The global distribution of angiosperm genome size is shaped by climate

Petr Bureš[1], Tammy L. Elliott[1,2], Pavel Veselý[1], Petr Šmarda[1], Félix Forest[3], Ilia J. Leitch[3], Eimear Nic Lughadha[3], Marybel Soto Gomez[3], Samuel Pironon[3,4], Matilda J. M. Brown[3], 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]Department of Biological Sciences, University of Cape Town, Cape Town 7700, South Africa

[3]Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, United Kingdom

[4]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

Word count: 6443
Introduction: 1507
Materials and Methods: 1990
Results: 1170
Discussion: 1776

No. of figures: 5
No. of tables: 2

Supporting Information
Supporting Figures: 13
Supporting Tables: 14
Supporting Datasets: 2

Summary

(1) Angiosperms, which inhabit diverse environments across all continents, exhibit significant variation in genome sizes, making them an excellent model system for examining hypotheses about the global distribution of genome size. These include the previously proposed large-genome-constraint, mutational-hazard, polyploidy-mediated, and climate-mediated hypotheses.

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

(3) We observed that angiosperms with large range sizes generally had small genomes, supporting the large-genome-constraint hypothesis. Climate was shown to exert a strong influence on genome size distribution along the global latitudinal gradient, while the frequency of polyploidy and the type of growth form had negligible effects. In contrast to the unimodal patterns along the global latitudinal gradient shown by plant size traits and polyploid proportions, the increase in angiosperm genome size from the equator to 40-50°N/S is probably mediated by different (mostly climatic) mechanisms than the decrease in genome sizes observed from 40–50°N northwards.

(4) Our analysis suggests that the global distribution of genome sizes in angiosperms is mainly shaped by climatically-mediated purifying selection, genetic drift, relaxed selection, and environmental filtering.

Keywords: C-value, chromosome size, geographic range size, flowering plants, glaciation, latitudinal gradient, large genome constraint hypothesis, nuclear DNA content, polyploid proportion, 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. Angiosperms exhibit tremendous variation in genome sizes (more than 2,400-fold; Pellicer *et al.*, 2018) and are found across all continents, with the majority of species being narrow endemics while a minority are widespread cosmopolitan species (Enquist *et al.*, 2019). This makes angiosperms a powerful model system for studying the underlying drivers that shape genome size evolution and its distribution across the globe. 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 that consistent geographic data has recently become available for most known species through the World Checklist of Vascular Plants (WCVP; Govaerts *et al.*, 2021), it is now possible to examine hypotheses seeking to understand the causal links between angiosperm genome size, distribution, and environment at a global scale.

Key proximal mechanisms generating changes in genome size are polyploidization followed by re-diploidization (Wendel, 2000; Leitch & Leitch, 2008; Soltis *et al.*, 2015; Guignard *et al.*, 2016; Šmarda *et al.*, 2019) and 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).

The 'large-genome-constraint' hypothesis (LGCH) suggests that species with large genomes might face selection pressure against them due to their negative impact on plant anatomy and physiology (Vinogradov, 2003; Knight *et al.*, 2005). This is because more genomic material occupies a larger volume, influencing the minimum cell size (Cavalier-Smith, 2005; Šímová & Herben, 2012; Bhadra *et al.*, 2023). Consequently, plants with larger genomes tend to have larger seeds (Knight & Ackerly, 2002; Beaulieu *et al.*, 2007; Carta *et al.*, 2022; Bhadra *et al.*, 2023), a trait linked to smaller distributional ranges (Sonkoly *et al.*, 2022). Additionally, they possess larger stomatal guard cells (Beaulieu et al., 2008; Veselý *et al.*, 2012; Bhadra *et al.*, 2023), which close and open more slowly (Drake et al., 2013; Kardiman & Ræbild, 2018; Lawson & Matthews, 2020). This might be disadvantageous in, for example, arid environments that demand efficient water management (Veselý *et al.*, 2020; Bureš *et al.*, 2023; Šmarda *et al.*, 2023). Larger cells also limit the mesophyll surface area packed into the leaf volume leading to lower CO_2 diffusion and rates of photosynthesis

(Théroux-Rancourt *et al.*, 2021). Species with large genomes also experience slower rates of cell division (Francis *et al.*, 2008; Šímová & Herben, 2012) and have higher phosphorus (P) and/or nitrogen (N) requirements (Šmarda *et al.*, 2013; Peng *et al.*, 2022). Large genomes may thus limit species' dispersal abilities and have narrower ecological niches, potentially resulting in smaller geographic ranges (Sheth *et al.*, 2020). In contrast, smaller genomes offer more flexibility in cell size (Beaulieu *et al.*, 2007; Beaulieu *et al.*, 2008; Veselý *et al.*, 2012; Meyerson *et al.*, 2020; Bhadra *et al.*, 2023), have faster rates of cell division (Francis *et al.*, 2008; Šímová & Herben, 2012), and lower P and N demands (Šmarda *et al.*, 2013; Peng *et al.*, 2022) allowing greater plasticity in range size.

Although TE insertions can occasionally have adaptive effects (Casacuberta & González, 2013; Schrader & Schmitz, 2019), they are mostly neutral or deleterious (Deniz *et al.*, 2019). Thus, TE insertions mostly become fixed via genetic drift rather than by natural selection or intragenomic selection favoring TE accumulation (Werren, 2011; Deniz *et al.*, 2019). As the relative importance of natural selection versus random genetic drift depends on population size, the mutational-hazard hypothesis (MHH) posits that genome growth via TEs occurs 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). The relative importance of natural selection and genetic drift also appears to hold for species range size in both plants and animals (Corbett-Detig *et al.*, 2015), likely because of the positive abundance-occupation relationship (Gaston *et al.*, 2002) where species with larger populations tend to have large distributional ranges (e.g., Brown, 1984; Johnson, 1998; Gaston, 2003; Webb *et al.*, 2012; Drovetski *et al.*, 2014; Spence *et al.*, 2021; Guo *et al.*, 2022; Ten Caten *et al.*, 2022).

Considering the potential effects of genetic drift and natural selection on genome size and their interplay with range size, the LGCH predicts that species with large ranges should not have large genomes, resulting in a triangular relationship (Fig. 1a). On the other hand, the MHH predicts that species genome sizes should decrease with increasing geographic ranges, producing a negative relationship (Fig. 1b). Although effective population size is affected by complex factors and range size is a relatively crude proxy, the high statistical power provided by the large amount of currently available data on species genome size and distribution should help overcome this imprecision.

Polyploidization is another major contributor to plant genome size evolution (Wendel, 2000; Leitch & Leitch, 2008; Soltis *et al.*, 2015; Guignard *et al.*, 2016; Šmarda *et al.*, 2019), which, in newly formed polyploids (neopolyploids), leads to multiplication of the genome size and chromosome number (Mandáková & Lysák, 2018). However, over time, polyploids undergo post-polyploid diploidization that includes chromosome fusions and genome downsizing (Mandáková & Lysák, 2018), thereby blurring the clear correlation between genome size and chromosome number (Choi *et al.*, 2020; Roddy *et al.*, 2020). Because not all of the duplicated portion of the genome is eliminated during the post-polyploid diploidization (Bowers *et al.*, 2003; Paterson *et al.*, 2004; Wang *et al.*, 2015), repeated polyploidization-diploidization cycles may lead to a gradual increase in genome size over time, especially in regions where polyploids originate more frequently. The proportion of neopolyploids at different latitudes across the globe shows a U-shaped pattern, being low in the tropics and increasing polewards (Rice *et al.*, 2019). The latitudinal U-shape in the proportion of neopolyploids 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 at 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 rate of the repeated polyploidization-diploidization cycles (Wendel, 2015; Wendel *et al.*, 2016; Clark & Donoghue, 2017) should persist over geological time scales and a U-shaped latitudinal distribution of genome size would gradually emerge in this scenario (Fig. 1c; polyploid-mediated hypothesis: PMH).

Latitudinal gradients encompass climatic and other environmental variables that could also be important factors contributing to genome size variation. These factors include temperature, precipitation, aridity, seasonality, ultraviolet-B radiation (UV-B), and 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 distributions 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 influence where plants grow. For example, in the case of temperature, large genomes might be predicted to be excluded from areas with both the lowest and highest mean temperature (underpinned, in part, by the impact of genome size on the rate of cell division) (Fig. 1d; *climate-mediated hypothesis*: CMH).

Here, we test the following hypotheses (Fig. 1): 1) the *large-genome constraint hypothesis* (LGCH), which predicts species which occupy large geographical ranges cannot have large genomes; 2) the *mutational-hazard hypothesis* (MHH), which predicts that genome size decreases with increasing geographic range size; 3) the *polyploid-mediated hypothesis* (PMH), which predicts an increase in genome size from the equator to the poles; and 4) the *climate-mediated hypothesis* (CMH), which predicts the exclusion of large genomes from both ends of the climatic spectrum. We achieve this by combining the largest dataset compiled to date for angiosperm genome size (16,017 species) with newly-available data on the global distribution of angiosperms from the WCVP, and mapping the global distribution of angiosperm 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 (WCVP; 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. 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 exceptional cases when the WCVP taxonomic framework differed from the 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 the genome size data (Dataset S1). We discarded taxa that were imprecisely 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).

Distributional range size estimation

Distribution range sizes were calculated as the extent of occurrence (EOO) for each species based on the Global Biodiversity Information Facility (GBIF) distribution data. To obtain EOO estimates in square kilometers, we first cleaned the data for species occurrences from GBIF following Elliott *et al.* (2022). Then, we calculated EOO (Dataset S1) using the 'eoo' function in the R package RANGEMAP v.0.1.18 (Cobos *et al*., 2022), with the 'polygons' option set to 'simple_wmap("simplest")' to omit oceans from the calculations. In addition, as an alternative measure of range size, we calculated the number of occupied TDWGs flagged as native for each species (Dataset S1).

Genome size compilation

We extracted genome size estimates from several sources, including (1) research papers published between 2012 and 2022 (or older studies that were absent from Release 8.0 of the 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. For multiple estimates presented for a species in the same publication, the genome size values were averaged. Finally, in cases where publications used nomenclature that conflicted 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 = 978 Mbp (Doležel *et al.*, 2003).The genome size per TDWG was calculated as the average of the 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: (i) the same publications as the genome size data when both estimates were reported together; (ii) the Chromosome Counts Database (CCDB: Rice *et al.*, 2015); and (iii) publications reporting only chromosome number (Dataset S1). We first ensured the estimations were not pseudo-replicated and then we selected the most prevalent number for a species. We report the median value for a species when it was not possible 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 by dividing the 2C genome size (in Mbp) by the diploid (2n) chromosome number. As mean chromosome size removes the correlation between genome size and chromosome number, we used it 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 used in tests of MHH and LGCH

We used one hundred species-level trees of all angiosperms comprising all 329,798 species recognized by version 6 of the World Checklist of Vascular Plants (Forest, 2023) pruned to species in our dataset.

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): (i) annuals (= therophytes; 12 % of species in the dataset), (ii) geophytes (11 %), (iii) non-geophytes (perennial herbs = hemicryptophytes + parasites + hydrophytes + epiphytes; 47 %), (iv) woody plants (= chamaephyte + phanerophytes; 30 %), using standard floras or The World Checklist of Selected Plant Families (WCSP, 2017). For each TDWG, we calculated the percentage of species belonging to the four growth forms (Dataset S2).

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, 2018) 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., UV-B-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.

Statistical analyses

We applied a series of linear regressions to test our four hypotheses (Fig. 1). The LGCH and MHH were modeled with 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). In PHYLOLM, we used the weighted Akaike information criterion (AICw; Akaike, 1978; Wagenmakers & Farrel, 2004) to select between seven evolutionary explicit models of trait evolution: Brownian motion, Pagel’s lambda, kappa, and delta, two Ornstein-Uhlenbeck models with an ancestral state estimated at the root or having the stationary distribution at the root, and the early burst model. The best model was Pagel’s lambda with AICw = 1 (averaged across all 100 trees), which we used to optimize branch lengths based on the data (model = ‘lambda’) using maximum likelihood estimation. To examine whether the association between range size 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 the best of our knowledge, a tool has 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), employing R packages CAPER v.1.0.1 (Orme, 2013) and QUANTREG v.5.93 (Koenker *et al.*, 2022).

To examine how genome size is associated with latitude (testing the PMH and CMH hypotheses), 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 AIC values 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 the ratio of the number of species for which we have genome size data (or the proportion of polyploids) and the number of all species in the TDWG (Dataset S2). To evaluate causal relationships between the effects of GST and percentage of growth forms on mean genome size across TDWGs, we employed a path analysis approach using the R package LAVAAN v.4.2.3 (Rosseel, 2012).

Results

Sampling bias

We compiled the largest genome size dataset to date, encompassing >5% of known angiosperm species (Dataset S1). Large datasets of phylogenetic representation and traits, including genome size data, are latitudinally biased, with northern latitudes being 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 in the WCVP to that of the taxa in our genome size dataset. Both datasets show an increase in range size from south to north (Fig. S2).

Genome size and range size (LGCH, 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 ($b = -0.007$, $P = 1.31e-06$), was flatter than that from the OLS ($b = -0.039$, $P < 2e-16$), due to a strong phylogenetic signal (Pagel's $\lambda = 0.916$) 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; in accordance with the triangular relationship. Although the slopes started decreasing at the genome size quantile 0.5 for the ordinary quantile regression (Fig. 2c), in the phylogenetic quantile regression, the slope decreased continuously with increasing quantiles (Fig. S3b). When we used the number of occupied TDWGs as a measure of range size (instead of the EOO), we observed very similar results (Fig. S4, S5, Tables S4, S5), suggesting that, at least for our dataset, TDWG counts provide a reasonable proxy for range size.

We also 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. S6, Tables S6-S8). For analyses with number of occupied TDWGs as a measure of range size, see Fig. S7, Tables S9-S11). 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 S6).

Genome size, neopolyploidy, latitude, and climate

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. S8. 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 the proportion of polyploid species exhibited very different latitudinal distributions (Fig. 3), with the proportion of polyploid species explaining 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. S9). 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. S10).

To assess which climatic parameters might 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 high or low temperatures and larger for 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 S12), highlighting the importance of freezing temperatures. Furthermore, if the MHH is combined with CMH by adding the range size to the model with GST, the explained variance increases to 46.14 % (Table S12).

We also 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 S12). When analyzed only for TDWGs with latitudinal centroids of at least 48.93° (the threshold at which genomes start decreasing northward), genome size decreases with a shortening of the growing season, but the relationship is not significant (P=0.481; Table S12). UVB1 (mean annual UVB) explained 34.6 % of the variation in mean genome size across TDWGs (Table S12).

Due to genome size variations among different plant growth forms (Bennett, 1987; Beaulieu *et al.*, 2008; Veselý *et al.*, 2013), and the presence of latitudinal trends in growth form proportions (Taylor *et al.*, 2023; Fig. S11 here), we investigated whether the observed S-shape (Fig. 3a) might be attributed to differences in the percentages of different growth forms within TDWGs with increasing latitudes. Annuals, geophytes, and non-geophyte herbs all exhibited the S-shape in mean genome size, varying only in magnitude (Fig. 4). Woody plants, however, had slightly larger genomes in the tropics compared to temperate or arctic regions (Fig. 4). These growth form patterns remained consistent across both species (Fig. 4a) and TDWG means (Fig. 4b-e). As sole predictor, the percentage of growth forms explained from 2% of genome size variance (in annuals) to 21% (in non-geophytes) (Table S13). However, when growth form percentage was added to the model with GST, the effects of non-geophytes and annuals became insignificant, with geophytes and woody plants contributing only 3.4% and 1.4% additional explained variance, respectively (Table S14). This significant drop in the explanatory power of growth forms suggests that GST directly influences both growth form percentages and mean genome size within TDWGs. This was confirmed through path analysis, which revealed that while GST strongly impacts genome size and the percentages of non-geophytes and woody plants, growth forms have minimal or negligible effects on the distribution of genome sizes across the globe (Fig. S12).

Discussion

Support for the LGCH, while not ruling out the MHH

We revealed a triangular relationship between range size and genome size, with a negative association between range size and genome size that is accentuated as genome sizes increase (Fig. 2b, 2c), supporting the LGCH (Fig. 1a). This relationship indicates that large-genomed species are restricted to occupying smaller ranges, which is likely due to the nucleotypic effects of their genomes hindering their dispersal distance and limiting their ecological niche (Knight & Ackerly, 2002; Beaulieu *et al.*, 2007, 2008; Veselý *et al.*, 2012; Carta *et al.*, 2022; Bhadra *et al.*, 2023). This places large-genomed species at a disadvantage compared to their smaller-genomed counterparts that have greater nucleotypic plasticity (Mayerson *et al.*, 2020; Bhadra *et al.*, 2023) and may thus occupy both large and small ranges (Fig. 2a). It is notable that the most pronounced S-shape in the latitudinal distribution of genome size (see *Genome size decreases* [...] *but not in the south* section below) is in geophytes (Fig. 4c), whose genomes are the largest among the analyzed growth forms (Fig. 4a). Although the triangular relationship we observed does not show support for the MHH, the LGCH does not necessarily rule out the MHH. Notably, the largest genomes are found in the southern hemisphere (Fig. 3a), where angiosperms in our dataset have the smallest ranges (Fig. S2) and could thus be most susceptible to genetic drift (Fig. 5). Genetic drift could facilitate genome growth in smaller-ranged species (as proposed in the MHH), which could further reduce the range size of large-genomed species (LGCH) and throw them into a deadly descending spiral toward extinction. This is supported by evidence showing that large-genomed species are at higher risk of extinction (Vinogradov, 2003; Soto Gomez *et al.*, 2023 in this issue).

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 (e.g., Avdulov, 1931; Bennett, 1976), Fabaceae (e.g., Stebbins, 1966; Bennett, 1976; Souza *et al.*, 2019), Orchidaceae (e.g., Trávníček *et al.*, 2019), Zygophyllaceae (e.g., Vidal-Russell *et al.*, 2022) and at broader phylogenetic scales across angiosperms (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), possibly because the maintenance of large cells becomes more difficult with increasing temperatures (Sabath *et al.*, 2013). Our finding that small genomes are prevalent in the tropics 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 low latitudes could be a result of the DNA-damaging effects of UV-B radiation (Bennett, 1976), which is generally highest in the tropics (Beckmann *et al*., 2014) and might result in selecting for smaller chromosomes that absorb less energy, therefore decreasing radiosensitivity (Sparrow *et al.*, 1967). This idea is supported by recent findings showing that plants with holocentric chromosomes, which tolerate fragmentation (Zedek & Bureš, 2019), are less stressed (Zedek *et al.*, 2020; 2021) and more competitive (Zedek *et al.*, 2022) under higher UV-B doses. Moreover, homologous recombination used to repair UV-B-induced damage might increase rates of DNA deletion, thereby further promoting genome downsizing (Schubert & Vu, 2016). However, as UV-B radiation intensity (which explained 34.6 % of the variation) 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 play a role in constraining the genome size of tropical plants, as many tropical soils are low in nutrients (especially phosphorus; Vitousek *et al*., 2010), and yet building and maintaining cells in plants with large genomes is expensive in terms of N and P. This may result in species with large genomes being less competitive in the nutrient-poor tropical soils, resulting in their exclusion from these environments (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 North pole, but not the South

Differences in genome size trends across latitudinal gradients in the northern versus southern hemisphere may be explained by the larger temperature gradient in the north, which could be partially associated with differences in the distribution of landmasses and major water bodies in the two hemispheres. Large areas of Eurasia and North America extend beyond 50°N and are surrounded by less water and more land masses than regions in the southern hemisphere, therefore experiencing weaker buffering effects from the ocean. If the distribution of genome sizes in plants followed a similar pattern to the distribution of polyploid species, genomes would be expected to be smaller near the equator and increase polewards. We find that this trend holds, but only up to a certain, presumably limiting, low temperature threshold, beyond which genome sizes decrease towards the high northern latitudes. In the southern hemisphere, this low-temperature threshold is probably not reached (Fig. 1a). The existence of such a latitudinal breakpoint in genome size was previously predicted (Bennett *et al.*, 1982). The main drivers of selection pressure against larger genomes in polar regions were predicted to be: (i) slower cell divisions mediated by lower temperatures (Francis & Barlow, 1988) and (ii) longer generation times mediated by lower temperatures and/or by shorter growing seasons (Bennett *et al.*, 1982; Bennett, 1987). Indeed, temperature variables alone explain a relatively large proportion (up to ~40 %) of the variation in the global distribution of genome sizes (Table 2, Table S1).

Several authors have hypothesized that the decrease in temperatures toward the poles can result in a 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 the 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 a difference in the proportion of polyploids between the southern and northern hemispheres (Fig. 3b), suggesting that unreduced gamete production might not be an important variable in explaining latitudinal variation in polyploidy (but see below).

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. S13). During glacial migrations, species with large genomes might have been more prone to extinction 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 the proportion of polyploids 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 polyploids 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). Nevertheless, our results show that any effect of length of growing season in the Arctic on genome size, is likely to be minor (Table S12).

Relatively large genomes in temperate regions

Temperate regions offer mild conditions between the extremes of the tropics and arctic regions discussed above. For instance, there are not very high nor low temperatures, lower doses of UV-B radiation than in the tropics, and the area was not as extensively glaciated as arctic regions. The temperate climate might thus relax selective pressures against larger genomes, thereby increasing the overall range and mean genome sizes of plants growing in temperate zones of both hemispheres (Fig. 4a).

Latitudinal gradient in genome size is not underpinned by contrasting proportions of different growth forms in different regions

Although the proportion of growth forms, especially perennial herbs and woody plants, may have significantly contributed to the global distribution of polyploids (Rice *et al.*, 2019), the impact of different growth forms on the global distribution of genome size appears weak and mostly mediated by temperature (Fig. S12, Table S14). The independence of global genome size distribution on growth forms is further supported by the observation that annuals, geophytes, and non-geophytes all exhibited the S-shape in mean genome size (Fig. 4). Woody plants showed a different pattern, but their genome size still decreased 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 why their genomes did not increase in temperate regions. Also, the absence of relationship between extinction risk and genome size in woody plants (Soto Gomez *et al.*, 2023) could suggest that genome size dynamics operate differently in woody vs herbaceous species.

Conclusions and future directions

Our study found support for the large genome constraint hypothesis in explaining the global distribution of genome sizes but could not rule out the mutation hazard hypothesis in also contributing to explaining the distribution patterns observed. In addition, we show a small effect of polyploidy and growth forms and a large effect of climate, especially temperature, on the distribution of genome size. Overall, our findings indicate that mainly purifying selection, genetic drift, relaxed selection, and environmental filtering influenced by climate are likely to have shaped the global distribution of angiosperm genomes sizes (Fig. 5). Further research should be directed at determining the relative contributions of long-term processes shaping the global distribution of genome sizes, such as glaciation cycles, UV-B-caused genome erosion, or polyploidization-rediploidization cycles. We also advocate more thorough investigation of links between environmental factors and genome size at finer regional or local scales. For instance, the use of vegetation plots combined with species Ellenberg indicator values would enable a more in-depth understanding of the complex interplay between genome size and both biotic (e.g., competition) and abiotic (e.g., altitude, temperature, soil reaction and moisture) factors in influencing a species habitat and niche and its resilience to environmental changes.

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 analysis, and PV assigned growth forms and contributed to the analysis. FF prepared phylogenetic trees. JŠ performed flow cytometric measurement generating unpublished data used in this study. PŠ, FF, IJL, ENL, MSG, SP, and MJMB contributed to 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. This work was financially supported by the Czech Science Foundation, grant no. GA20-15989S.

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

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.

Bhadra S, Leitch IJ, Onstein RE. 2023. From genome size to trait evolution during angiosperm radiation. *Trends in Genetics*: TIG, S0168-9525(23)00164-6. Advance online publication. doi: 10.1016/j.tig.2023.07.006.

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.

Bowers JE, Chapman BA, Rong JK, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438. doi: 10.1038/nature01521.

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.

Bureš P, Ozcan M, Šmerda J, Michálková E, Horová L, Plačková K, Šmarda P, Elliott TL, Veselý P, Ćato S, Norouzi M, Sheidai M, František Zedek. 2023. Evolution of genome size and GC content in the tribe Carduinae (Asteraceae): rare descending dysploidy and polyploidy, limited environmental control and strong phylogenetic signal. *Preslia* **95**: 185–213. doi: 10.23855/preslia.2023.185.

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.

Choi IY, Kwon EC, Kim NS. 2020. The C- and G-value paradox with polyploidy, repeatomes, introns, phenomes and cell economy. *Genes Genomics* **42**: 699–714. doi: 10.1007/s13258-020-00941-9.

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.

Cobos ME, Barve V, Barve N, Jiménez-Valverde A, Nuñez-Penichet C. 2022. “Rangemap: An R Package to Explore Species’ Geographic Ranges”. *Biodiversity Informatics* **17**: 59–66. https://doi.org/10.17161/bi.v17i.16271.

Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biology* **13**: e1002112. doi: 10.1371/journal.pbio.1002112.

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.

Drake PL, Froend RH, Franks PJ. 2013. Smaller, faster stomata: scaling of stomatal size, rate of response, and stomatal conductance. *Journal of Experimental Botany* **64**: 495–505. doi: 10.1093/jxb/ers347.

Drovetski SV, Aghayan SA, Mata VA, Lopes RJ, Mode NA, Harvey JA, Voelker G. 2014. Does the niche breadth or trade-off hypothesis explain the abundance-occupancy relationship in avian Haemosporidia? *Molecular Ecology* **23**: 3322–3229. doi: 10.1111/mec.12744.

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* **130**: 999–1014. 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.

Forest F. 2023. Species-level phylogenetic trees of all angiosperm species (100 trees). **https://zenodo.org/record/7600341.**

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, Greenwood JJ, Gregory RD, Quinn RM, Lawton JH. 2002. Abundance–occupancy relationships. *Journal of Applied Ecology* **37**: 39–59. doi: 10.1046/j.1365-2664.2000.00485.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.

Grime JP, Shacklock JML, Brand SR. 1985. Nuclear DNA contents, shoot phenology and species co-existence in a limestone grassland community. *New Phytologist* **100**: 435–445.

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.

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.

Guo Q, Qian H, Zhang J. 2022. On the relationship between species diversity and range size. *Journal of Biogeography* **49**: 1911–1919. doi: 10.1111/jbi.14477.

Gustafsson A. 1948. Polyploidy, life form and vegetative reproduction. *Hereditas* **34:** 1–22. doi: 10.1111/j.1601-5223.1948.tb02824.x.

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.

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.

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.

Kardiman R, Ræbild A. 2018. Relationship between stomatal density, size and speed of opening in Sumatran rainforest species. *Tree Physiology* **38**: 696–705.

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.

Lawson T, Matthews J. 2020. Guard Cell Metabolism and Stomatal Function. *Annual Review of Plant Biology* **71**: 273–302. doi: 10.1146/annurev-arplant-050718-100251.

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.

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.

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.

Meyerson LA, Pyšek P, Lučanová M, Wigginton S, Tran CT, Cronin JT. 2020. Plant genome size influences stress tolerance of invasive and native plants via plasticity. *Ecosphere* **11**: e03145. doi: 10.1002/ecs2.3145.

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.

Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences, USA* **101**: 9903–9908. doi: 10.1073/pnas.0307901101.

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.

Peng Y, Yang J, Leitch IJ, Guignard MS, Seabloom EW, Cao D, Zhao F, Li H, Han X, Jiang Y, Leitch AR, Wei C. 2022. Plant genome size modulates grassland community responses to multi-nutrient additions. *New Phytologist* **236**: 2091–2102. doi: 10.1111/nph.18496.

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.

Rosseel Y. 2012. "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* **48**: 1–36. doi: 10.18637/jss.v048.i02.

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.

Schrader L, Schmitz, J. 2019. The impact of transposable elements in adaptive evolution. *Molecular Ecology* **28**: 1537–1549.

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.

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.

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.

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.

Soto Gomez M, Brown MJM, Pironon S, Veselý P, Bureš P, Elliott TL, Zedek F, Pellicer J, Forest F, Nic Lughadha E, Leitch IJ. 2023. Genome size is positively correlated with extinction risk in herbaceous angiosperms. *New Phytologist* **XXX**: XXX–XXX.

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.

Spence ES, Fant JB, Gailing O, Griffith MP, Havens K, Hipp AL, Kadav P, Kramer A, Thompson P, Toppila R, Westwood M, Wood J, Zumwalde BA, Hoban S. 2021. Comparing Genetic Diversity in Three Threatened Oaks. *Forests* **12**: 561. doi: 10.3390/f12050561.

Stebbins GL. 1940. The significance of polyploidy in plant evolution. *American Naturalist* **74:** 54–66.

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.

Šímová I, Herben T. 2012. Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *Proceedings of the Royal Society B: Biological Sciences* **279**: 867–875. doi: 10.1098/rspb.2011.1284.

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.

Šmarda P, Klem K, Knápek O, Veselá B, Veselá K, Holub P, Kuchař V, Šilerová A, Horová L, Bureš P. 2023. Growth, physiology, and stomatal parameters of plant polyploids grown under ice age, present-day, and future CO2 concentrations. *New Phytologist* in press. doi: 10.1111/nph.18955.

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.

Taylor A, Weigelt P, Denelle P, Cai L, Kreft H. 2023. The contribution of plant life and growth forms to global gradients of vascular plant diversity. *New Phytologist*: in press. doi: 10.1111/nph.19011.

Ten Caten C, Holian LA, Dallas T. 2022. Effects of occupancy estimation on abundance-occupancy relationships. *Biology Letters* **18**: 20220137. doi: 10.1098/rsbl.2022.0137.

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.

Théroux-Rancourt G, Roddy AB, Earles JM, Gilbert ME, Zwieniecki MA, Boyce CK, Tholen D, McElrone AJ, Simonin KA, Brodersen CR. 2021. Maximum CO2 diffusion inside leaves is limited by the scaling of cell size and genome size.

Proceedings of the Royal Society B: Biological Sciences **288**: 20203145. doi: 10.1098/rspb.2020.3145.

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.

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.

Wagenmakers EJ, Farrell S. 2004. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* **11**: 192–196. doi: 10.3758/bf03206482.

Wang X, Wang J, Jin D, Guo H, Lee T-H, Liu T, Paterson AH. 2015. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Molecular Plant* **8**: 885–898. doi: 10.1016/j.molp.2015.04.004.

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

Webb TJ, Freckleton RP, Gaston KJ. 2012. Characterizing abundance–occupancy relationships: there is no artefact. *Global Ecology and Biogeography* **21**: 952–957. doi: 10.1111/j.1466-8238.2011.00736.x.

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.

Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences USA* **108**(Suppl 2): 10863–10870. doi: 10.1073/pnas.1102343108.

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.

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, b) range size and (c, d) latitude based on four hypotheses outlined in the Introduction. The question mark in (d) 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). The association of the raw data between genome and range size is shown in (a), whereas both variables are log-transformed in the other two plots (b, c). The slope estimates from the quantile regression, including 95% confidence intervals (dark grey), are indicated in (c). The solid red line in (b) indicates 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) represent 95% confidence intervals.

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 and dark blue indicate TDWG regions with the highest and lowest temperatures in the coldest quarter, respectively (BIO11 from CHELSA). 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 the ratio of the number of species for which we have genome size data (or the proportion of polyploids) to the 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 Proposed major factors (to the left of arrows) 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 contribute to the observed latitudinal trend in genome size (see Table 2).

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 (as Extent of Occurrence, EOO) 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 represent 95% confidence intervals. Both genome and range size are transformed by $\log_{10}$ in (a) and (b).

Fig. S4 Associations between genome and range size per species when the number of occupied TDWG regions (instead of Extent of Occurrence, EOO) is used as a measure of range size. The association of the raw data between genome and range size is shown in (a), whereas both variables are log-transformed in the other two plots (b, c). The slope estimates from the quantile regression, including 95% confidence intervals (dark grey), are indicated in (c). The solid red line in (b) indicates 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 represent 95% confidence intervals.

Fig. S5 Associations between genome and range size per species considering phylogenetic relationships when the number of occupied TDWG regions (instead of the Extent of Occurrence, EOO) is used as a measure of range size. 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 represent 95% confidence intervals. Both genome and range size are transformed by log10 in (a) and (b).

Fig. S6 Associations between mean chromosome size and range size (as Extent of Occurrence, EOO) 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. S7 Associations between mean chromosome size and range size (as the number of occupied TDWG regions) 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. S8 Mean genome sizes (2C; Gbp) averaged per TDWG region for the two most species-rich monocot (a – Orchidaceae, b – Poaceae) and dicot (c – Asteraceae, d – Fabaceae) families. Dark red colors indicate relatively large mean genome sizes, whereas light yellow shades indicate TDWG regions with relatively small mean genome sizes.

Fig. S9 The global distribution of mean chromosome size in flowering plants calculated per TDWG 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. S10 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 region.

Fig. S11 Latitudinal distribution of the percentage of (a) nongeophyte, (b) annual, (c) geophyte, and (d) woody species in our genome size dataset (Dataset S2).

Fig. S12 Path analysis of causal relationships among the effects of the growing season temperature (GST) and percentages of species of different growth forms on the mean genome size in TDWG regions: (a) nongeophytes, (b) annuals, (c) geophytes, and (d) woody species. The numbers indicate standardized regression coefficients from the path analyses. The arrows show the direction of the causal effects, their thickness indicates the relative effects, the fading indicates significance of the effect and the color indicates positive (red) or negative (blue) effect.

Fig. S13 Mean genome sizes (2C; Gbp) across the global latitudinal gradient illustrating TDWG 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; growth form percentages in TDWG regions.

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

Table S2 Results of quantile regression of 2C genome size on range size (EOO).

Table S3 Results of phylogenetic quantile regression of 2C genome size on range size (EOO).

Table S4 Results of quantile regression of genome size on range size (TDWGs).

Table S5 Results of phylogenetic quantile regression of genome size on range size (TDWGs).

Table S6 Results of OLS and PGLS regressions of mean chromosome size on range size (EOO).

Table S7 Results of quantile regression of mean chromosome size on range size (EOO).

Table S8 Results of phylogenetic quantile regression of mean chromosome size on range size (EOO).

Table S9 Results of OLS and PGLS regressions of mean chromosome size on range size (TDWGs).

Table S10 Results of quantile regression of mean chromosome size on range size (TDWGs).

Table S11 Results of phylogenetic quantile regression of mean chromosome size on range size (TDWGs).

Table S12 Additional regressions of 2C genome size on other biologically relevant variables.

Table S13 Results of regressions of 2C genome size on percentage of growth forms in TDWGs.

Table S14 Results of regressions of 2C genome size on additive effects of GST and percentage of growth forms in TDWGs.

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	3.746	<3.708, 3.784>	191.71	<2E-16	0.012
$\log_{10}$(Range size)	-0.039	<-0.046, -0.033>	-12.24	<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	3.583	<3.5782, 3.5887>	<2E-16	0.916	0.002
$\log_{10}$(Range size)	-0.007	<-0.008, -0.0066>	1.31E-06		

Table 1: Results of ordinary least squares (OLS) and phylogenetic generalised least squares (PGLS) regression of 2C genome size on range size. bi - regression estimates of model terms; 95%CI - lower and upper 95% confidence intervals of the regression estimates; R2adj - R squared adjusted indicating explained variance. The OLS analysis was perfromed with 12,137 species. The PGLS analysis was perfromed with 12,123 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

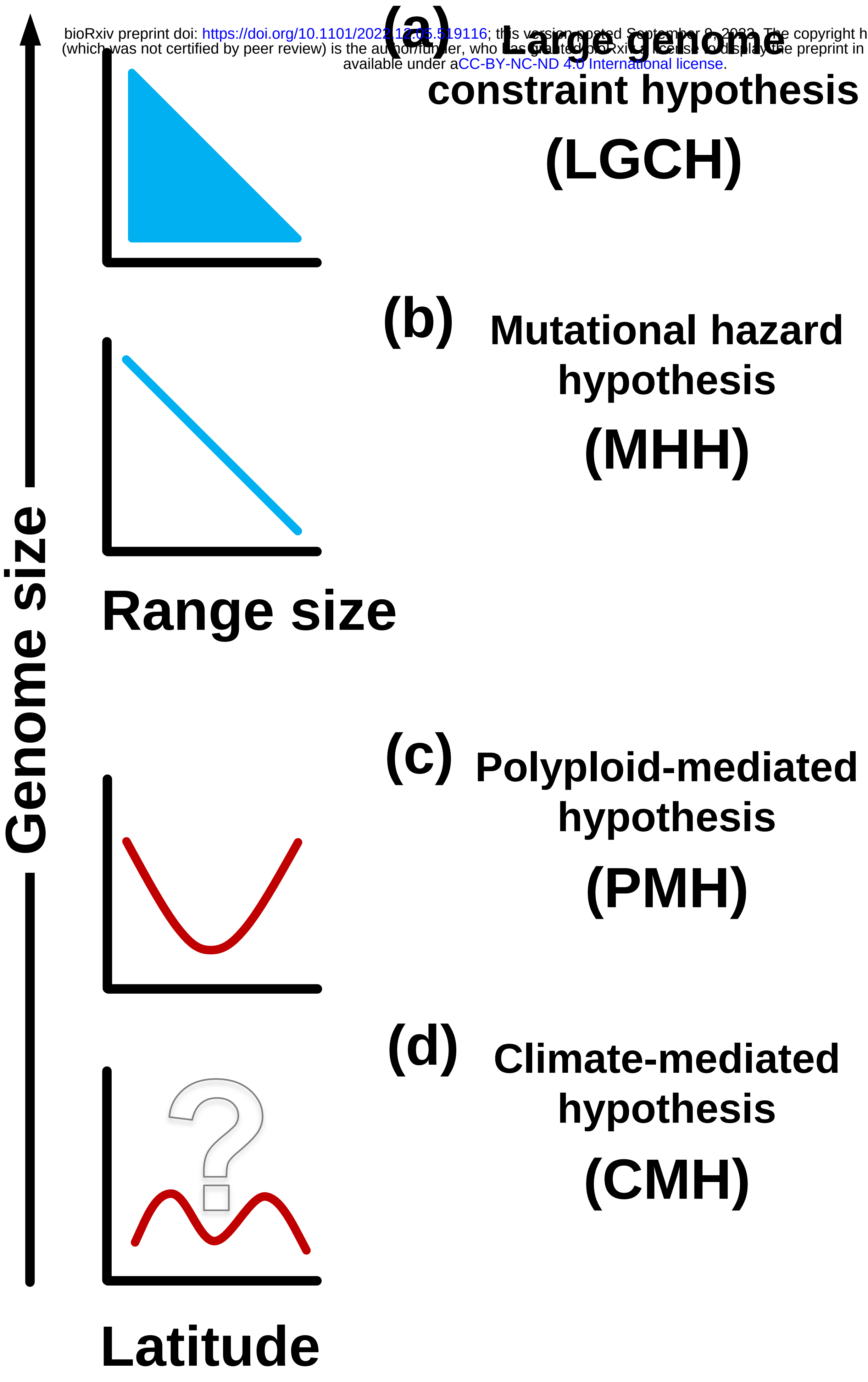
Model term	b_i	95%CI	t	P	R^2_{adj}
Intercept	3.372	<3.338, 3.407>	191.10	<2E-16	0.0177
Polyploid proportion	0.116	<0.033, 0.199>	2.75	0.006	

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

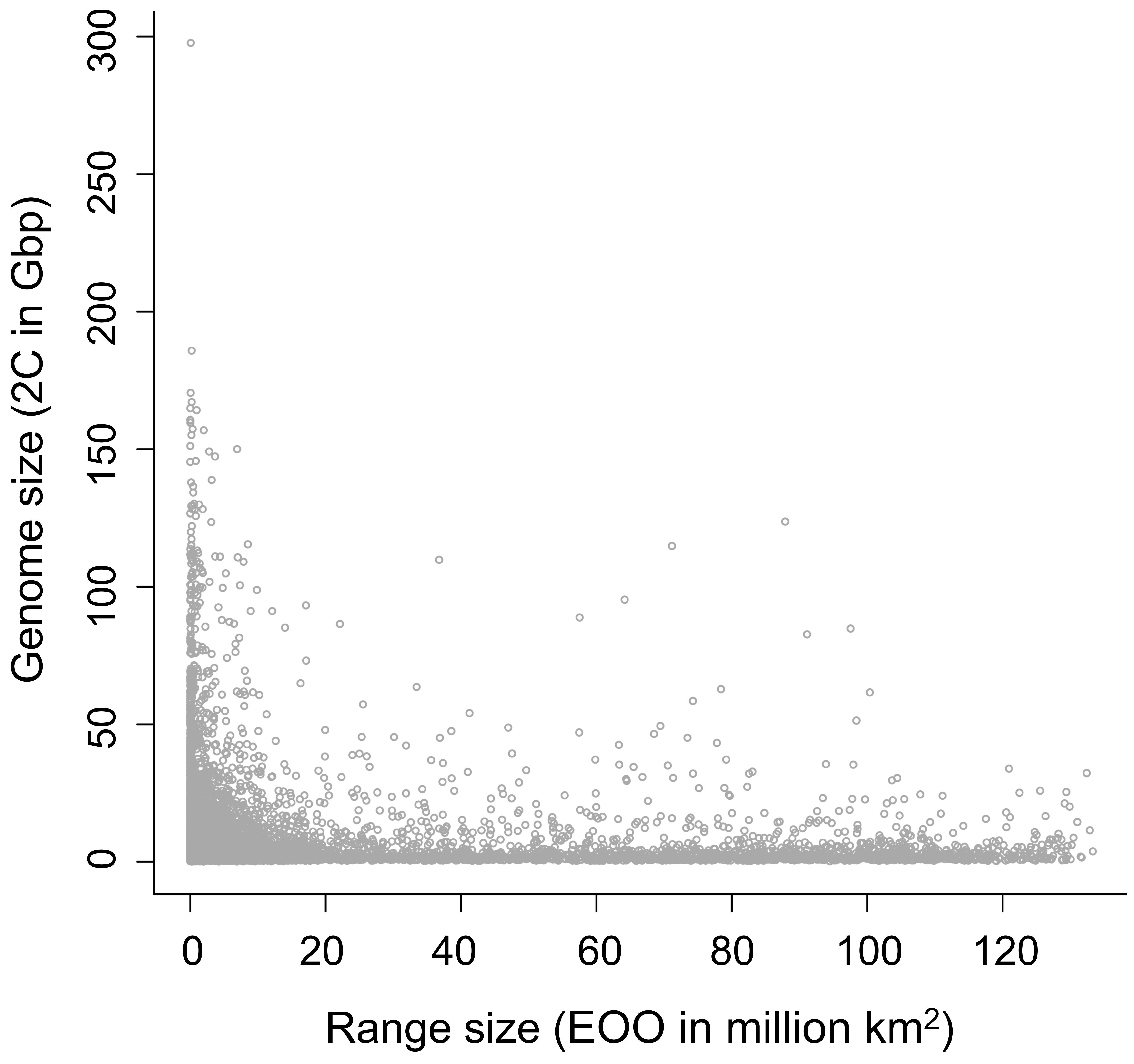
Model term	b_i	95%CI	t	P	R^2_{adj}
Intercept	3.444	<3.405, 3.483>	172.53	<2E-16	0.4075
GST	0.005	<-4.33E-05, -0.0099>	1.95	0.052	
GST^2	0.000	<-0.0005, -0.0003>	-5.34	<2E-07	

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

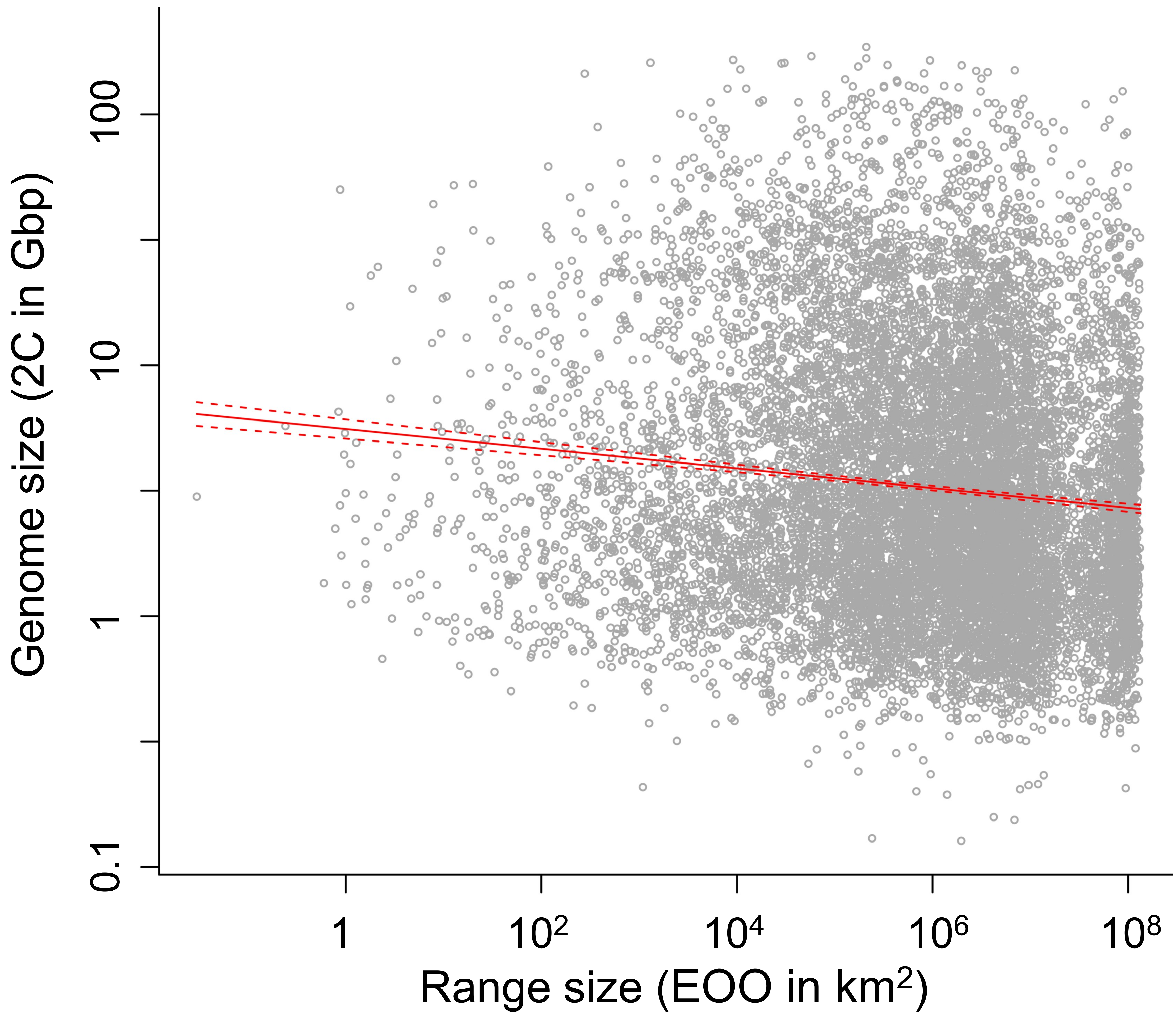
base R, but the parameter "raw" was set to "TRUE" to obtain parameter estimates corresponding to response variables.


(a)
Large genome constraint hypothesis
(LGCH)
(b)
Mutational hazard hypothesis
(MHH)
Range size
Genome size
(c)
Polyploid-mediated hypothesis
(PMH)
(d)
Climate-mediated hypothesis
(CMH)
Latitude

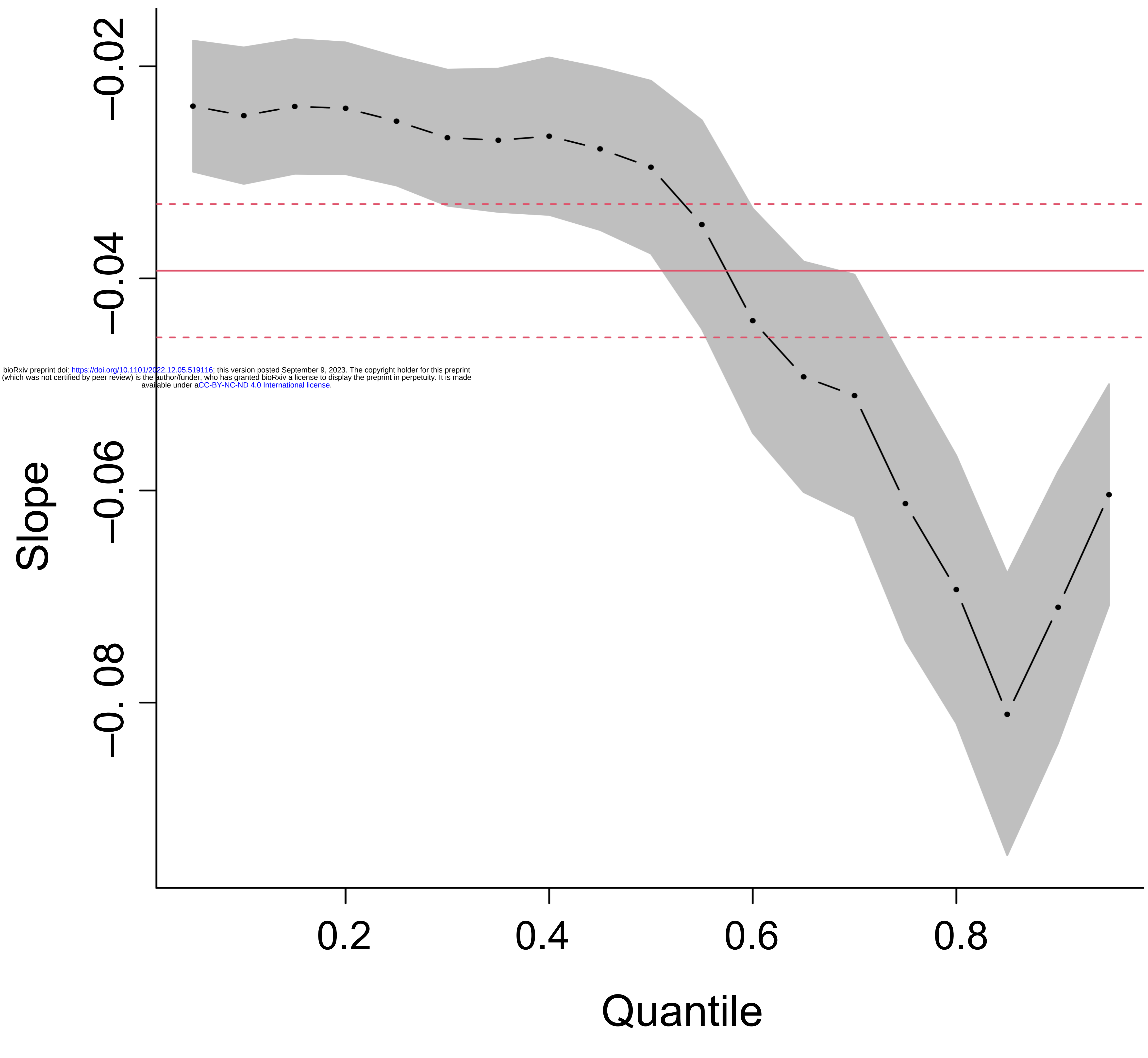
(a) Range size, EOO ~ Genome size per species



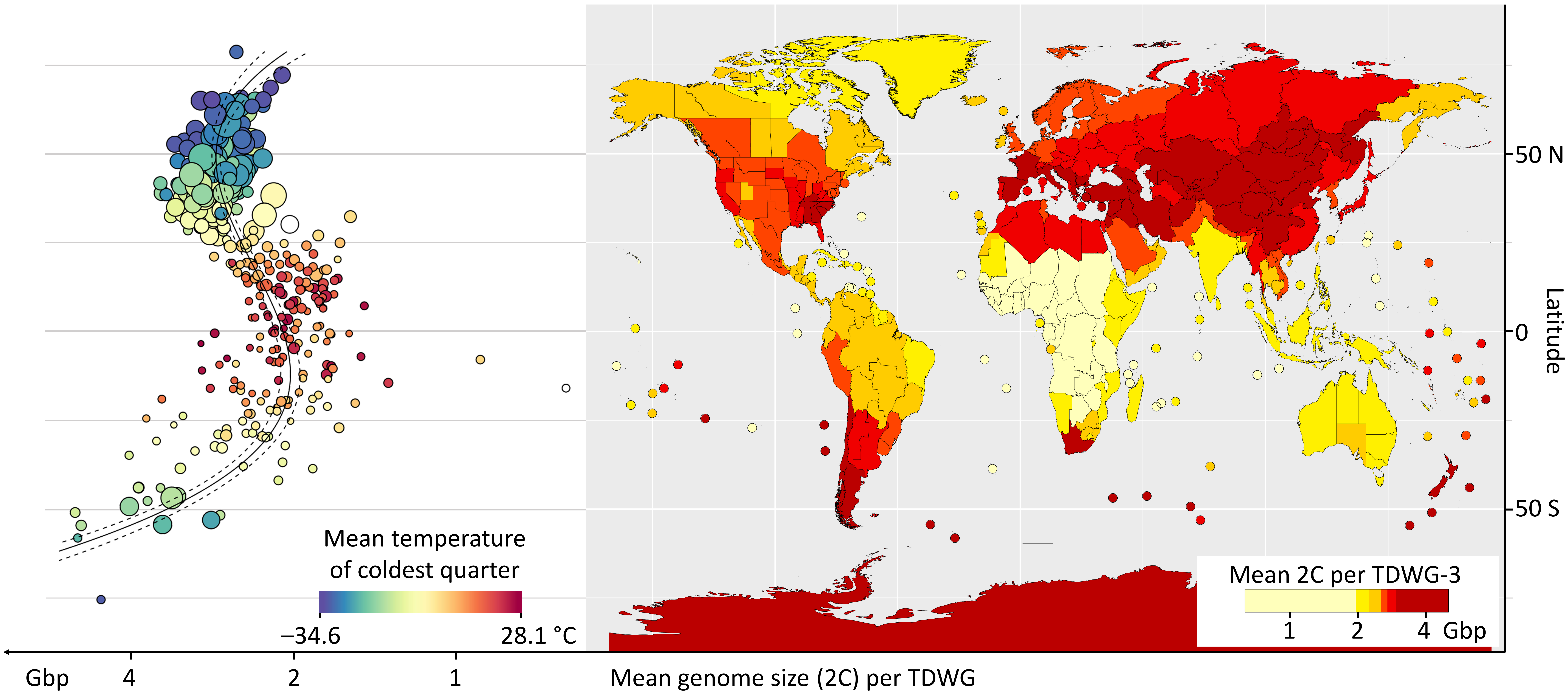
(b) OLS: Range size, EOO → Genome size per species



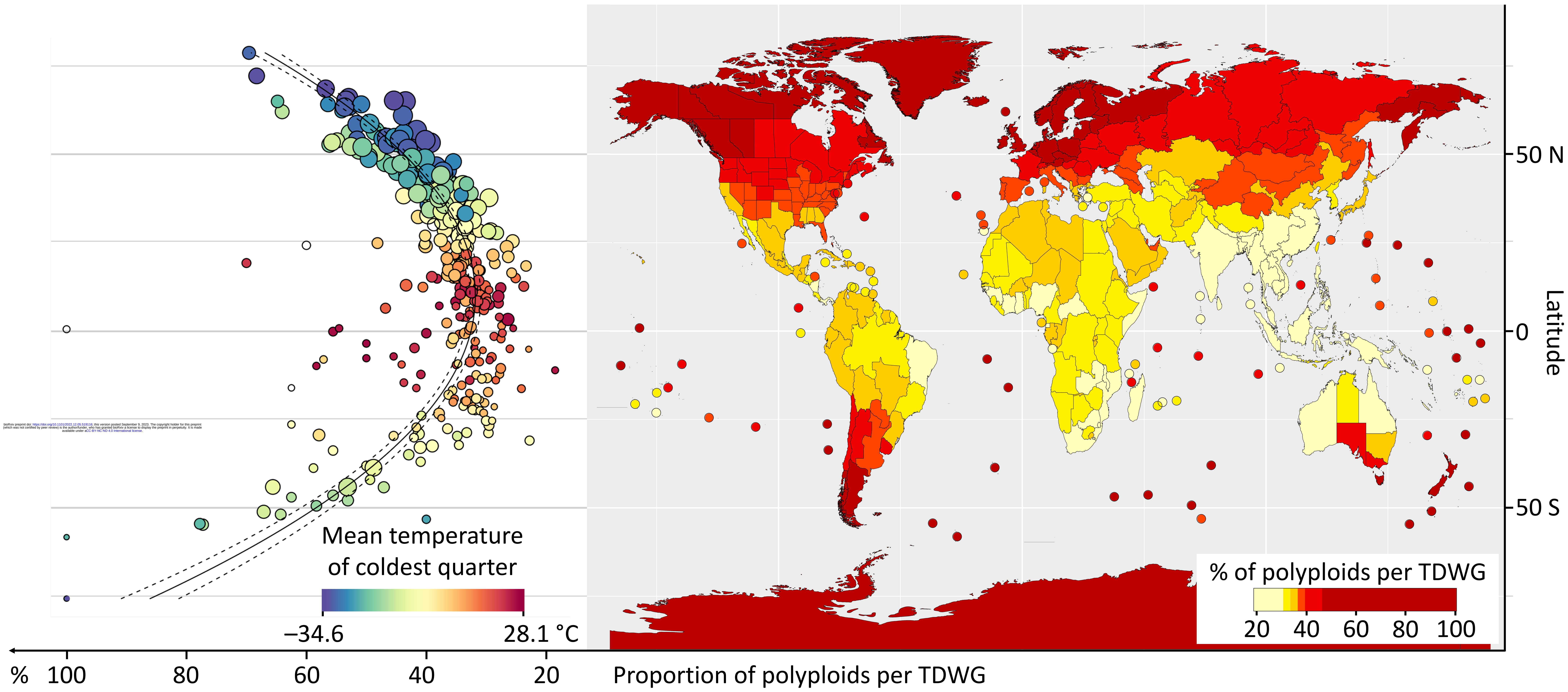
(c) Quantile regression per species



(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

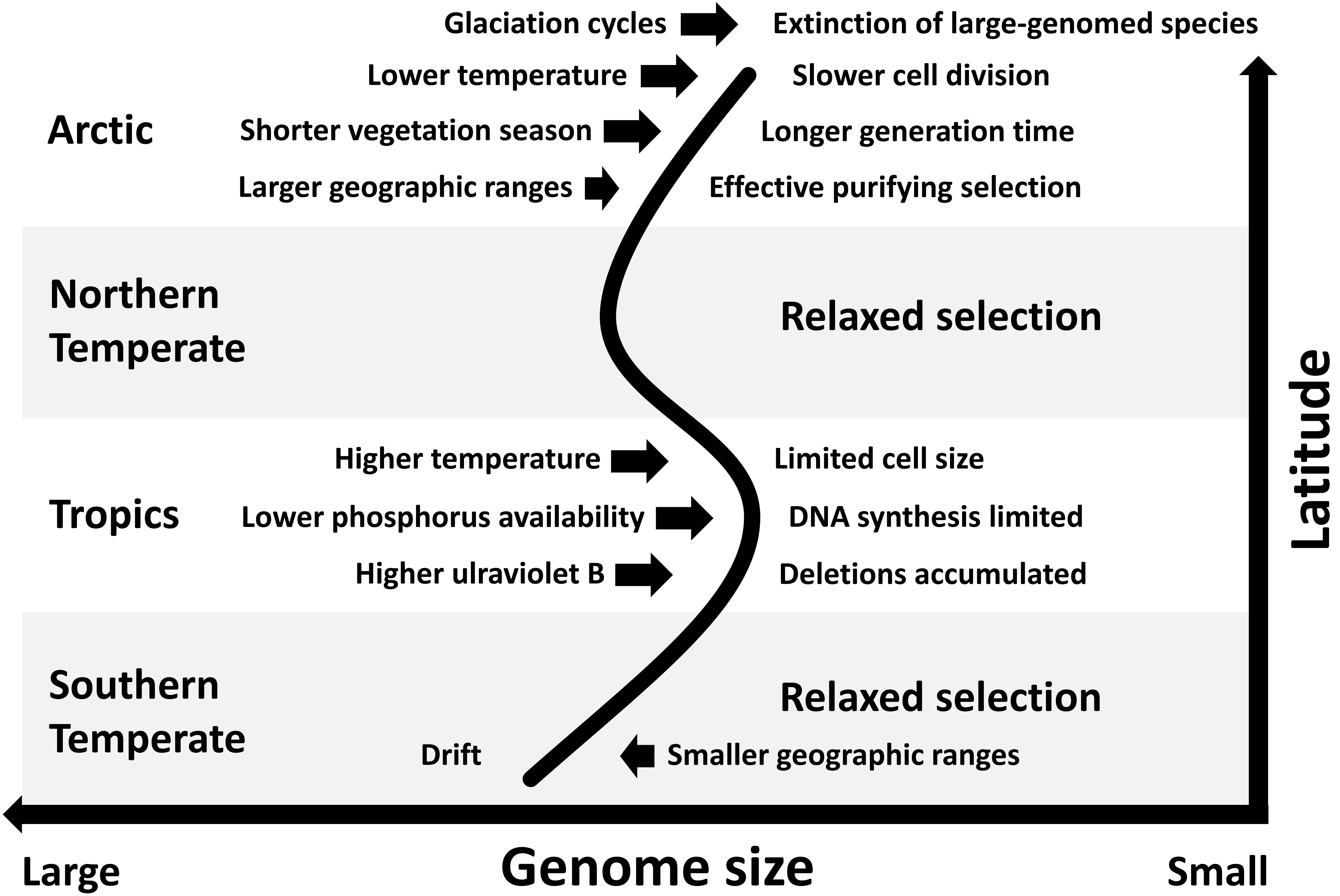
Latitudinal trends in mean genome sizes across TDWGs and life forms

(b) Annuals

(c) Geophytes

(d) Non-geophyte perennial herbs

(e) Woody plants


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 selection
Higher temperature
Limited cell size
Tropics
Lower phosphorus availability
DNA synthesis limited
Higher ulraviolet B
Deletions accumulated
Southern Temperate
Relaxed selection
Drift
Smaller geographic ranges
Latitude
Large
Genome size
Small