which 1370 are new to science, including the pattern of their DNA peaks. *Forum Geobotanicum* **8:** 24–78. doi: 10.3264/FG.2019.1022.

Figure captions

Fig. 1 Expected associations between genome size and (a) range size and (b, c) latitude based on three hypotheses outlined in the Introduction. The question mark in (c) indicates uncertainty about the potential shape of the curve. Given this uncertainty, we present a curve that could possibly result from the effects of temperature.

Fig. 2 Associations between genome and range size per species (a,b,c) and averaged per TDWG Level-3 regions (d). The association of the raw data between genome and range size is shown in (a), whereas both variables are log-transformed in the other three plots (b,c,d). The slope estimates from the quantile regression, including 95% confidence intervals (dark grey), are indicated in (c). The solid red lines in (b) and (d) indicate the fit of the ordinary least squares (OLS) regressions, while the solid red line in (c) indicates the slope value from the OLS analysis. Dashed red lines in (b, c, d) represent 95% confidence intervals. The size of points in (d) indicates the weights in the regression analysis (see Methods for details).

Fig. 3 The global distribution of mean genome size (a) and polyploid proportion (b) in flowering plants. Mean genome size (2C) and the proportion of polyploids were calculated per TDWG Level-3 region. The two plots on the left side show (a) the distribution of genome size and (b) the proportion of polyploids across latitude. Dark red indicates TDWG regions with the highest temperatures in the coldest quarter (BIO11 from CHELSA), and dark blue shows regions with the lowest temperatures. The solid line in the plot indicates the mean from the regression fit. Dashed lines indicate 95% confidence intervals. The size of points in the plots indicates the weights used in the regression analysis. The weight was calculated as "$\log_{10}$(number of species in a TDWG)" divided by the ratio of the number of species for which we have genome size data (or the proportion of polyploids) : number of all species in the TDWG The maps on the right side show the distribution of (a) mean genome size and (b) polyploid proportion, with dark red and light yellow TDWG regions indicating areas with relatively high and low values for each variable, respectively.

Fig. 4 The association of genome size (2C; Gbp) and latitude across four growth forms groupings. Plot (a) is based on species genome sizes (grey circles), whereas the latter four plots (b,c,d,e) represent the mean genome size calculated per TDWG for a given growth form. All results are based on polynomial regressions of the 3rd order, where solid lines represent the model estimates. The dashed lines in (a) show the 95% confidence intervals.

Fig. 5 Major factors (to the left of arrows) suggested affecting physiological, anatomical, and molecular response (to the right of arrows), resulting in expansions/contractions of the genome and thus ultimately forming the global latitudinal trend in genome size (S-shaped curve). Relatively high genome sizes in the temperate regions could be the result of relaxed selective pressure, whereas various drivers might be constraining or pushing this trait in tropical and arctic regions. The proportion of polyploid species (low in the tropics and increasing toward the poles), which is not included in the figure, could also weakly (Table 2) contribute to the observed latitudinal trend in genome size.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Pearson's correlation coefficients (*r*) among 29 climatic variables assessed to be included in the multiple linear regression model explaining genome size variation along the global latitudinal gradient. Dark red and dark blue circles indicate high and low *r* values, respectively. Larger circles in the upper triangle represent stronger correlations between variables (both negative and positive), whereas the numbers in the lower triangle indicate the *r* values.

Fig. S2 Global distribution of mean geographic range sizes for those species included in the genome size dataset (a) and for all species in the WCVP dataset (b) mapped per TDWG Level-3 region. The two plots on the left-hand side of the figure show the distribution of mean geographic range sizes across the global latitudinal gradient. Dark red shading in the maps on the right-hand side of the figure indicates relatively high mean range sizes of species included in each TDWG unit, whereas light yellows indicate TDWGs with species with relatively small range sizes.

Fig. S3 Associations between genome and range size per species considering phylogenetic relationships. The solid red line in (a) indicates the fit of the phylogenetic generalized least squares regression (PGLS), while the red line in (b) indicates the slope value from the phylogenetic quantile regression analysis. The slope estimates from the phylogenetic quantile regression, including 95% confidence intervals (error bars), are indicated in (b). Dashed red lines in both plots represent 95% confidence intervals. Both genome and range size are transformed by $\log_{10}$ in (a) and (b).

Fig. S4 Associations between mean chromosome size and range size per species. The solid red lines in (a) and (c) indicate the fit of the ordinary least squares (OLS) and phylogenetic generalized least squares (PGLS) regressions, respectively. The solid black circles and the gray shading in (b) represent the slope estimates and the 95% confidence intervals across 19 quantiles, whereas the hollow circles and the error bars in (d) indicate slope estimates and the 95% confidence intervals of the phylogenetic quantile regression. The horizontal red line in (b) represents the slope estimate of the OLS regression, while the horizontal red line in (d) shows the slope estimate of the PGLS regression. Dotted red lines in all four plots indicate the 95% confidence intervals of the slope estimates.

Fig. S5 Mean genome sizes (2C; Gbp) averaged per TDWG Level-3 region for the two most species-rich monocot (a,b) and dicot (c,d) families. Dark red colors indicate relatively large mean genome sizes, whereas light yellows indicate TDWGs with relatively small mean genome sizes.

Fig. S6 The global distribution of mean chromosome size in flowering plants calculated per TDWG Level-3 region. The plot on the left side shows the distribution of mean chromosome sizes across latitudes, with dark reds indicating TDWG regions with high temperatures in the coldest quarter (BIO11 from Bioclim) and dark blues showing regions with low temperatures. The size of points in the plots indicates the weights used in the regression analysis (see Methods for details). The map on the right side shows the distribution of mean chromosome

sizes mapped according to each TDWG region, where dark reds indicate relatively high values.

Fig. S7 Mean genome sizes (2C; Gbp) across the global latitudinal gradient for the Old World (a) and New World (b). Circles in both plots represent the genome size averaged per TDWG Level-3 region.

Fig. S8 Mean genome sizes (2C; Gbp) across the global latitudinal gradient illustrating TDWG Level-3 regions that were glaciated (blue) and non-glaciated (red) during the last glacial maximum (LGM) approximately 18,000 years before the present. We assessed the glaciation status of each TDWG region at the Last Glacial Maximum (LGM; ~18,000 years BP) using past climatic reconstructions from Ehlers (2015). We considered TDWG regions to be 'Glaciated' if their centroids were covered by the ice sheets during the LGM (Dataset S2).

Dataset S1 Dataset containing 16,017 angiosperm taxa, their genome sizes, chromosome numbers, chromosome sizes, geographic ranges, latitudinal centroids, and growth forms.

Dataset S2 Dataset containing 369 TDWGs (Botanical countries), their geographic centroids, counts of all angiosperm taxa and counts of angiosperm taxa with genomic traits; mean values for genome size, chromosome size, range size; mean values for genome size in growth forms; proportion of polyploid taxa; glaciation status

Table S1 Bioclim variables as they explain the variance in 2C genome size across TDWGs in the polynomial regression of a given order

Table S2 Results of quantile regression of genome size on range size

Table S3 Results of phylogenetic quantile regression of genome size on range size

Table S4 Results of OLS and PGLS regressions of mean chromosome size on range size

Table S5 Results of quantile regression of mean chromosome size on range size

Table S6 Results of phylogenetic quantile regression of mean chromosome size on range size

Table S7 Results of regressions of 2C genome size on different variables discussed in the Discussion

Table 1: Results of OLS and PGLS regressions of 2C genome size on range size

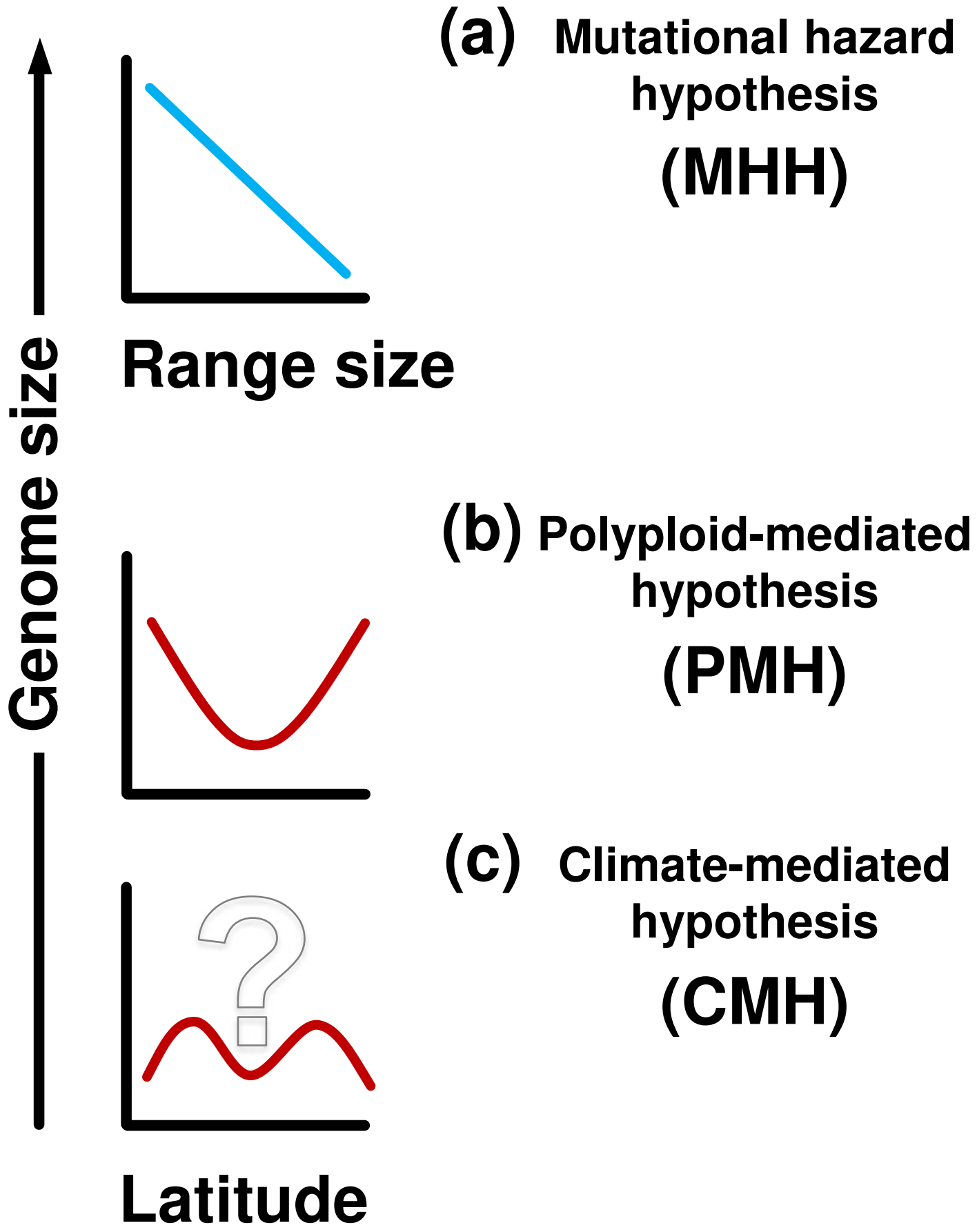
OLS model: $\log_{10}$(2C genome size) ~ $\log_{10}$(Range size)					
Model term	b_i	95%CI	t	P	R^2_{adj}
Intercept	3.604	<3.592, 3.617>	559.54	<2E-16	0.014
$\log_{10}$(Range size)	-0.104	<-0.118, -0.091>	-15.04	<2E-16	
PGLS model: $\log_{10}$(2C genome size) ~ $\log_{10}$(Range size)					
Model term	b_i	95%CI	P	lambda	R^2_{adj}
Intercept	3.560	<3.5594, 3.5604>	<2E-16	0.924	0.002
$\log_{10}$(Range size)	-0.021	<-0.0213, -0.0208>	<4E-7		

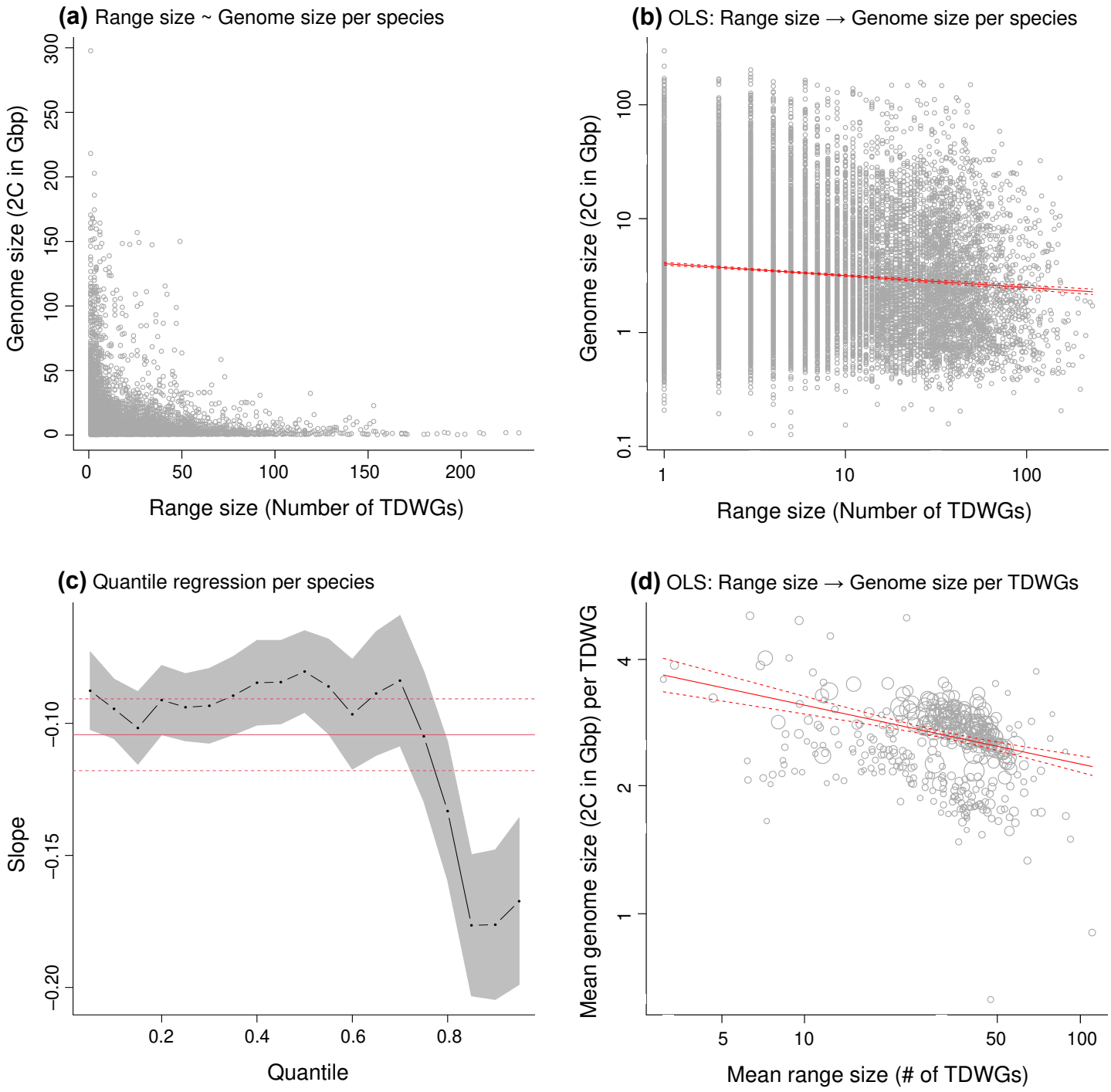
Table 1: Results of ordinary least squares (OLS) and phylogenetic generalized least squares (PGLS) regressions of 2C genome size on range size. b_i - regression estimates of model terms; 95%CI - lower and upper 95% confidence intervals of the regression estimates; R^2_{adj} - R squared adjusted indicating explained variance. The OLS analysis was performed with 15,984 species. The PGLS analysis was performed with 12,134 species. The PGLS was performed repeatedly with one hundred different trees (see Methods). Therefore, the values for PGLS are averages across these one hundred regressions.

Table 2: Results of linear and polynomial regressions of 2C genome size and polyploid proportion on various predictors.

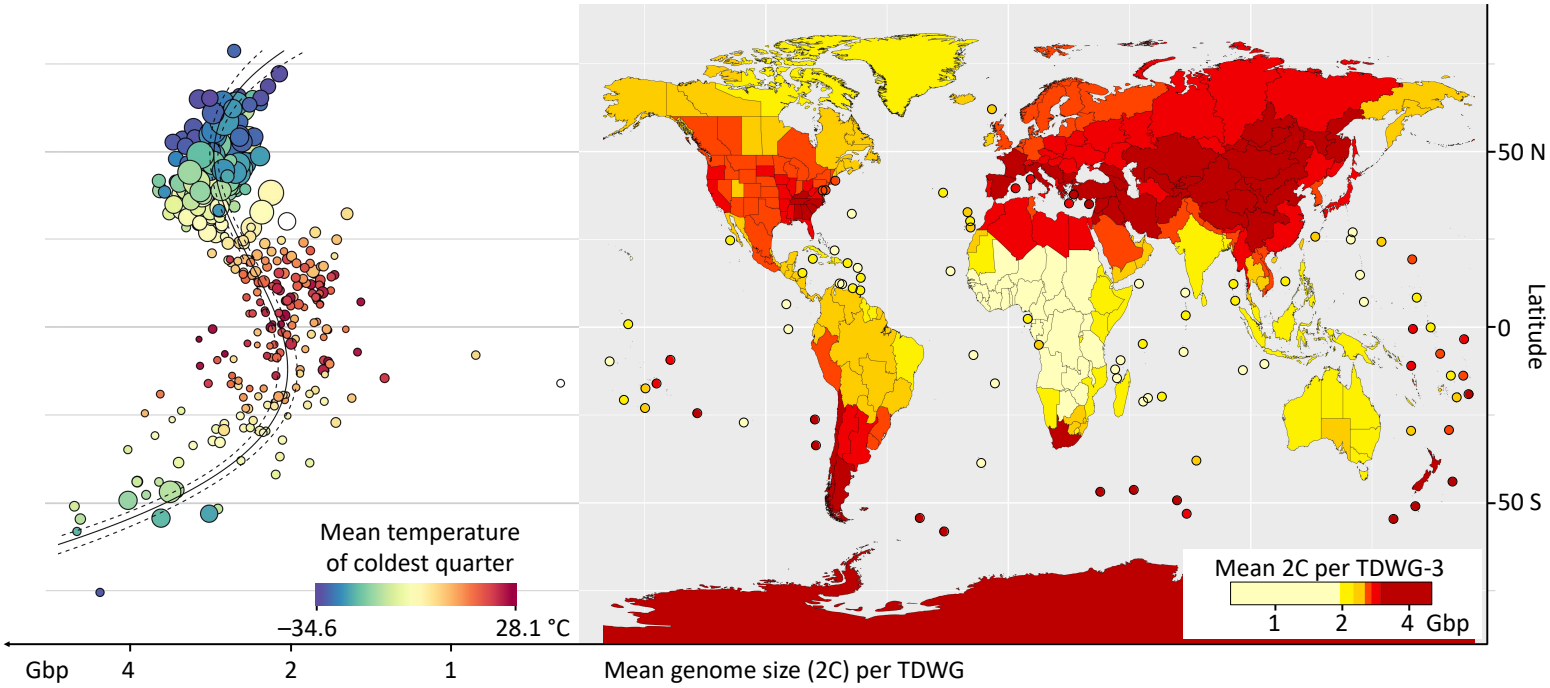
Polynomial regression (N=367): $\log_{10}$(2C genome size) ~ latitude + latitude^2 + latitude^3					
Model term	b_i	95%CI	t	P	R^2_{adj}
Intercept	3.320	<3.305, 3.335>	434.02	<2E-16	0.4012
Latitude	2.136E-03	<1.67E-03, 2.60E-03>	9.07	<2E-16	
Latitude^2	6.734E-05	<5.85E-05, 7.62E-05>	14.99	<2E-16	
Latitude^3	-1.215E-06	<-1.40E60, -1.03E-06>	-13.09	<2E-16	
Polynomial regression (N=368): polyploid proportion ~ latitude + latitude^2					
Model term	b_i	95%CI	t	P	R^2_{adj}
Intercept	0.344	<0.330, 0.357>	49.06	<2E-16	0.2947
Latitude	-1.11E-03	<-1.48E-03, -7.27E-04>	-5.75	1.95E-08	
Latitude^2	5.47E-05	<4.59E-05, 6.35E-05>	12.19	<2E-16	
Linear regression (N=367): $\log_{10}$(2C genome size) ~ polyploid proportion					
Model term	b_i	95%CI	t	P	R^2_{adj}
Intercept	3.372	<3.338, 3.407>	191.10	<2E-16	0.0177
Polyploid proportion	0.116	<0.033, 0.199>	2.75	0.006	
Polynomial regression (N=365): $\log_{10}$(2C genome size) ~ GST + GST^2					
Model term	b_i	95%CI	t	P	R^2_{adj}
Intercept	3.444	<3.405, 3.483>	172.53	<2E-16	0.4075
GST	0.005	<-4.33E-05, -0.0099>	1.95	0.052	
GST^2	0.000	<-0.0005, -0.0003>	-5.34	<2E-07	

Table 2: N - number of TDWGs included in the analysis; b_i - regression estimates of model terms; 95%CI - lower and upper limits of 95% confidence intervals of the regression estimates; R^2_{adj} - R squared adjusted indicating explained variance. GST is the mean temperature of the growing season. In the case of polynomial regressions, we fitted orthogonal polynomials using the "poly" function in base R, but the parameter "raw" was set to "TRUE" to obtain parameter estimates corresponding to response variables.

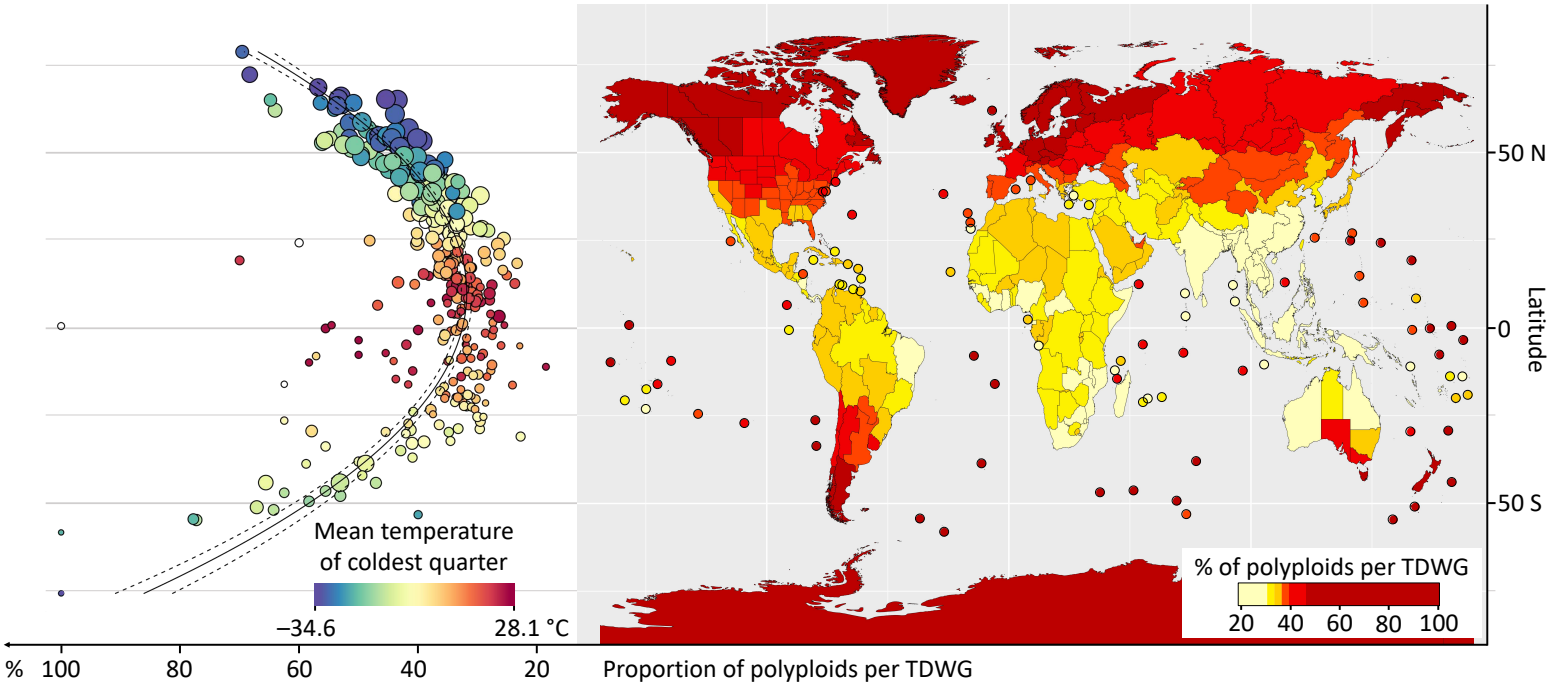

(a) Mutational hazard hypothesis
(MHH)
Range size
Genome size
(b) Polyploid-mediated hypothesis
(PMH)
(c) Climate-mediated hypothesis
(CMH)
Latitude

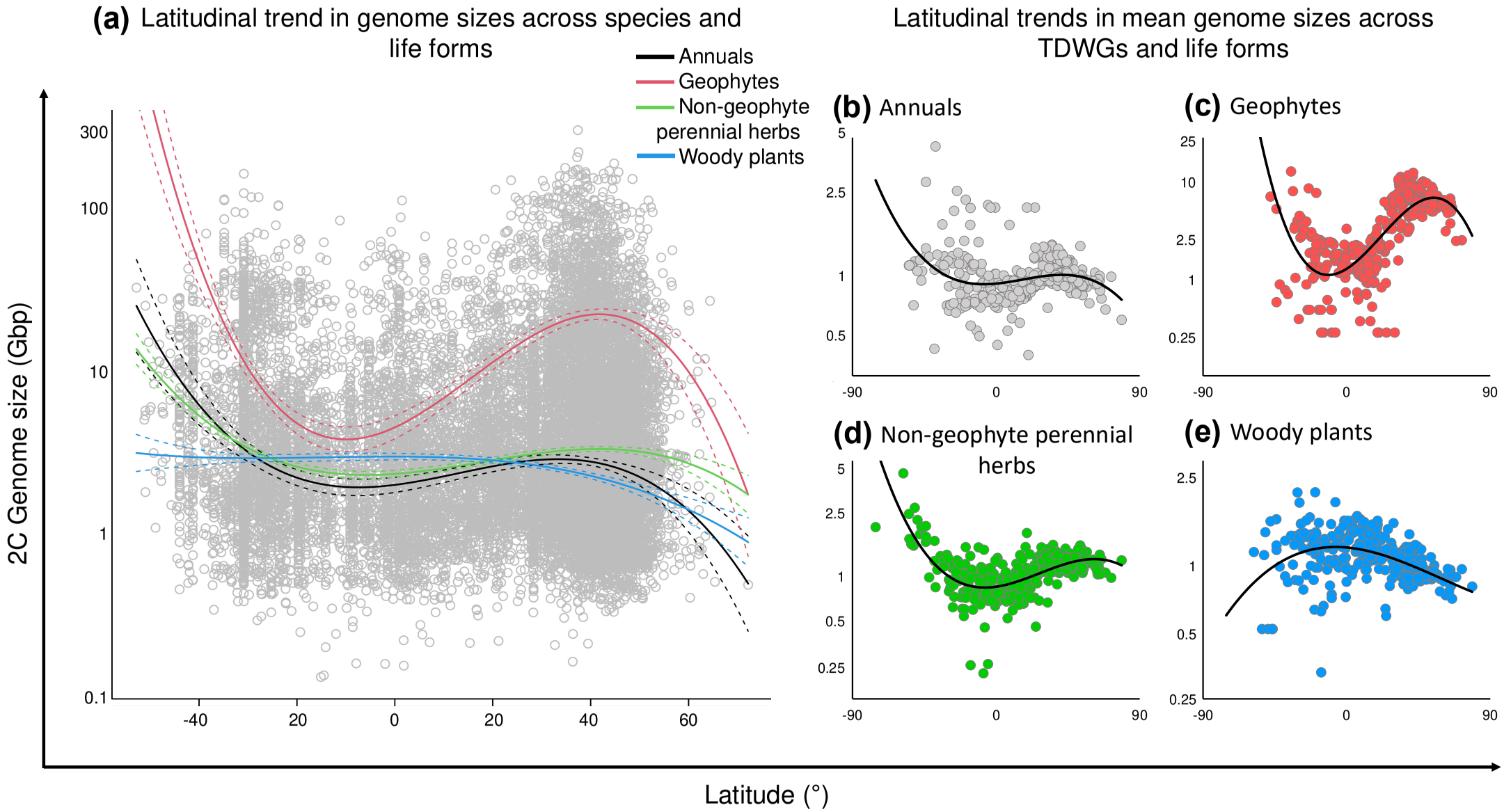

(a) Range size ~ Genome size per species
Genome size (2C in Gbp)
Range size (Number of TDWGs)
(b) OLS: Range size → Genome size per species
Genome size (2C in Gbp)
Range size (Number of TDWGs)
(c) Quantile regression per species
Slope
Quantile
(d) OLS: Range size → Genome size per TDWGs
Mean genome size (2C in Gbp) per TDWG
Mean range size (# of TDWGs)

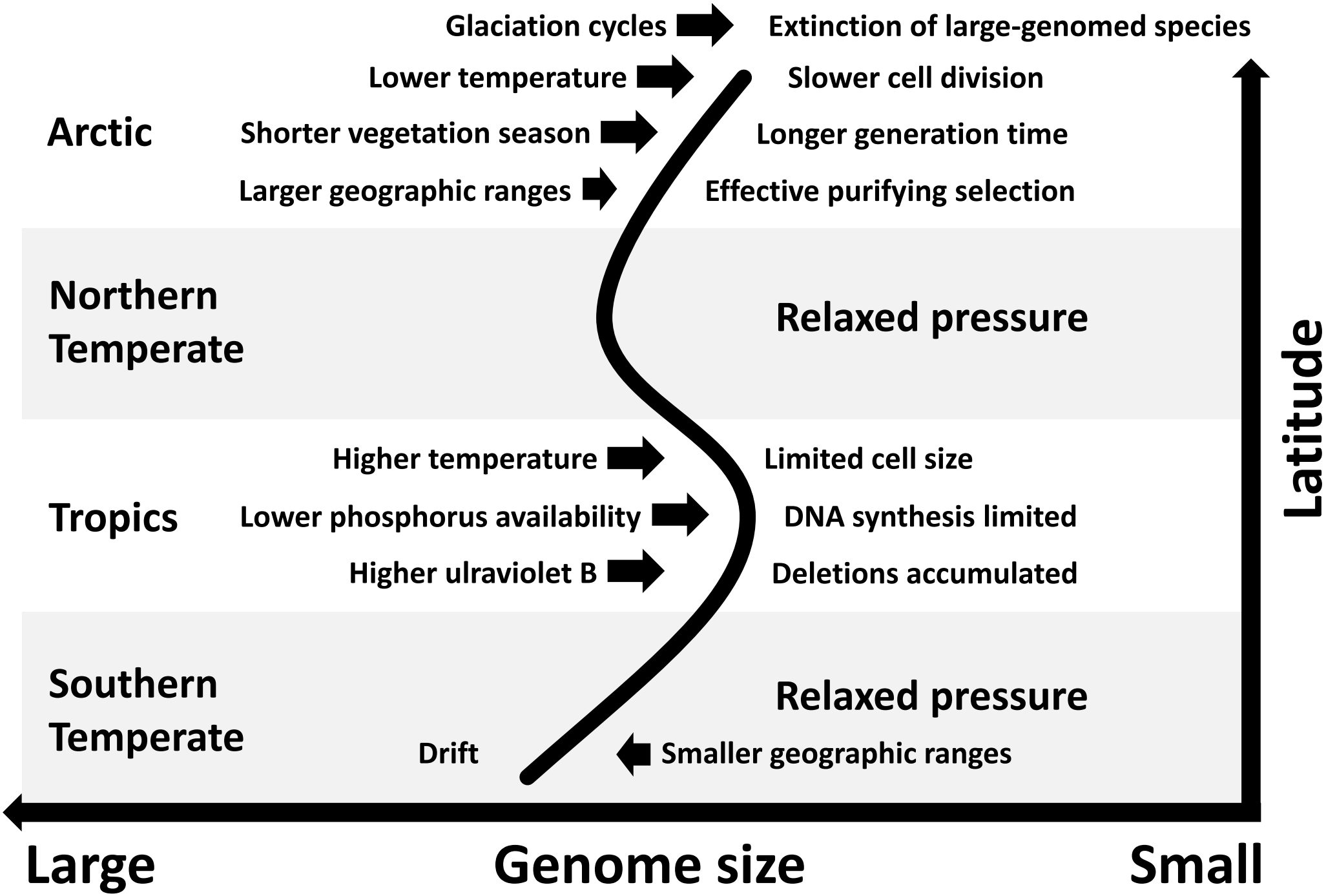
(a) Latitudinal gradient of mean genome size (2C) in flowering plants



(b) Latitudinal gradient of polyploid proportion (%) in flowering plants




(a) Latitudinal trend in genome sizes across species and life forms
Annuals
Geophytes
Non-geophyte perennial herbs
Woody plants
2C Genome size (Gbp)
300
100
10
1
0.1
-40
20
0
20
40
60
Latitude (°)
Latitudinal trends in mean genome sizes across TDWGs and life forms
(b) Annuals
(c) Geophytes
(d) Non-geophyte perennial herbs
(e) Woody plants
-90
0
90


Glaciation cycles
Extinction of large-genomed species
Lower temperature
Slower cell division
Arctic
Shorter vegetation season
Longer generation time
Larger geographic ranges
Effective purifying selection
Northern Temperate
Relaxed pressure
Higher temperature
Limited cell size
Tropics
Lower phosphorus availability
DNA synthesis limited
Higher ulraviolet B
Deletions accumulated
Southern Temperate
Relaxed pressure
Drift
Smaller geographic ranges
Latitude
Large
Genome size
Small

New Phytologist Supporting Information

Article Title: The global biogeography of genome size in angiosperms is shaped by climate and range size

Authors: Petr Bureš[1], Tammy L. Elliott[1], Pavel Veselý[1], Petr Šmarda[1,], Félix Forest[2], Ilia J. Leitch[2], Eimear Nic Lughadha[2], Marybel Soto Gomez[2], Samuel Pironon[2,3], Matilda Brown[2], František Zedek[1]

[1]Masaryk University, Faculty of Science, Department of Botany and Zoology, Kotlarska 2, Brno, Czech Republic;
[2]Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, United Kingdom; and
[3]UN Environment Programme World Conservation Monitoring Centre (UNEP-WCMC), Cambridge, Cambridgeshire, CB3 0DL, United Kingdom.

Article acceptance date: XX XXX XXXX

The Following Supporting Information is available for this article

Fig. S1 Pearson's correlation coefficients (*r*) among 29 climatic variables assessed to be included in the multiple linear regression model explaining genome size variation along the global latitudinal gradient. Dark red and dark blue circles indicate high and low *r* values, respectively. Larger circles in the upper triangle represent stronger correlations between variables (both negative and positive), whereas the numbers in the lower triangle indicate the *r* values.

Fig. S2 Associations between mean chromosome size and range size per species. The solid red lines in (a) and (c) indicate the fit of the ordinary least squares (OLS) and phylogenetic generalized least squares (PGLS) regressions, respectively. The solid black circles and the gray shading in (b) represent the slope estimates and the 95% confidence intervals across 19 quantiles, whereas the hollow circles and the error bars in (d) indicate slope estimates and the 95% confidence intervals of the phylogenetic quantile regression. The horizontal red line in (b) represents the slope estimate of the OLS regression, while the horizontal red line in (d) shows the slope estimate of the PGLS regression. Dotted red lines in all four plots indicate the 95% confidence intervals of the slope estimates.

Fig. S3 Mean genome sizes (2C; Gbp) averaged per TDWG Level-3 region for the two most species rich monocot (a,b) and dicot (c,d) families. Dark red colours indicate relatively large mean genome sizes, whereas light yellows indicate TDWGs with relatively small mean genome sizes.

Fig. S4 The global distribution of mean chromosome size in flowering plants calculated per TDWG Level-3 region. The plot on the left side shows the distribution of mean chromosome sizes across latitudes, with dark reds indicating TDWG regions with high temperatures in the coldest quarter (BIO11 from Bioclim) and dark blues showing regions with low temperatures. The size of points in the plots indicates the weights used in the regression analysis (see Methods for details). The map on the right side shows the distribution of mean chromosome sizes mapped according to each TDWG region, where dark reds indicate relatively high values.

Fig. S5 Mean genome sizes (2C; Gbp) across the global latitudinal gradient for the Old World (a) and New World (b). Circles in both plots represent the genome size averaged per TDWG Level-3 region.

Fig. S6 Global distribution of mean geographic range sizes for those species included in the genome size dataset (a) and for all species in the WCVP dataset (b) mapped per TDWG Level-3 region. The two plots on the left hand side of the figure show the distribution of mean geographic range sizes across the global latitudinal gradient. Dark red shading in the maps on the right hand side of the figure indicate relatively high mean range sizes of species included in each TDWG unit, whereas light yellows indicate TDWGs with species with relatively small range sizes.

Fig. S7 Associations between genome and range size per species considering phylogenetic relationships. The solid red line in (a) indicates the fit of the phylogenetic generalized least squares regression (PGLS), while the red line in (b) indicates the slope value from the phylogenetic quantile regression analysis. The slope estimates from the phylogenetic quantile regression, including 95% confidence intervals (error bars), are indicated in (b). Dashed red lines in both plots represent 95% confidence intervals. Both genome and range size are transformed by $\log_{10}$ in (a) and (b).

Fig. S8 Mean genome sizes (2C; Gbp) across the global latitudinal gradient illustrating TDWG Level-3 regions that were glaciated (blue) and non-glaciated (red) during the last glacial maximum (LGM) approximately 18,000 years before present.

Table S1 Bioclim variables as they explain the variance in 2C genome size across TDWGs in the polynomial regression of a given order.

Table S2 Results of quantile regression of genome size on range size.

Table S3 Results of phylogenetic quantile regression of genome size on range size.

Table S4 Results of OLS and PGLS regressions of mean chromosome size on range size.

Table S5 Results of quantile regression of mean chromosome size on range size.

Table S6 Results of phylogenetic quantile regression of mean chromosome size on range size.

Table S7 Results of regressions of 2C genome size on different variables discussed in the Discussion.

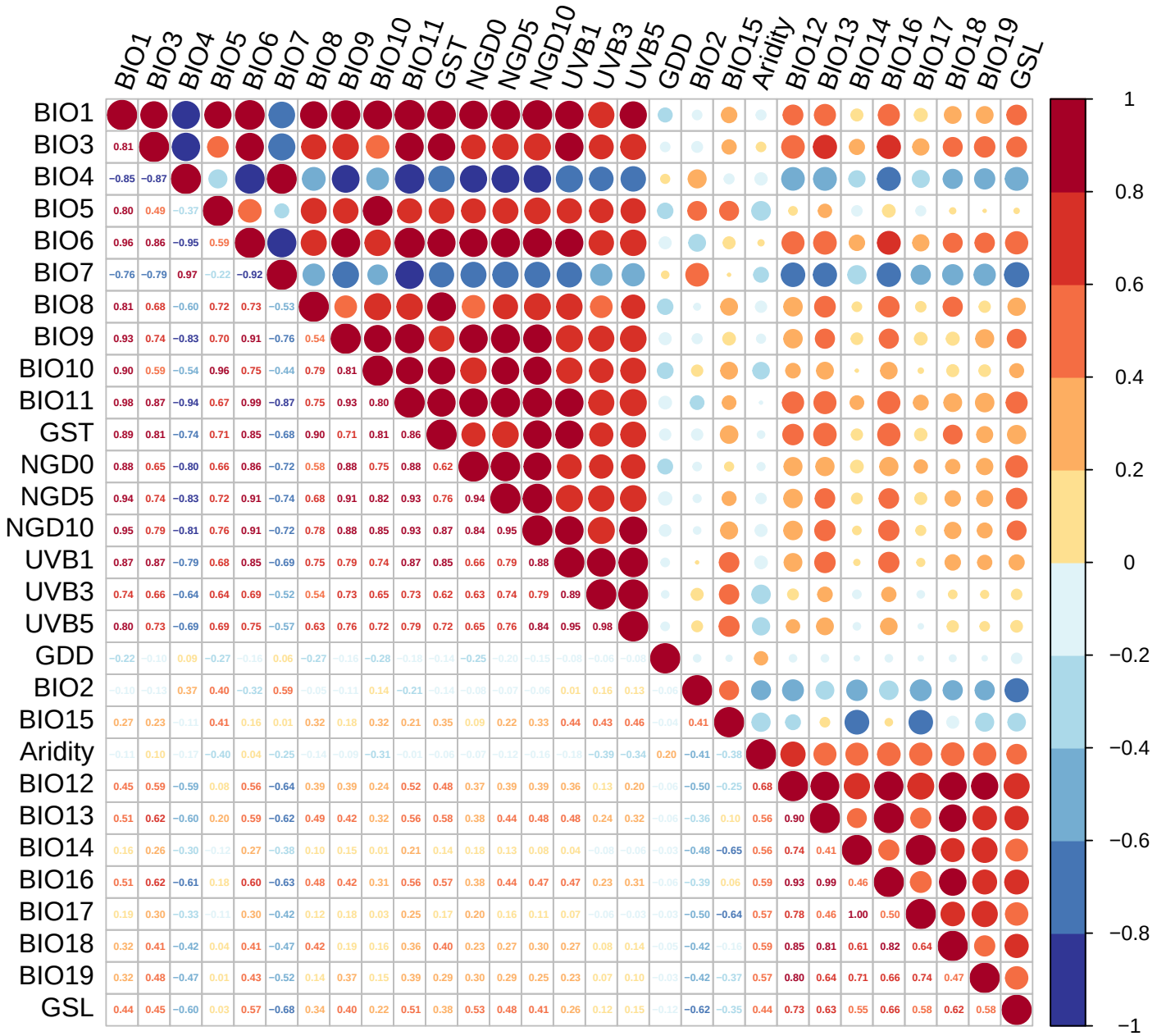


Fig. S1 Pearson's correlation coefficients (*r*) among 29 climatic variables assessed to be included in the multiple linear regression model explaining genome size variation along the global latitudinal gradient. Dark red and dark blue circles indicate high and low *r* values, respectively. Larger circles in the upper triangle represent stronger correlations between variables (both negative and positive), whereas the numbers in the lower triangle indicate the *r* values.

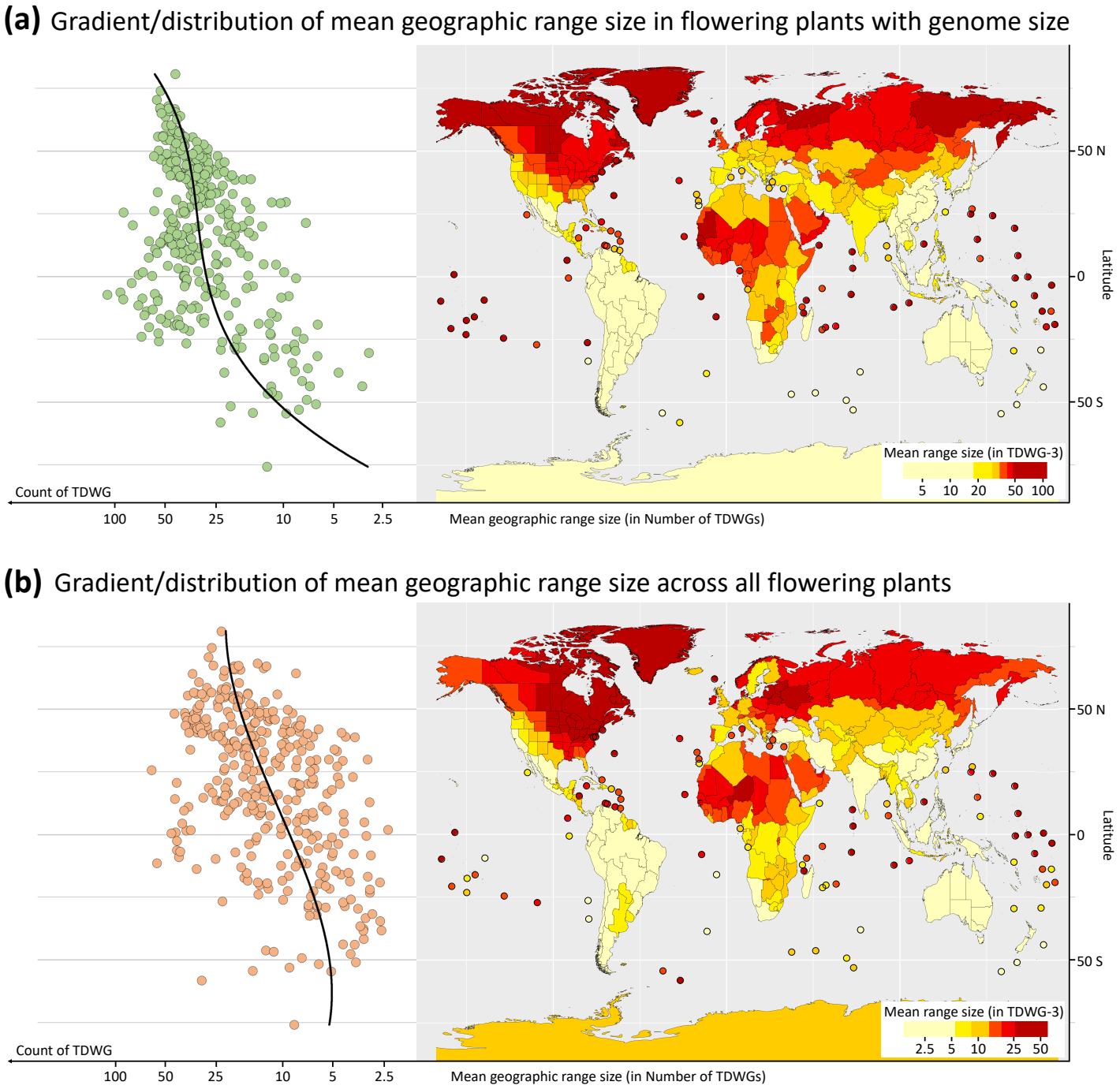


Fig. S2 Global distribution of mean geographic range sizes for those species included in the genome size dataset (a) and for all species in the WCVP dataset (b) mapped per TDWG Level-3 region. The two plots on the left hand side of the figure show the distribution of mean geographic range sizes across the global latitudinal gradient. Dark red shading in the maps on the right hand side of the figure indicate relatively high mean range sizes of species included in each TDWG unit, whereas light yellows indicate TDWGs with species with relatively small range sizes.

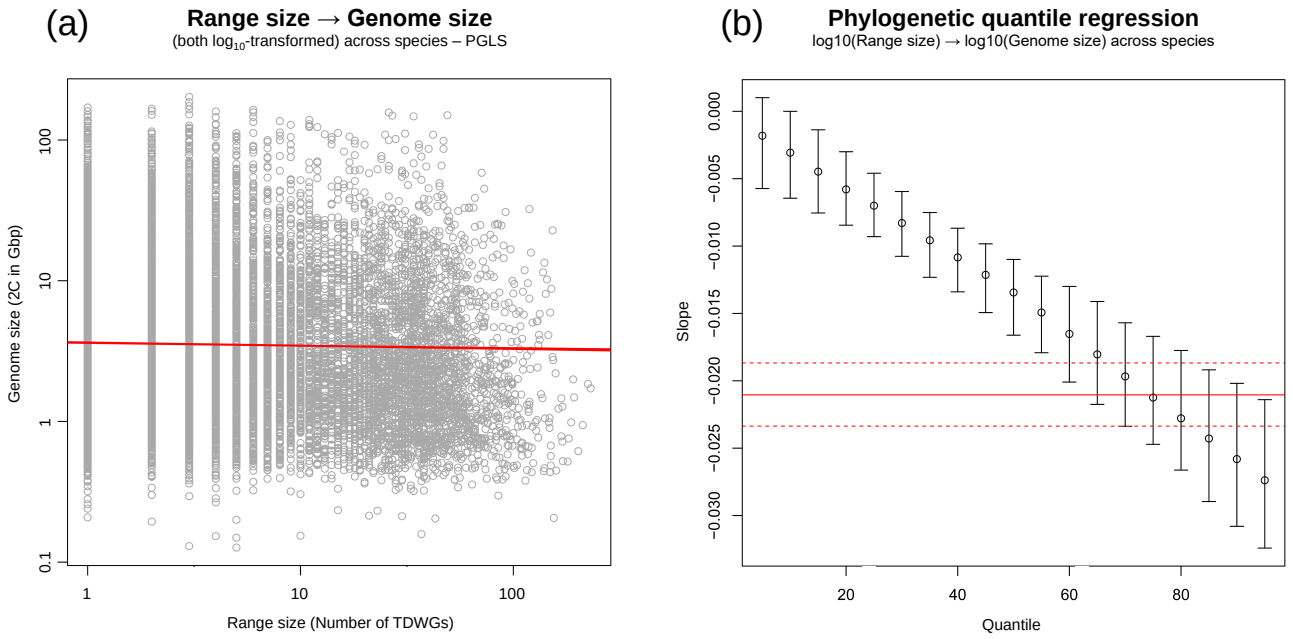


Fig. S3 Associations between genome and range size per species considering phylogenetic relationships. The solid red line in (a) indicates the fit of the phylogenetic generalized least squares regression (PGLS), while the red line in (b) indicates the slope value from the phylogenetic quantile regression analysis. The slope estimates from the phylogenetic quantile regression, including 95% confidence intervals (error bars), are indicated in (b). Dashed red lines in both plots represent 95% confidence intervals. Both genome and range size are transformed by $\log_{10}$ in (a) and (b).

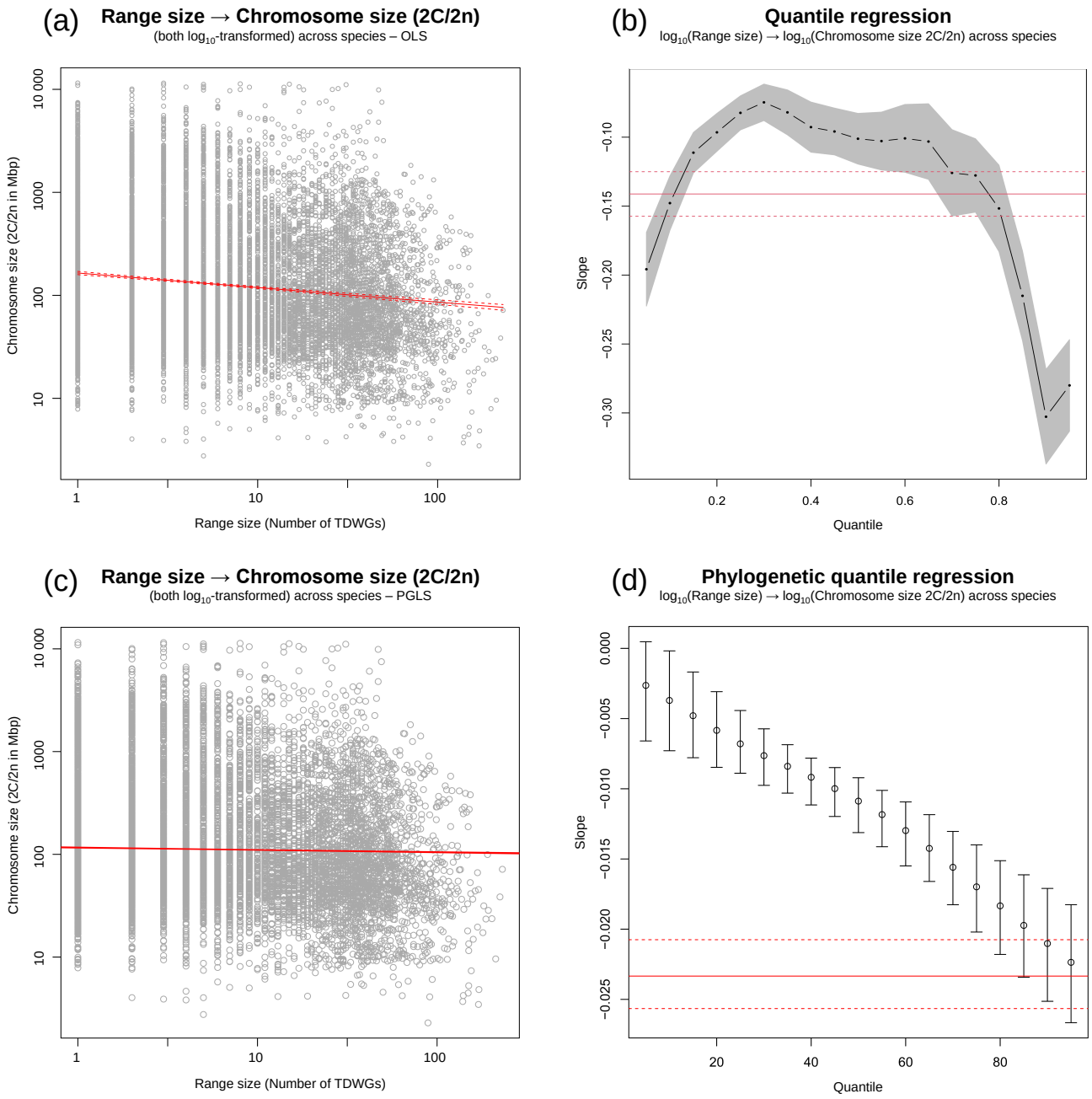


Fig. S4 Associations between mean chromosome size and range size per species. The solid red lines in (a) and (c) indicate the fit of the ordinary least squares (OLS) and phylogenetic generalized least squares (PGLS) regressions, respectively. The solid black circles and the gray shading in (b) represent the slope estimates and the 95% confidence intervals across 19 quantiles, whereas the hollow circles and the error bars in (d) indicate slope estimates and the 95% confidence intervals of the phylogenetic quantile regression. The horizontal red line in (b) represents the slope estimate of the OLS regression, while the horizontal red line in (d) shows the slope estimate of the PGLS regression. Dotted red lines in all four plots indicate the 95% confidence intervals of the slope estimates.

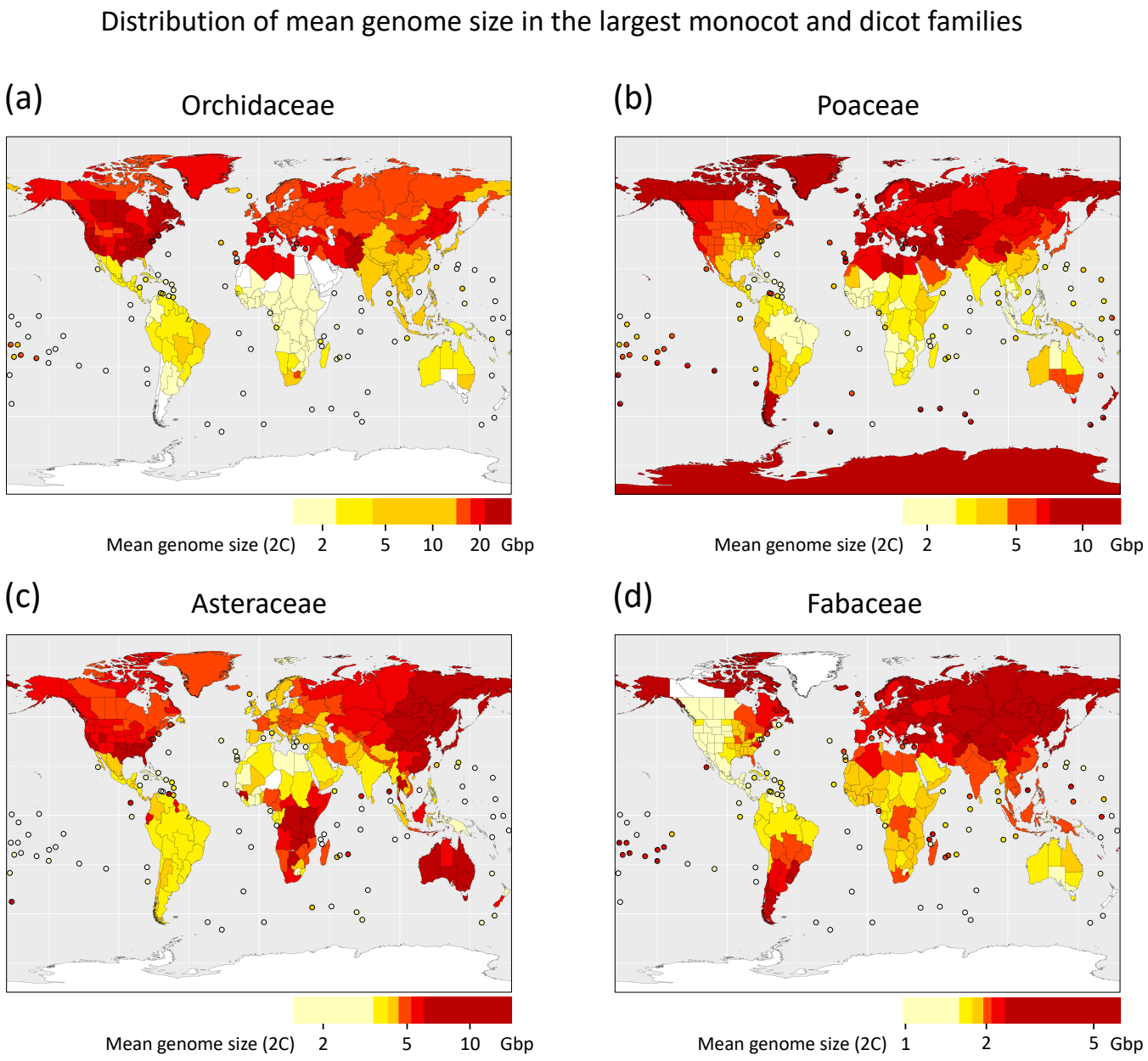


Fig. S5 Mean genome sizes (2C; Gbp) averaged per TDWG Level-3 region for the two most species rich monocot (a,b) and dicot (c,d) families. Dark red colours indicate relatively large mean genome sizes, whereas light yellows indicate TDWGs with relatively small mean genome sizes.

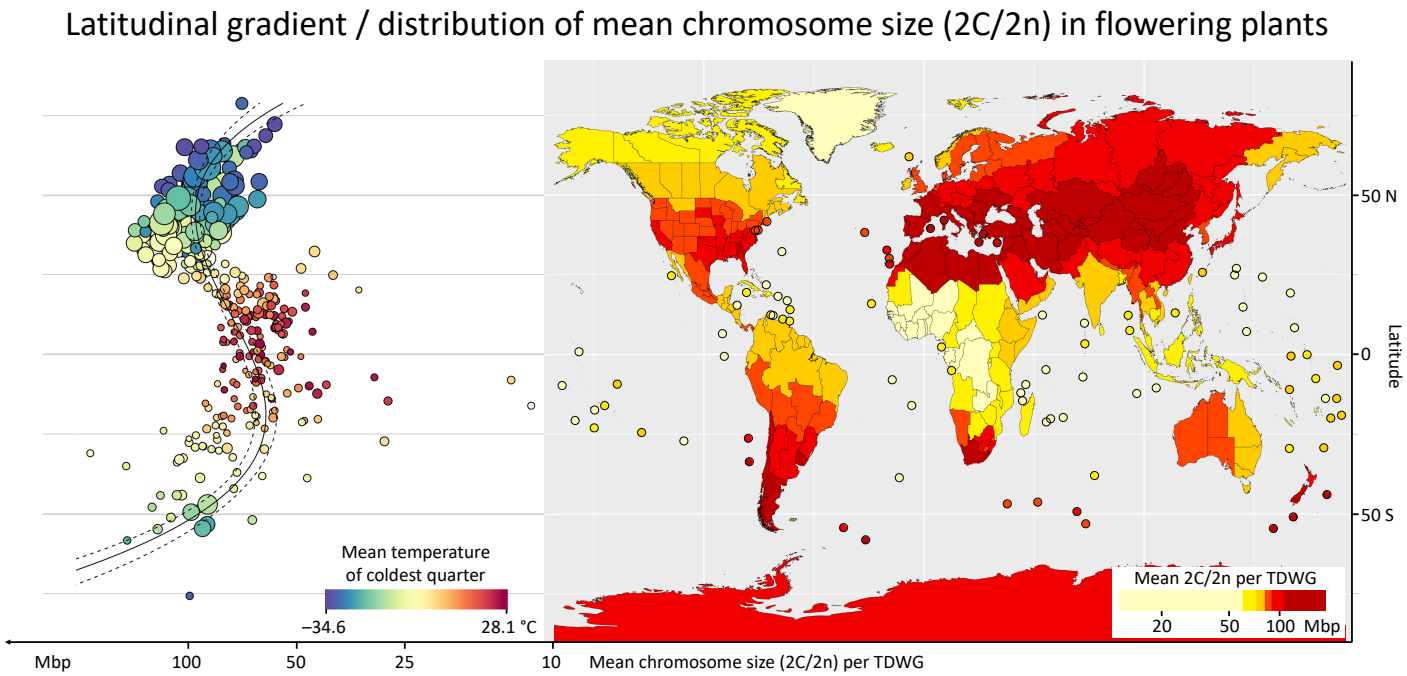


Fig. S6 The global distribution of mean chromosome size in flowering plants calculated per TDWG Level-3 region. The plot on the left side shows the distribution of mean chromosome sizes across latitudes, with dark reds indicating TDWG regions with high temperatures in the coldest quarter (BIO11 from Bioclim) and dark blues showing regions with low temperatures. The size of points in the plots indicates the weights used in the regression analysis (see Methods for details). The map on the right side shows the distribution of mean chromosome sizes mapped according to each TDWG region, where dark reds indicate relatively high values.

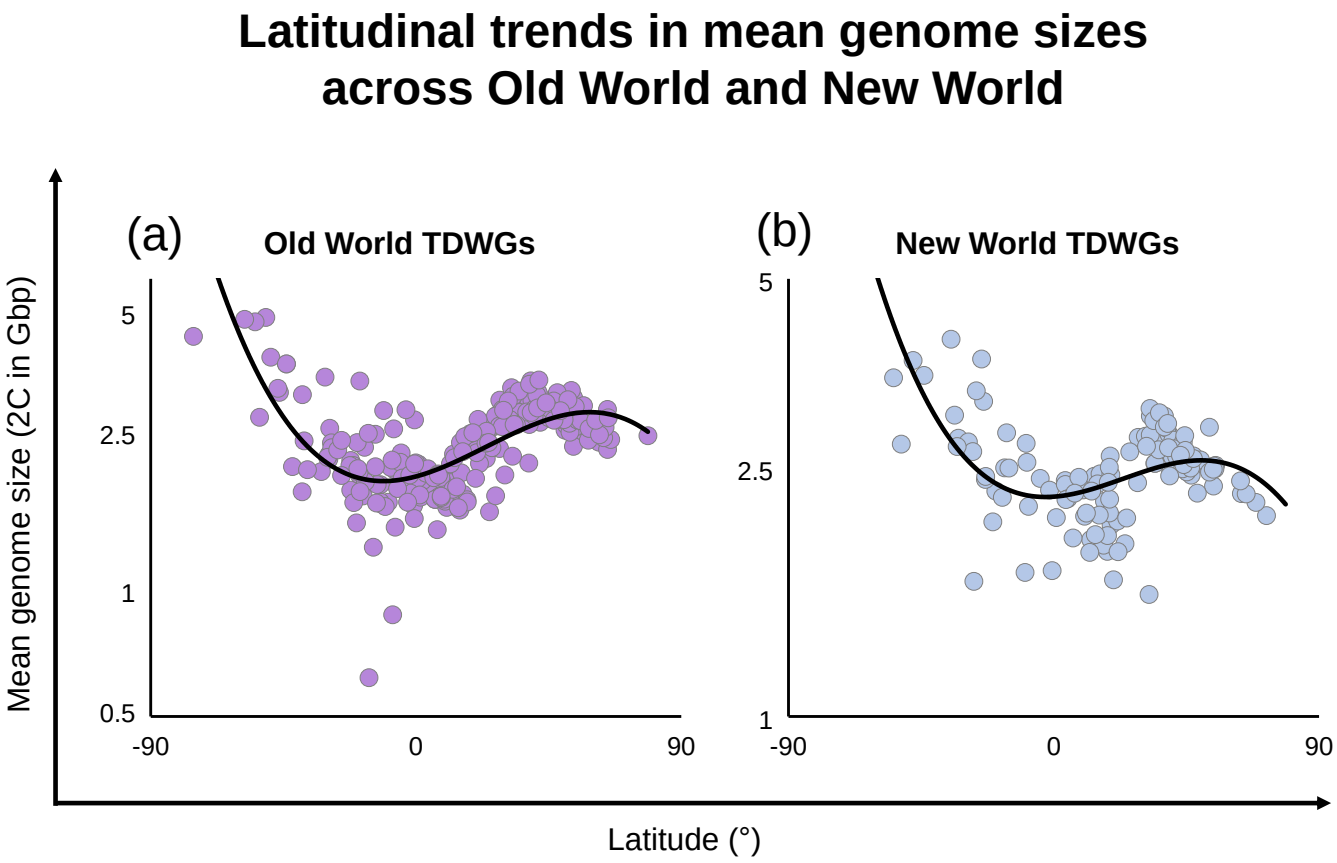


Fig. S7 Mean genome sizes (2C; Gbp) across the global latitudinal gradient for the Old World (a) and New World (b). Circles in both plots represent the genome size averaged per TDWG Level-3 region.

Latitudinal trends in mean genome sizes across glaciated and non-glaciated TDWGs

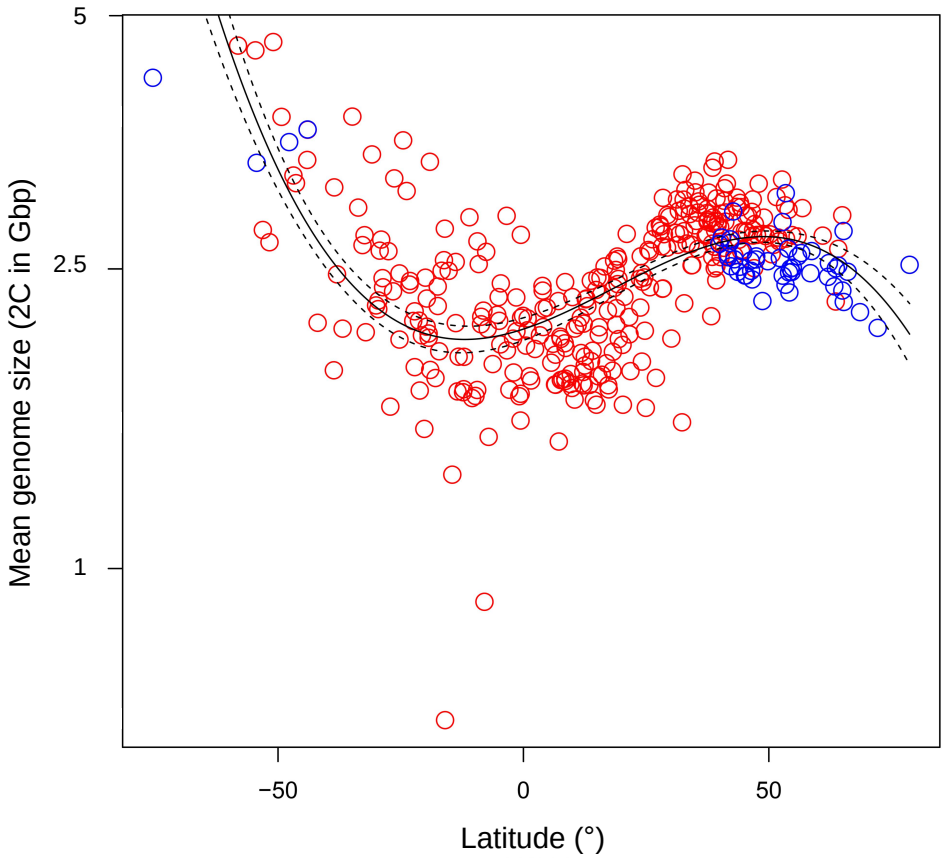


Fig. S8 Mean genome sizes (2C; Gbp) across the global latitudinal gradient illustrating TDWG Level-3 regions that were glaciated (blue) and non-glaciated (red) during the last glacial maximum (LGM) approximately 18,000 years before present.

Table S1 Bioclim variables as they explain the variance in 2C genome size across TDWGs in the polynomial regression of a given order.

	Bioclim variable	N	Order	R^2_{adj}	Explanation of the Bioclim variable
	BIO1	364	2	0.317	Mean annual air temperature
	BIO3	364	3	0.314	Isothermality
	BIO4	364	2	0.212	Temperature seasonality
	BIO5	364	1	0.046	Mean daily maximum air temperature of the warmest month
	BIO6	364	2	0.367	Mean daily minimum air temperature of the coldest month
	BIO7	364	1	0.136	Annual range of air temperature
	BIO8	364	2	0.326	Mean daily mean air temperatures of the wettest quarter
	BIO9	364	2	0.086	Mean daily mean air temperatures of the driest quarter
Group 1	BIO10	364	1	0.085	Mean daily mean air temperatures of the warmest quarter
	BIO11	364	2	0.372	Mean daily mean air temperatures of the coldest quarter
	GST	365	2	0.408	Mean temperature of the growing season
	NGD0	367	1	0.024	Number of days at which mean daily air temperature $>$ 0℃
	NGD5	363	2	0.157	Number of days at which mean daily air temperature $>$ 5℃
	NGS10	358	2	0.326	Number of days at which mean daily air temperature $>$ 10℃
	UVB1	367	2	0.346	Annual Mean UV-B
	UVB3	367	3	0.145	Mean UV-B of Highest Month
	UVB5	367	3	0.185	Sum of Monthly Mean UV-B during Highest Quarter
Group 2	GDD0	367	1	-0.002	Growing degree days heat sum above 0℃
Group 3	BIO2	364	1	0.033	Mean diurnal air temperature range
Group 4	BIO15	364	1	0.034	Precipitation seasonality
Group 5	Aridity	364	1	0.001	Aridity index
	BIO12	364	1	0.066	Annual precipitation amount
	BIO13	364	1	0.126	Precipitation amount of the wettest month
	BIO14	364	1	-0.003	Precipitation amount of the driest month
Group 6	BIO16	364	1	0.118	Mean monthly precipitation amount of the wettest quarter
	BIO17	364	1	-0.002	Mean monthly precipitation amount of the driest quarter
	BIO18	364	1	0.051	Mean monthly precipitation amount of the warmest quarter
	BIO19	364	1	0.016	Mean monthly precipitation amount of the coldest quarter
	GSL	365	1	0.031	Length of the growing season

Variables were selected as predictors of 2C genome size in multiple regression based on the highest explained variance in their group. Aridity and GDD0 were not selected because of low explained variance. GST, BIO2, BIO13, and BIO15 were then included into a backward model selection process to select the best combination of variables predicting 2C genome size (see Methods for details).

Table S2 Results of quantile regression of genome size on range size.

Quantile regression model: $\log_{10}$(2C genome size) $\sim$ $\log_{10}$(Range size)

Quantile	Model term	b_i	se(b_i)	t	P
5%	Intercept	2.899	0.007	412.607	0
	$\log_{10}$(Range size)	-0.088	0.008	-11.669	0
10%	Intercept	3.011	0.006	492.020	0
	$\log_{10}$(Range size)	-0.094	0.006	-16.411	0
15%	Intercept	3.087	0.006	495.326	0
	$\log_{10}$(Range size)	-0.102	0.007	-14.502	0
20%	Intercept	3.146	0.005	584.281	0
	$\log_{10}$(Range size)	-0.091	0.007	-13.615	0
25%	Intercept	3.208	0.007	482.003	0
	$\log_{10}$(Range size)	-0.094	0.007	-14.444	0
30%	Intercept	3.266	0.006	510.241	0
	$\log_{10}$(Range size)	-0.093	0.007	-12.910	0
35%	Intercept	3.319	0.007	509.107	0
	$\log_{10}$(Range size)	-0.090	0.008	-11.845	0
40%	Intercept	3.374	0.007	460.997	0
	$\log_{10}$(Range size)	-0.085	0.008	-10.338	0
45%	Intercept	3.438	0.007	472.582	0
	$\log_{10}$(Range size)	-0.084	0.008	-10.462	0
50%	Intercept	3.491	0.007	475.151	0
	$\log_{10}$(Range size)	-0.080	0.008	-10.180	0
55%	Intercept	3.557	0.009	385.882	0
	$\log_{10}$(Range size)	-0.086	0.009	-9.384	0
60%	Intercept	3.644	0.010	360.321	0
	$\log_{10}$(Range size)	-0.097	0.011	-9.106	0
65%	Intercept	3.720	0.010	355.561	0
	$\log_{10}$(Range size)	-0.089	0.012	-7.413	0
70%	Intercept	3.815	0.012	318.872	0
	$\log_{10}$(Range size)	-0.084	0.013	-6.651	0
75%	Intercept	3.934	0.013	298.276	0
	$\log_{10}$(Range size)	-0.105	0.013	-8.347	0
80%	Intercept	4.077	0.015	272.455	0
	$\log_{10}$(Range size)	-0.133	0.013	-9.934	0
85%	Intercept	4.253	0.015	290.314	0
	$\log_{10}$(Range size)	-0.176	0.014	-12.933	0
90%	Intercept	4.411	0.013	336.729	0
	$\log_{10}$(Range size)	-0.176	0.014	-12.187	0
95%	Intercept	4.607	0.016	296.578	0
	$\log_{10}$(Range size)	-0.167	0.016	-10.398	0

b_i – regression estimates of model terms.

Table S3 Results of phylogenetic quantile regression of genome size on range size.

Phylogenetic quantile regression model: $\log_{10}$(2C genome size) $\sim$ $\log_{10}$(Range size)

Quantile	Model term	Slope	Lower 95% CI	Upper 95% CI
5%	$\log_{10}$(Range size)	-0.002	-0.006	0.001
10%	$\log_{10}$(Range size)	-0.003	-0.006	0.000
15%	$\log_{10}$(Range size)	-0.004	-0.008	-0.001
20%	$\log_{10}$(Range size)	-0.006	-0.008	-0.003
25%	$\log_{10}$(Range size)	-0.007	-0.009	-0.005
30%	$\log_{10}$(Range size)	-0.008	-0.011	-0.006
35%	$\log_{10}$(Range size)	-0.010	-0.012	-0.008
40%	$\log_{10}$(Range size)	-0.011	-0.013	-0.009
45%	$\log_{10}$(Range size)	-0.012	-0.015	-0.010
50%	$\log_{10}$(Range size)	-0.013	-0.017	-0.011
55%	$\log_{10}$(Range size)	-0.015	-0.018	-0.012
60%	$\log_{10}$(Range size)	-0.017	-0.020	-0.013
65%	$\log_{10}$(Range size)	-0.018	-0.022	-0.014
70%	$\log_{10}$(Range size)	-0.020	-0.023	-0.016
75%	$\log_{10}$(Range size)	-0.021	-0.025	-0.017
80%	$\log_{10}$(Range size)	-0.023	-0.027	-0.018
85%	$\log_{10}$(Range size)	-0.024	-0.029	-0.019
90%	$\log_{10}$(Range size)	-0.026	-0.031	-0.020
95%	$\log_{10}$(Range size)	-0.027	-0.032	-0.021

The intercept value is always zero because the phylogenetic quantile regression works with phylogenetic independent contrasts whose regression has to be forced through the origin. 95%CI – lower and upper 95% confidence intervals of the regression estimates.

Table S4 Results of OLS and PGLS regressions of mean chromosome size on range size.

OLS model: $\log_{10}$(mean chromosome size) $\sim$ $\log_{10}$(Range size)

Model term	b_i	95% CI	t	P	$R^2{}_{adj}$
Intercept	2.217	2.201, 2.232	278.40	<2e-16	0.021
$\log_{10}$(Range size)	-0.141	-0.157, -0.125	-17.14	<2e-16	

PGLS model: $\log_{10}$(mean chromosome size) $\sim$ $\log_{10}$(Range size)

Model term	b_i	95% CI	P	lambda	$R^2{}_{adj}$
Intercept	2.066	2.066, 2.067	1.57E-58	0.962	0.004
$\log_{10}$(Range size)	-0.023	-0.024, -0.023	4.32E-09		

b_i – regression estimates of model terms; 95%CI – lower and upper 95% confidence intervals of the regression estimates; $R^2{}_{adj}$ – R squared adjusted indicating explained variance. The OLS analysis was conducted with 13,996 species, while the PGLS analysis was performed with 10,670 species, iterated over 100 different phylogenetic trees (see Methods). Therefore, the values for PGLS are averages across these 100 regressions.

Table S5 Results of quantile regression of mean chromosome size on range size.

Quantile regression model: $\log_{10}$(mean chromosome size) $\sim$ $\log_{10}$(Range size)

Quantile	Model term	b_i	se(b_i)	t	P
5%	Intercept	1.491	0.010	152.814	0
	$\log_{10}$(Range size)	-0.196	0.014	-14.364	0
10%	Intercept	1.601	0.007	240.353	0
	$\log_{10}$(Range size)	-0.148	0.010	-14.692	0
15%	Intercept	1.659	0.006	268.413	0
	$\log_{10}$(Range size)	-0.111	0.008	-14.814	0
20%	Intercept	1.713	0.006	299.498	0
	$\log_{10}$(Range size)	-0.097	0.007	-13.752	0
25%	Intercept	1.754	0.005	344.908	0
	$\log_{10}$(Range size)	-0.082	0.006	-13.073	0
30%	Intercept	1.797	0.007	268.692	0
	$\log_{10}$(Range size)	-0.075	0.007	-11.056	0
35%	Intercept	1.857	0.009	213.299	0
	$\log_{10}$(Range size)	-0.082	0.008	-9.853	0
40%	Intercept	1.932	0.010	202.68	0
	$\log_{10}$(Range size)	-0.093	0.009	-9.981	0
45%	Intercept	2.000	0.009	229.814	0
	$\log_{10}$(Range size)	-0.096	0.009	-11.055	0
50%	Intercept	2.065	0.010	216.425	0
	$\log_{10}$(Range size)	-0.101	0.009	-10.771	0
55%	Intercept	2.136	0.010	206.85	0
	$\log_{10}$(Range size)	-0.103	0.011	-9.59	0
60%	Intercept	2.216	0.012	179.613	0
	$\log_{10}$(Range size)	-0.101	0.013	-8.028	0
65%	Intercept	2.314	0.015	158.329	0
	$\log_{10}$(Range size)	-0.103	0.014	-7.37	0
70%	Intercept	2.450	0.016	150.79	0
	$\log_{10}$(Range size)	-0.126	0.016	-7.864	0
75%	Intercept	2.567	0.013	193.387	0
	$\log_{10}$(Range size)	-0.128	0.014	-9.406	0
80%	Intercept	2.714	0.019	139.61	0
	$\log_{10}$(Range size)	-0.152	0.016	-9.471	0
85%	Intercept	2.925	0.021	142.058	0
	$\log_{10}$(Range size)	-0.215	0.017	-12.974	0
90%	Intercept	3.214	0.019	168.46	0
	$\log_{10}$(Range size)	-0.303	0.018	-17.297	0
95%	Intercept	3.424	0.013	257.932	0
	$\log_{10}$(Range size)	-0.280	0.017	-16.527	0

b_i – regression estimates of model terms.

Table S6 Results of phylogenetic quantile regression of mean chromosome size on range size.

Phylogenetic quantile regression model: $\log_{10}$(2C/2n mean chromosome size) $\sim$ $\log_{10}$(Range size)

Quantile	Model term	Slope	Lower 95% CI	Upper 95% CI
5%	$\log_{10}$(Range size)	-0.003	-0.007	0.000
10%	$\log_{10}$(Range size)	-0.004	-0.007	0.000
15%	$\log_{10}$(Range size)	-0.005	-0.008	-0.002
20%	$\log_{10}$(Range size)	-0.006	-0.008	-0.003
25%	$\log_{10}$(Range size)	-0.007	-0.009	-0.004
30%	$\log_{10}$(Range size)	-0.008	-0.010	-0.006
35%	$\log_{10}$(Range size)	-0.008	-0.010	-0.007
40%	$\log_{10}$(Range size)	-0.009	-0.011	-0.008
45%	$\log_{10}$(Range size)	-0.010	-0.012	-0.008
50%	$\log_{10}$(Range size)	-0.011	-0.013	-0.009
55%	$\log_{10}$(Range size)	-0.012	-0.014	-0.010
60%	$\log_{10}$(Range size)	-0.013	-0.015	-0.011
65%	$\log_{10}$(Range size)	-0.014	-0.017	-0.012
70%	$\log_{10}$(Range size)	-0.016	-0.018	-0.013
75%	$\log_{10}$(Range size)	-0.017	-0.020	-0.014
80%	$\log_{10}$(Range size)	-0.018	-0.022	-0.015
85%	$\log_{10}$(Range size)	-0.020	-0.023	-0.016
90%	$\log_{10}$(Range size)	-0.021	-0.025	-0.017
95%	$\log_{10}$(Range size)	-0.022	-0.027	-0.018

The intercept value is always zero because the phylogenetic quantile regression works with phylogenetic independent contrasts whose regression has to be forced through the origin. 95%CI – lower and upper 95% confidence intervals of the regression estimates.

Table S7 Results of regressions of 2C genome size on different variables discussed in the Discussion.

Polynomial regression (N=367): $\log_{10}$(2C genome size) $\sim$ UVB1 + UVB1^2

Model term	b_i	95% CI	t	P	R^2_{adj}
Intercept	3.414	<3.361, 3.467>	127.37	<2E-16	0.346
UVB1	3.89E-05	<5.95E-06, 7.18E-05>	2.320	0.021	
UVB1^2	-1.03E-08	<-1.48E-08, -5.84E-09>	-4.533	<8E-06	

Polynomial regression (N=362): $\log_{10}$(2C genome size) $\sim$ GST + GST^2 + BIO11 + BIO11^2

Model term	b_i	95% CI	t	P	R^2_{adj}
Intercept	3.508	<3.455, 3.562>	128.493	<2E-16	0.464
GST	-2.518E-03	<-9.0499E-03, 4.0133E-03>	-0.758	0.449	
GST^2	-1.760E-04	<3.8406E-04, 3.1972E-05>	-1.664	0.097	
BIO11	3.941E-04	<-6.5567E-04, 1.4439E-03>	0.738	0.461	
BIO11^2	-4.994E-05	<-9.9424E-05, -4.4982E-07>	-1.985	0.048	

Polynomial regression (N=365): $\log_{10}$(2C genome size) $\sim$ GST + GST^2 + $\log_{10}$(range size)

Model term	b_i	95% CI	t	P	R^2_{adj}
Intercept	3.619	<3.567, 3.671>	136.876	<2E-16	0.516
GST	7.487E-03	<0.003, 0.012>	3.243	0.001	
GST^2	-4.624E-04	<-0.001, -0.0003>	-6.847	3.26E-11	
$\log_{10}$(Range size)	-0.129	<0.157, -0.101>	-9.06	<2E16	

Polynomial regression (N=365): $\log_{10}$(2C genome size) $\sim$ GSL

Model term	b_i	95% CI	t	P	R^2_{adj}
Intercept	3.426	<3.401, 3.450>	272.350	<2E-16	0.031
GSL	-1.62E-05	<-2.520E-04, -7.620E-04>	-3.561	0.0004	

Polynomial regression (N=47): $\log_{10}$(2C genome size) $\sim$ GSL [only for TDWGs from 48.93° northward]

Model term	b_i	95% CI	t	P	R^2_{adj}
Intercept	3.413	<3.386, 3.440>	256.050	<2E-16	-0.011
GSL	4.94E-05	<-9.06E-05, 1.89E-04>	0.711	0.481	

UVB1 – mean annual UVB; GST – mean temperature of growing season; range size – range size of species presented in TDWGs for which we had 2C genome size data; GSL – mean length of the growing season; b_i – regression estimates of model terms; 95% CI – lower and upper 95% confidence intervals of the regression estimates; R^2_{adj} – R squared adjusted indicating explained variance.