Zhang Haiyang (Orcid ID: 0000-0001-7951-0502)
Lü Xiaotao (Orcid ID: 0000-0001-5571-1895)
Xing Dingliang (Orcid ID: 0000-0001-9133-6674)
Xu Zhuwen (Orcid ID: 0000-0002-8898-6577)

Journal: Ecology
Manuscript type: Article

β-diversity in temperate grasslands is driven by stronger environmental filtering of plant species with large genomes

Authors: Hai-Yang Zhang[a, b, c*], Xiao-Tao Lü[b1], Cun-Zheng Wei[d*], Jeff R. Powell[c], Xiao-Bo Wang[b,e], Dingliang Xing[f], Zhu-Wen Xu[g], Huan-Long Li[d], Xing-Guo Han[b, d, h]

Affiliations:
[a]College of Life Sciences, Hebei University, Baoding, China;
[b]Erguna Forest-Steppe Ecotone Research Station, CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China;
[c]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales, Australia;
[d]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China;
[e]Center for Grassland Microbiome, State Key Laboratory of Grassland Agroecosystems, and College of Pastoral, Agriculture Science and Technology, Lanzhou University, Lanzhou, 730020, China
[f]Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China;
[g]Department of Ecology, School of Ecology and Environment, Inner Mongolia University, No. 235 West College Road, 010021, Hohhot, China;
[h]University of Chinese Academy of Sciences, Beijing, China

[*]These authors contributed equally to this work.

Corresponding author: Hai-Yang Zhang, E-mail: haiyang.zhang@westernsydney.edu.au

Handling Editor; James T. Cronin

Open Research: Data (Zhang, 2022) are available in Figshare at https://doi.org/10.6084/m9.figshare.20452569.

Abstract

Elucidating mechanisms underlying community assembly and biodiversity patterns is central to ecology and evolution. Genome size (GS) has long been hypothesized to potentially affect species' capacity to tolerate environmental stress and might therefore help drive community assembly. However, its role in driving β-diversity (i.e., spatial variability in species composition) remains unclear. We measured GS for 161 plant species and community composition across 52 sites spanning a 3200-km transect in the temperate grasslands of China. By correlating the turnover of species composition with environmental dissimilarity, we found that resource filtering (i.e., environmental dissimilarity that includes precipitation, soil nitrogen and phosphorus concentrations) affected β-diversity patterns of large-GS species more than small-GS species. By contrast, geographical distance explained more variation of β-diversity for small-GS than for large-GS species. In a 10-year experiment manipulating levels of water, nitrogen, and phosphorus, adding resources increased plant biomass in species with large-GS, suggesting that large-GS species are more sensitive to the changes of resource availability. These findings highlight the role of GS in driving community assembly and predicting species responses to global change.

Keywords: Biodiversity conservation, community assembly, dispersal, genome size diversity, nutrient addition, species distribution.

Introduction

Disentangling the drivers of β-diversity (the site-to-site variability in species composition) provides insights into the processes that govern community assembly (Chase 2010, Mori et al. 2018). β-diversity can arise from community assembly processes involving deterministic selection, when environmental heterogeneity creates different niches that shape the occurrences of species in a community, and stochastic aspects related to dispersal limitations and ecological drift (Vellend et al. 2014, Mori et al. 2018). Interspecific variations in plant traits determine the capacity for individuals to grow, reproduce, and disperse within and among habitats, and therefore play important roles in determining the relative importance of deterministic selection (McGill et al. 2006).

A fundamental character that significantly varies across angiosperms [2400-fold; Pellicer and Leitch (2020)] and correlates with diverse phenotypic characters at the cellular and organismal level (Herben et al. 2012, Pellicer et al. 2018) is genome size (GS, i.e. nuclear DNA content). GS has received relatively little attention in the context of its role in community assembly but could have functional consequences for species' environmental tolerance, dispersal capacity, and interactions with other species (Knight and Ackerly 2002, Herben et al. 2012). The potential impact of plant GS on community assembly processes is starting to be recognized (Pellicer et al. 2018), but to date, it remains unclear whether and to what extent the environmental drivers of species composition will be different for large and small GS communities over a large scale.

A key determinant of β-diversity is environmental filtering. The strength of the relationship between local environmental conditions and species' environmental requirements affects the establishment and persistence of species (van Breugel et al. 2019). Environmental

filtering is hypothesized to differ for large- vs. small-GS species (Faizullah et al. 2021). First, small-GS species grow faster due to a short cell cycle duration and are subject to fewer material costs for packing DNA, allowing them to achieve optimal growth across a wider range of environments. Second, according to the 'large genome constraint hypothesis', the optimal growth for large-GS species is only achievable under conditions of stress-free or high resource availability (Knight et al. 2005). It has been hypothesized that there will be selection for species with small genomes in nutrient-depleted soils as a way to reduce the biochemical cost of synthesising DNA, which is rich in nitrogen (N) and phosphorus (P) (Leitch and Leitch 2008). Indeed, there is evidence that large-GS species became more dominant under conditions with higher nutrient availability (Šmarda et al. 2013, Guignard et al. 2016). Recently, Peng et al. (2022) manipulated nutrient availability in the Inner Mongolia grassland and showed that aboveground net primary productivity increased predominantly from large-GS species with the addition of N, and N plus P. Thus, we expect that environmental filtering would have stronger effects on large-GS species than that on small-GS ones.

Finally, empirical evidence largely supports the positive correlation between generation time and genome size (Grotkopp et al. 2004); because of this, plants with larger genomes are disproportionally more frequent among perennial herbs (Stebbins 1971). Therefore, life forms should be considered given that genome size in a community can depend on the relative proportions of annual and perennial species. To assess the roles that genome size plays in plant community assembly, we used data from 520 plant communities in 52 sites (10 communities per site) along a 3200-km transect in the temperate grasslands of northern China (Figure 1). We measured plant GS [the amount of DNA in a gamete nucleus or 1C-value, representing the DNA content of the whole complement of chromosomes for the organism, irrespective of the degree of

generative polyploidy; C-values have been used as a reference value for GS, for details see Greilhuber et al. (2005)] for 161 herbaceous species occurring along the transect (Appendix S1: Figure S1). Generalized dissimilarity models [GDMs (Fitzpatrick et al. 2013)] were used to quantify the effects of GS, environmental variation, and geographical distance on β-diversity along the gradient. We hypothesized that environmental filtering would play a more important role in driving β-diversity for large-GS than for small-GS species. In addition, to confirm the findings of our observational transect study, we analysed data from a 10-yr field experiment manipulating resource availability to examine whether large-GS species would be favored over small-GS species following water, nitrogen and phosphorus addition.

Materials and Methods

Study sites

The transect study was conducted across a 3200-km scale of northern China's grasslands, extending from the Xinjiang Uygur Autonomous Region in the west to eastern Inner Mongolia (42.89° N to 49.19° N, 83.45° E to 120.36° E; Figure 1). There are four vegetation types along this transect, including alpine-, desert-, typical- and meadow-steppe from west to east (Appendix S1: Table S1). Plant species richness decreased from the eastern to the western end of the transect. Dominant soil types are classified as aeolian chestnut soil in the east to brown calcic soil, grey desert soil, and sandy soil to the west.

Plant community and soil sampling

We conducted transect sampling during July and August 2012 by sampling 52 sites. To ensure that plant samples were collected during the same period of phenology at each site, we sampled sequentially from the west to the east along this transect since this coincides roughly

with a decreasing temperature trend (i.e., delayed growing peak period). Study sites were investigated from west to east along the entire transect with an interval of 50~100 km. Sampling sites were generally far from cities, under natural conditions, with little human disturbance, and represented the local natural vegetation (Wang et al. 2014, Wang et al. 2017). At each site, two 50 m × 50 m large plots were established, and five 1 m × 1 m quadrats were selected within each large plot (each corner and the center of the plot). For each quadrat, we clipped the aboveground tissues of living plants, sorted into species, and stored in paper bags. Community data from 10 quadrats were pooled together to represent the local species pool for each site. All these analyses on plant community were based on species presence/absence data. Soil samples from each quadrat were collected with five soil cores (2.5 cm diameter × 10 cm depth) from the upper 10 cm layer. Five soil cores were collected from four corners 10 cm away from the edge plus one from the center of the quadrat. Soils from the five soil cores were combined for each quadrat and sieved (2.0 mm mesh) to remove roots and rocks, homogenized by hand, and preserved for subsequent chemical analysis.

Environmental variables

At each site, spatial geographical coordinates were recorded using a handheld GPS (eTrex Venture, Garmin, Olathe, Kansas, USA). We targeted how water and nutrient availability could affect community composition in this transect study. Mean annual precipitation (MAP) that indicates long-term average water availability of each sampling site was obtained from WorldClim 2.0 (1-km spatial resolution) (Fick and Hijmans 2017). Total N concentration of soil samples was determined using wet oxidation and a modified Kjeldahl procedure, and total P concentration was measured by colorimetric analysis with ammonium molybdate and persulfate oxidation (Murphy and Riley 1962).

Plant genome size measurements

Plant species were first identified by a group of plant taxonomists, then all the species names were further standardized into the accepted names according to The Plant List (version 1.1; www.theplantlist.org). We recorded 286 herbaceous species in the transect; of these we were able to obtain measurements of GS from 169 species during subsequent visits to representative study sites belonging to the Chinese Grassland Long-term Research Stations (at least one site for each grassland type: alpine grassland, desert steppe, typical steppe, and meadow steppe). These stations represented the typical vegetation and species pool for each grassland type and were convenient to re-visit. Samples were collected to measure genome sizes during the growing seasons of 2017-2019 (from July to September). We sampled plant species that occurred at each grassland type of the transect, focusing primarily on (but not limited to) the more common ones (details for the common species and species richness of each study site are supplied in Zhang, 2022). Genome size measurements were conducted according to Doležel et al. (2007) and details are supplied in Appendix S1: Section S1. Eventually, we had genome size information for 161 common species (Appendix S1: Table S2), and these species contributed at least 80% biomass and richness in each of the 52 sites. All subsequent analyses were based on these 161 species from the 52 sites along the transect (for the longitude distribution of each species, see Appendix S1: Figure S2).

Ideally, one should measure the genomes sizes of individual species in every site along the transect to check the intra-specific variation (Šmarda and Bureš 2010, Meyerson et al. 2016). However, it is a challenge to do so over this 3200 km transect. In our case, we first assumed the genomes sizes were relatively stable at the species level within the same grassland type. Then, we screened the intra-specific variation of GS for species that occurred in more than one

grassland type across the transect to account for the potential effects of intra-specific variation or ploidy on our analysis. To do so, 25 species were selected, and their GS were measured across multiple sites (2-5 sites, depending on their occurrence frequency) to check the GS intra-specific variation (Appendix S1: Table S3). Among the 25 species, no evidence of substantial intra-specific variation in GS or ploidies was found, except for, *Agropyron cristatum*, *Allium mongolicum*, *Artemisia frigida*, *Potentilla anserina*, and *Potentilla tanacetifolia*. For those five species, GS showed substantial difference at the meadow steppe (in the case of *P. anserina*, alpine) but was quite stable at other grassland types. For each of these five species, we treated them as two different pseudo-species and define their GS based on which grassland types they mainly occurred. Overall, the plant GS (1C values) varied 137-fold from the smallest *Cardamine hirsuta* (0.23 pg/1C) to the largest *Allium ramosum* (31.50 pg/1C), with a median and mean of 1.61 and 3.39 pg/1C, respectively (Appendix S1: Figure S1; Table S2). By comparison, for the 25 plant species that had multiple GS measures, GS ranged from 0.35 pg/1C in *Carex korshinskyi* to 31.50 pg/1C in *Allium ramosum*.

Generalized Dissimilarity Modelling (GDM)

We used the GDM approach to analyse β-diversity patterns along environmental gradients, which is a matrix regression technique for modelling turnover in species composition between sites as a function of the spatial and/or environmental distance between them (Ferrier et al. 2007, Fitzpatrick et al. 2013, Mokany et al. 2022). GDM formulates the relationship between β-diversity and environmental and/or spatial distance using generalized linear modelling with a link function of the form:

$$\beta_{ij} = 1 - e^{-\eta_{ij}} \quad (1)$$

where β_{ij} is the community dissimilarity between sites *i* and *j*, and η_{ij} is the environmental distance between those sites (Mokany et al. 2022). GDM uses I-spline basis functions to transform each of the predictor variables (Ferrier et al. 2007), so that summed absolute difference in the transformed predictor values provides a predicted ecological distance (η). In our case, environmental distances were calculated based on the dissimilarity of MAP, soil N and P between sites. The advantage of GDMs is that they allow nonlinear relationships between dissimilarity and distance (Fitzpatrick et al. 2013). In addition, another major advantage of GDM over other modeling methodologies is that it can explicitly and simultaneously take into account the influence of both geographic distance (i.e. spatial autocorrelation) and environmental variables on explaining biological variation. GDM provides information on the relative importance of predictor variables by means of response curves (I-spline). The I-spline associated with each variable describes the relationship between beta diversity and that gradient. The total amount of compositional turnover associated with each variable, holding all other variables constant, can be inferred from the maximum height of the I-spline associated with it (Fitzpatrick et al. 2013, Mokany et al. 2022). Thus, the influence of geographic distance relative to other variables in explaining variation in β-diversity can be assessed even when localities are spatially autocorrelated.

We plotted the partial effect of each predictor against the level of a given predictor to visualize the results of each GDM (holding all other predictors constant). The shape of the line shows how β-diversity varies along each environmental or spatial gradient, i.e. how the effect of a given predictor on β-diversity varies at a given level of that predictor. Furthermore, we also determined the proportion of deviance uniquely attributable to environmental resources or distance, by comparing the deviance explained by a GDM containing all of the variables and a

GDM with all variables except environment or distance, respectively. The unique deviance explained by resource or distance was calculated as the difference in deviance explained by these models. We then converted this to a percentage by dividing the deviance explained by the full GDM. These percentages can indicate the relative importance of geographic distance among sites (linked to dispersal limitation processes) and environmental difference (linked to niche differentiation processes) in determining β-diversity.

GDMs were fitted to the β-diversity for turnover component ($\beta_{turnover}$) and total (β_{total}), separately, using the gdm function in the gdm R-package (Manion et al. 2017). The overall regional species composition dissimilarity (β_{total}) was estimated using the Sørensen dissimilarity index (Baselga 2010). β_{total} was then partitioned to quantify the portion of dissimilarity originating from 'pure' species turnover (Simpson's dissimilarity; $\beta_{turnover}$) or from differences in species richness ($\beta_{richness}$) where $\beta_{total} = \beta_{turnover} + \beta_{richness}$ (Baselga 2010). In our study, species turnover among grasslands was more important than pure richness differences of assemblages ($\beta_{total} = 0.86$, $\beta_{turnover} = 0.77$, $\beta_{richness} = 0.09$; Appendix S1: Figure S3); in other words, differences in prevalence among species in our study were more important than differences in species richness among sites in driving variance in species composition. Thus, we mainly focus on how $\beta_{turnover}$ changed along the environmental gradient. However, we also calculated the β_{total} to examine changes of the total β-diversity via the betapart package in R (Baselga and Leprieur 2015). The results were generally similar for $\beta_{turnover}$ and β_{total} (Figure 2-3 vs. Appendix S1: Figure S4-5) and therefore we only reported the results for $\beta_{turnover}$ in the man text for simplicity.

For the GDM analysis, we first separated plant species into large and small GS groups based on the median GS value of our studied species (1.61 pg/1C value) and run the model analysis separately. Given that one cannot include a continuous trait or character in GDM

analysis, here we use the grouping strategies (i.e., separating plants into large or small GS groups) to explore whether GS mediates the processes of community assembly. We then repeated such analysis by using different GS separating thresholds, i.e. sequential threshold analysis. Plant communities were separated into the large- and small-GS subsets also by the mean (3.39 pg/1C) GS of the studied species or by the median (2.50 pg/1C) or mean (5.90 pg/1C) GS of the global terrestrial plants in the Kew C-value database (https://cvalues.science.kew.org/). In addition, as GS has a fundamental relationship with generation time and plant growth forms, we also ran the GDM analysis separately for annual and perennial species. Because our grasslands were dominated by perennial species (136 perennial vs. 25 annual species), we also repeated the sequential threshold GDM analysis for perennial species only.

Quantile regression

To directly explore whether GS is associated with environmental conditions, we conducted quantile regression (QR) with the 'quantreg' package (Koenker 2015). Cade and Noon (2003) provide a detailed discussion of quantile regression methods. Briefly, QR is a statistical method for modelling linear relationships across a series of quantiles of the response variable and it does not assume equal variances across covariate values nor does it make assumptions about the distribution of errors. Therefore, it is always useful for models with unevenly distributed observations, and where more than one slope can describe how the response and explanatory variables are related (Cade and Noon 2003). In our case, we estimated the quantile-specific linear coefficient (slope) starting from quantile 5^{th} to 95^{th} (quantiles increase every 10 percent) of the plant GS distribution using quantile regression and bootstrap resampling

to estimate the 95% confidence intervals of the coefficients. QR was applied to the correlation between GS and MAP, soil N or soil P, respectively.

Water, nitrogen and phosphorus addition experiment

To further test whether large-GS species would be favored over small-GS species by increasing resource availability, we obtained data from a long-term field resources manipulation experiment that was conducted in the typical steppe of Duolun, Xilingol in northern China (42.0°N, 116. 3°E). We chose this site because the species in this local site (n = 43) highly overlap with the species in our transect study, with 29 species of this local site occurring also in our transect study. The site has a MAP of 360 mm and MAT of 2.1°C and is close to the eastern side of the transect described above. The study design is described by Xu et al. (2012). As we aimed to assess the effects of resource limitation generally and not of the specific resources (for which limitation may differ by site), we selected only two treatments from this experimental setup: no (control) vs. combined multiple resource addition. The control plots only received ambient precipitation while the resource addition plot received additional water and nutrients on top of the ambient precipitation from 2007 to 2016 as described by (Xu et al. 2012). Water was added weekly (15 mm via sprinkler) during the growing season, which added up to 180 mm yearly. Nutrients were added in May and July as $CO(NH_2)_2$ and $Ca(H_2PO_4)_2$, with rates of 10 g N m^{-2} yr^{-1} and 10 g P m^{-2} yr^{-1}. Each treatment was applied at the plot level (8 m × 8 m) and had 7 replicates. Aboveground biomass was harvested at the end of each growing season from a 0.3 m × 0.3 m quadrat randomly chosen from within each plot. Biomass was sorted by species and was oven-dried at 65 °C for 48h. From May to July 2017, GS was measured for each species following the same procedure as in the transect study. We measured the GS of the same species from different treatment plots. At this local scale, resource addition treatment affected very little

the genome size of a given plant species. Thus, GS values were averaged at the species level. Note that from the plant community record, 50 species occurred during the experimental period and 43 of them were sampled and measured for GS in 2017. The measured species contributed more than 97% plant aboveground biomass at the community level for all studied plots. Community weighted mean (CWM) GS was calculated based on aboveground biomass of the plant species that had GS data in each plot. We used the log-transformed GS to calculate CWM due to GS being a strongly right-skewed variable. We first grouped the plant species into large and small GS groups by their median value (pg/1C value) to check how the aboveground productivity of plants with large or small GS responds to resource addition. Phylogenetic generalized linear mixed models (PGLMM) with Bayesian estimation were used to test for any effect of GS and resource addition on aboveground biomass at the species level. GS was included as a continuous variable in the PGLMM statistical analysis. Although we only had five annual species in our studied community, we still considered potential influences of growth forms in data analysis. The analyses were performed by fitting Markov Chain Monte Carlo generalized mixed models from R package MCMCglmm (Hadfield 2010) with plant phylogeny and plot as random factors. All analyses were conducted in R software 3.6.2 (R Core Team 2020).

Results

β-diversity patterns along the transect gradient

We found that MAP and soil N and P concentrations increased from western to eastern of our transect, except for sites from the alpine steppe that had relatively high MAP, soil N and P concentrations (Figure 1b-d). We found β-diversity increased with increasing environmental

dissimilarity (Figure 2a). After separating the species pool based on the median GS of our study, drivers of β-diversity in small and large GS groups were strikingly different, with β-diversity of small GS groups mainly driven by distance and MAP while that of large GS groups was driven by distance, MAP, and soil P conditions (Figure 2b, c). MAP explained more variation of β-diversity for small (8.63%) than large GS (7.77%) groups while soil N and P explained more for large (10.06%) than small (0.85%) groups (Figure 2d-f).

We observed qualitatively similar results when using different separating thresholds (Fig. 3). When summing the variation explained by resources, including MAP, soil N and P, we found that resources explained more variation of β-diversity for species with large genomes than that for small genomes (Figure 3a). By contrast, distance explained more variation of β-diversity for species with small genomes than that for large genomes when using threshold 2.5 pg/1C (Figure 2b), and the differences became larger as the selected threshold increased (Figure 3b). More interestingly, the resources explained more while distance explained less variation of β-diversity for species with large genomes, leading to a significantly increased resources/distance ratio (Figure 3c) as the threshold increases.

Perennial species had significantly larger GS (3.76 pg/1C) than annual (1.50 pg/1C) species (Appendix S1: Figure S6a). The drivers of β-diversity in annual and perennial species were different, with β-diversity of annual species mainly driven by distance and soil N while that of perennial species was driven by distance, MAP, and soil P concentrations (Appendix S1: Figure S6b-f). Even for perennial species only, we found that resources explained more variation of β-diversity for species with large GS than that with small GS. By contrast, distance explained more variation of β-diversity for species with small GS than that with large GS, regardless of which threshold we chose.

Quantile regression showed that plant GS did not correlate with MAP for most quantiles except for the negative correlations at the very upper quantiles. By contrast, plant GS positively correlated with soil N and P when quantiles ranging from 20% to 70% and then shifted to negative correlations at the very upper quantiles (Figure 4).

Biomass response to the water and nutrient addition over 10 years

We found that the CWM GS increased significantly after resource addition (Figure 5a; Appendix S1: Table S4). The enhancement of plant biomass was higher for large-GS species (3.9 times) than for small-GS species (2.2 times; Figure 5bc). MCMCglmm results showed that although GS had a non-significant effect on plant biomass ($P = 0.096$), its interaction with resource addition was significant ($P < 0.001$; Table 1).

Discussion

Pairwise species turnover among communities ($\beta_{turnover}$) showed a positive correlation with predicted environmental distances, calculated by combining MAP, soil N and P, in line with previous findings (Robroek et al. 2017, Mori et al. 2018). For each specific environmental predictor, soil N and soil P but not MAP explained more variation in $\beta_{turnover}$ for the large-GS species (Figure 2). These results indicate that plant GS can influence the relative importance of resource heterogeneity and geographic distance (which can indicate, in part, dispersal limitation) in driving plant β-diversity. Because the environmental filtering and dispersal processes have been broadly recognized as two main processes in driving species compositions (Mori et al. 2018), we conclude that genome size plays a critical role in affecting community assembly and species diversity over a large spatial scale.

Plant genome size can have a fundamental connection with growth forms such as annuals or perennials (Herben et al. 2012, Greilhuber and Leitch 2013). In our study, perennial plants had significantly larger genomes than annual plants (Figure S6), which was in agreement with previous study (Herben et al. 2012). One would expect that fundamental differences between annual and perennial growth forms may explain the distinct community assembly processes between the large and small GS groups (Figure S6). However, we provided two pieces of evidence to reject this explanation. First, we found community assembly was different for perennial and annual species, but the underlying patterns in perennial/annual comparisons were different from the large/small GS comparisons. For the perennial/annual comparisons, resource filtering contributed equally to the variation of β-diversity for both growth forms while distance contributed more to β-diversity in perennial than in annual plants. Second, we found that even when only perennial species were used to run the sequential threshold GDM analysis, results were consistent with the one revealed by using all species together, i.e., geographical distance explained more variation of β-diversity for small than large GS groups while soil N and P explained more for species with large GS than small GS (Figure 2d-f).

By choosing different separating thresholds from a small to a large GS value when conducting the large-small GS comparisons, we showed that resources explained more variation and geographic distance explained less variation as the threshold increased, indicating that the species with the largest genomes were most strongly impacted by resource filtering (Figure 3). Previous studies used the quantile regression approach to directly test the linkage between plant GS and their environmental condition (Knight and Ackerly 2002, Faizullah et al. 2021). For example, Knight and Ackerly (2002) found that species with small GS predominated in all environments while species with large GS occurred at intermediate July maximum temperatures

and increased in frequency with increasing annual precipitation. In our study, we showed that plant GS correlated more with soil nutrients than precipitation across the gradient of quantiles. In addition, we revealed a positive correlation between GS and soil nutrients in most quantiles, except for the upper quantiles where negative correlations emerged (Figure 4). These results suggest that plants with large genomes might not be always favored by increased resource availability, but they do show higher sensitivity to the changes in resource availability. Regardless of whether plants with large GS can be selected with or against along resource gradient, results from Knight and Ackerly (2002) and ours both suggest that relationships between GS and environmental filters are stronger for species with large GS.

Our local resource addition experiment provided additional insights into how resource filtering can affect plant community and their community-level GS. Over the 10-yr course of our experiment, resource addition generally increased plant biomass production and the extent of such enhancement was greater in plants with large GS than with small GS (Figure 5), supported by the significant interactive effect between GS and resource addition ($P < 0.001$). This was in line with the large genome constraint hypothesis and suggest that plants with large GS might have a high resource demand (Knight et al. 2005). Previous studies found that the large GS species became more dominant after nutrient addition for 72 (Šmarda et al. 2013) and 160 yr (Guignard et al. 2016) respectively at the time of data analysis, supporting the general pattern across 10 years in our experiment. More importantly, our results were in line with a recent study conducted also at the Inner Mongolia grassland (Peng et al. 2022), where the interactions between GS and nutrient addition can influence ANPP and the effect was apparent after one year of nutrient addition. In addition to the nutrient use or demand, genome size directly affects minimum cell size and its variation has consequences for leaf gas exchanges and water use

efficiency (Beaulieu et al. 2008, Simonin and Roddy 2018). For example, Simonin and Roddy (2018) found that stomatal density, guard cell length, leaf vein density and stomatal conductance are more variable among species with small genomes. The variance in these traits might assist the plants with small genomes to finely tune their traits across a wider range of environmental gradients and increase fitness under environmental constraints. However, more work is needed to clarify whether this can be the underlying mechanism impacting the results in the current study. Moreover, the biomass fluctuation was larger in plants with large GS than with small GS after adding resources across the 10-yr period (Figure 5b, c). Results from the 10-yr resources manipulation study can represent responses to natural resource fluctuations over the relatively short term, while results from our transect study can represent the consequences for selection over long-term ecological scales. Taken together, these results confirm that the growth of the plant with large GS tends to be more sensitive to increased resource availability.

We limited our analysis to GS variation across plant species. However, variation in GS arises as a consequence of the amplification/deletion of transposable elements or polyploidy (i.e., whole-genome duplication) (Leitch et al. 2008). The variation of both GS and ploidy levels can potentially affect resource uptake and utilization of plants and their stressful tolerance in the environments (Lavergne et al. 2010, Meyerson et al. 2016, Pyšek et al. 2018). However, our study lacks ploidy information for most species and thus we cannot rule out the possible importance of ploidy in driving plant community assembly. Nevertheless, our work combines spatial investigation and temporal monitoring to provide robust evidence that the enhancement of resource availability can shift plant community and their community-level GS in grasslands. Knowledge of the patterns and dynamics of variation in GS is largely based on the analysis of European and North American floras, and species from the dryland ecosystem with relatively

low precipitation or world's important center of plant diversity, such as the northern grassland of China, are neglected (Herben et al. 2012). Our study opens a new avenue to understand how GS contributes to community shifts along gradients of environmental change and to gather more mechanistic and predictive insights into community assembly processes.

Acknowledgements

This work was supported by Natural Science Foundation of China (42130515 and 32171543), National Key Research and Development Program (2022YFF1300603) and Australian Research Council (FT190100590).

Author Contributions: XH, XL, CW, and HZ designed the study; HZ, XL, CW, XW, ZX, and HL performed the data collection; HZ performed modeling work and analysed output data; HZ wrote the first draft of the manuscript, and all authors contributed substantially to revisions.

References

Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography **19**:134-143.

Baselga, A., and F. Leprieur. 2015. Comparing methods to separate components of beta diversity. Methods in Ecology and Evolution **6**:1069-1079.

Beaulieu, J. M., I. J. Leitch, S. Patel, A. Pendharkar, and C. A. Knight. 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytologist **179**:975-986.

Cade, B. S., and B. R. Noon. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment **1**:412-420.

Chase, J. M. 2010. Stochastic community assembly causes higher biodiversity in more productive environments. Science **328**:1388-1391.

Doležel, J., J. Greilhuber, and J. Suda. 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nature Protocols **2**:2233.

Faizullah, L., J. A. Morton, E. I. Hersch-Green, A. M. Walczyk, A. R. Leitch, and I. J. Leitch. 2021. Exploring environmental selection on genome size in angiosperms. Trends in Plant Science **26**:1039-1049.

Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Diversity and Distributions **13**:252-264.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology **37**:4302-4315.

Fitzpatrick, M. C., N. J. Sanders, S. Normand, J.-C. Svenning, S. Ferrier, A. D. Gove, and R. R. Dunn. 2013. Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. Proceedings of the Royal Society B: Biological Sciences **280**:20131201.

Greilhuber, J., J. Doležel, M. A. Lysak, and M. D. Bennett. 2005. The origin, evolution and proposed stabilization of the terms 'genome size'and 'C-value'to describe nuclear DNA contents. Annals of Botany **95**:255-260.

Greilhuber, J., and I. J. Leitch. 2013. Genome size and the phenotype. Pages 323-344 *in* I. J. Leitch, J. Greilhuber, J. Doležel, and J. F. Wendel, editors. Plant Genome Diversity, Volume 2, Physical Structure, Behaviour and Evolution of Plant Genomes. Springer.

Grotkopp, E., M. Rejmánek, M. J. Sanderson, and T. L. Rost. 2004. Evolution of genome size in pines (Pinus) and its life-history correlates: supertree analyses. Evolution **58**:1705-1729.

Guignard, M. S., R. A. Nichols, R. J. Knell, A. Macdonald, C. A. Romila, M. Trimmer, I. J. Leitch, and A. R. Leitch. 2016. Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. New Phytologist **210**:1195-1206.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. Journal of Statistical Software **33**:1-22.

Herben, T., J. Suda, J. Klimešová, S. Mihulka, P. Říha, and I. Šímová. 2012. Ecological effects of cell-level processes: genome size, functional traits and regional abundance of herbaceous plant species. Annals of Botany **110**:1357-1367.

Knight, C. A., and D. D. Ackerly. 2002. Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. Ecology Letters **5**:66-76.

Knight, C. A., N. A. Molinari, and D. A. Petrov. 2005. The large genome constraint hypothesis: evolution, ecology and phenotype. Annals of Botany **95**:177-190.

Koenker, R. 2015. quantreg: Quantile Regression. R package version 5.11.http://cran.r-project.org/package=quantreg.

Lavergne, S., N. J. Muenke, and J. Molofsky. 2010. Genome size reduction can trigger rapid phenotypic evolution in invasive plants. Annals of Botany **105**:109-116.

Leitch, A., and I. Leitch. 2008. Genomic plasticity and the diversity of polyploid plants. Science **320**:481-483.

Leitch, I. J., L. Hanson, K. Lim, A. Kovarik, M. Chase, J. Clarkson, and A. Leitch. 2008. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). Annals of Botany **101**:805-814.

Manion, G., M. Lisk, S. Ferrier, K. M. Lugilde, M. C. Fitzpatrick, M. M. C. Fitzpatrick, I. Rcpp, and I. Rcpp. 2017. Package 'gdm'. A toolkit with functions to fit, plot, and summarize Generalized Dissimilarity Models: CRAN Repository.

McGill, B. J., B. J. Enquist, E. Weiher, and M. Westoby. 2006. Rebuilding community ecology from functional traits. Trends in Ecology & Evolution **21**:178-185.

Meyerson, L. A., J. T. Cronin, G. P. Bhattarai, H. Brix, C. Lambertini, M. Lučanová, S. Rinehart, J. Suda, and P. Pyšek. 2016. Do ploidy level and nuclear genome size and latitude of origin modify the expression of *Phragmites australis* traits and interactions with herbivores? Biological Invasions **18**:2531-2549.

Mokany, K., C. Ware, S. N. Woolley, S. Ferrier, and M. C. Fitzpatrick. 2022. A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. Global Ecology and Biogeography **31**:802-821.

Mori, A. S., F. Isbell, and R. Seidl. 2018. β-diversity, community assembly, and ecosystem functioning. Trends in Ecology & Evolution **33**:549-564.

Murphy, J., and J. P. Riley. 1962. A modified single solution method for the determination of phosphate in natural waters. Analytica Chimica Acta **27**:31-36.

Pellicer, J., O. Hidalgo, S. Dodsworth, and I. J. Leitch. 2018. Genome size diversity and its impact on the evolution of land plants. Genes **9**:88.

Pellicer, J., and I. J. Leitch. 2020. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. New Phytologist **226**:301-305.

Peng, Y., J. Yang, I. J. Leitch, M. S. Guignard, E. W. Seabloom, D. Cao, F. Zhao, H. Li, X. Han, Y. Jiang, A. R. Leitch, and C. Wei. 2022. Plant genome size modulates grassland community responses to multi-nutrient additions. New Phytologist **https://doi.org/10.1111/nph.18496**.

Pyšek, P., H. Skálová, J. Čuda, W. Y. Guo, J. Suda, J. Doležal, O. Kauzál, C. Lambertini, M. Lučanová, and T. Mandáková. 2018. Small genome separates native and invasive populations in an ecologically important cosmopolitan grass. Ecology **99**:79-90.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Robroek, B. J., V. E. Jassey, R. J. Payne, M. Martí, L. Bragazza, A. Bleeker, A. Buttler, S. J. Caporn, N. B. Dise, and J. Kattge. 2017. Taxonomic and functional turnover are decoupled in European peat bogs. Nature communications **8**:1-9.

Simonin, K. A., and A. B. Roddy. 2018. Genome downsizing, physiological novelty, and the global dominance of flowering plants. PLoS Biology **16**:e2003706.

Šmarda, P., and P. Bureš. 2010. Understanding intraspecific variation in genome size in plants. Preslia **82**:41-61.

Šmarda, P., M. Hejcman, A. Březinová, L. Horová, H. Steigerová, F. Zedek, P. Bureš, P. Hejcmanová, and J. Schellberg. 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.

Stebbins, G. L. 1971. Chromosomal evolution in higher plants. London, UK: Addison-Wesley.

van Breugel, M., D. Craven, H. R. Lai, M. Baillon, B. L. Turner, and J. S. Hall. 2019. Soil nutrients and dispersal limitation shape compositional variation in secondary tropical forests across multiple scales. Journal of Ecology **107**:566-581.

Vellend, M., D. S. Srivastava, K. M. Anderson, C. D. Brown, J. E. Jankowski, E. J. Kleynhans, N. J. Kraft, A. D. Letaw, A. A. M. Macdonald, and J. E. Maclean. 2014. Assessing the relative importance of neutral stochasticity in ecological communities. Oikos **123**:1420-1430.

Wang, C., X. Wang, D. Liu, H. Wu, X. Lü, Y. Fang, W. Cheng, W. Luo, P. Jiang, J. Shi, H. Yin, J. Zhou, X. Han, and E. Bai. 2014. Aridity threshold in controlling ecosystem nitrogen cycling in arid and semi-arid grasslands. Nature communications **5**:4799.

Wang, X., X. Lü, J. Yao, Z. Wang, Y. Deng, W. Cheng, J. Zhou, and X. Han. 2017. Habitat-specific patterns and drivers of bacterial β-diversity in China's drylands. The ISME Journal **11**:1345-1358.

Xu, Z., S. Wan, H. Ren, X. Han, M.-H. Li, W. Cheng, and Y. Jiang. 2012. Effects of water and nitrogen addition on species turnover in temperate grasslands in northern China. Plos One **7**:e39762.

Zhang, H. 2022. β-diversity in temperate grasslands is driven by stronger environmental filtering of plant species with large genomes. figshare. Dataset. https://doi.org/10.6084/m9.figshare.20452569.v1

Table 1 Phylogenetic generalized linear mixed model ($MCMC_{GLMM}$) coefficients (posterior mean), lower and upper 95% credible intervals (CIs) of parameters, the effective sample size taken from the chain, with significant pMCMC values (≤ 0.05 in bold). The $MCMC_{GLMM}$ model took the phylogenetic relationships among species, and plot included as random factors to test the effects of genome size (GS, log-transformed), resources addition (WNP, i.e., water, nitrogen and phosphorus), their interactions (GS: WNP) and plant growth form (annual vs. perennial) on the plant biomass in the resource addition experiment.

Fixed effects	Posterior mean	Lower, upper 95% CI	Effective sample size	pMCMC
Intercept	0.90	-23.77, 26.49	1007.0	0.973
GS	12.07	-2.18, 24.57	650.2	0.089
WNP	7.48	5.83, 9.30	900.0	< 0.001***
GS : WNP	9.51	6.12, 12.73	900.0	< 0.001***
Growth form	-2.41	-18.99, 12.26	900.0	0.776

Figure captions

Figure 1 Geographic distribution of 52 study sites across the 3200-km climatic gradient in the grasslands of northern China, with grey, black, green and blue points indicating alpine, desert, typical and meadow steppes, respectively. Duolun station, where we conducted our 10-yr resources manipulation experiment, as indicated by the red point. (b-d) The changes of the three environmental variables (mean annual precipitation, MAP; total soil nitrogen, N; and phosphorus, P) along the transect.

Figure 2 Results from the generalized dissimilarity model (GDM, for details see methods) based on species presence/absence data. (a) Relationship between $\beta_{turnover}$ and environmental dissimilarity based on resources of precipitation (mean annual precipitation, MAP), soil N, and soil P. (b) Relative importance of variables (i.e., percentage) for explaining variation in $\beta_{turnover}$. (c-f) Partial ecological distances (i.e. effects on $\beta_{turnover}$) showing the individual effects of each variable on $\beta_{turnover}$ for species with large (blue) and small (yellow) genomes. The separation of plant species into groups by genome size was based on the median genome size value of our studied species (1.61 pg/1C value). Locations on each line associated with larger values on the y-axis indicate an increased explanation associated with that variable. The shape of each curve indicates how the change rate of species turnover varies along the environmental gradient.

Figure 3 Results from the generalized dissimilarity model (GDM) for $\beta_{turnover}$ (based on Simpson index) by using different genome size thresholds. Plant communities were separated into the large- and small-genome size (1C DNA content, pg/1C) subsets by either the median (1.61 pg/1C) or mean (3.39 pg/1C) genome size for the 161 species of the present study or from the median (2.50 pg/1C) or mean (5.90 pg/1C) genome size of the global terrestrial plants in the Kew database (see *Materials and Methods*). Note that among the 161 species, five species had

two genome size values for each species to account for intra-specific or ploidy variation, leading to 166 values of genome sizes. (a) Variation driven by resources without the spatial component. (b) Variation driven by distance without the resources (i.e., rainfall, soil nitrogen and phosphorus) component. (c) The ratio of variation explained by resources versus geographic distance was calculated. (d) Measured genome size for the 166 species with genome sizes ranked from small to large along the x-axis, with red dashed lines indicating thresholds used in the analyses; the numbers inside parentheses indicate the number of species in each group.

Fig. 4 The relationship between genome size (pg/1C, log-transformed, GS) and mean annual precipitation (MAP), soil N and P concertation. The slope of the relationship between GS and the MAP (a), soil N (b) and soil P (c) concentrations across the 5^{th} to 95^{th} quantiles of the data. Blue solid lines are the least squares estimates for the coefficients, the double blue dashed lines are the 95% confidence intervals for the least squares estimates. The grey area represents the 95% confidence interval for the quantile regression estimates plotted as the red line.

Figure 5 (a) Comparison of community weighted means of genome size (1C DNA content; pg, log-transformed) for plots without (open circles; control) and with (filled circles) additional water, nitrogen and phosphorus (+Water+N+P) in a temperate steppe from 2007 to 2016 (except in 2011). Error bars indicate ± one standard error (n = 7). (b-c) Biomass production (g/m^2) of species with large (blue; panel b) and small (yellow; panel c) genomes for plots without and with resources addition. Plant species were separated into large vs. small genomes size (GS) groups based on the median value of the species at this local site (1.75 pg/1C value).

Figure 1

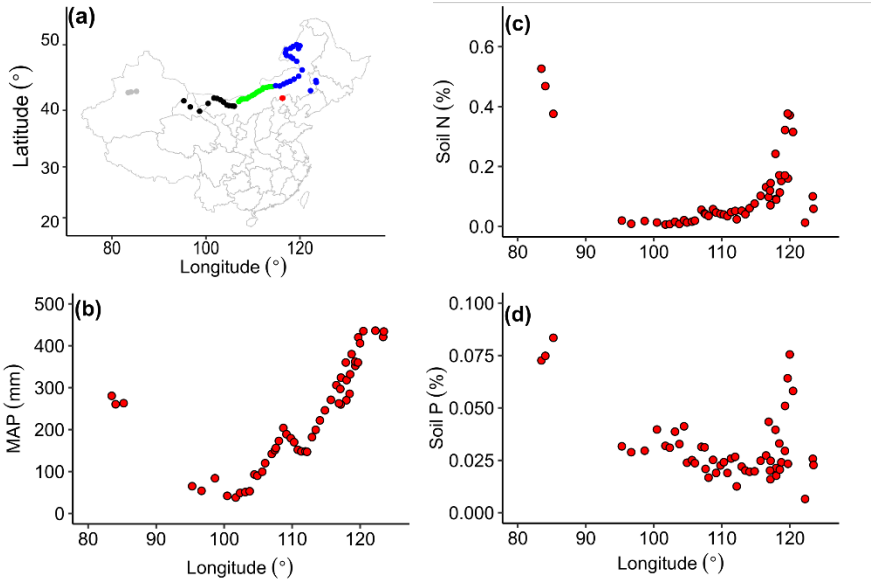

Figure 2

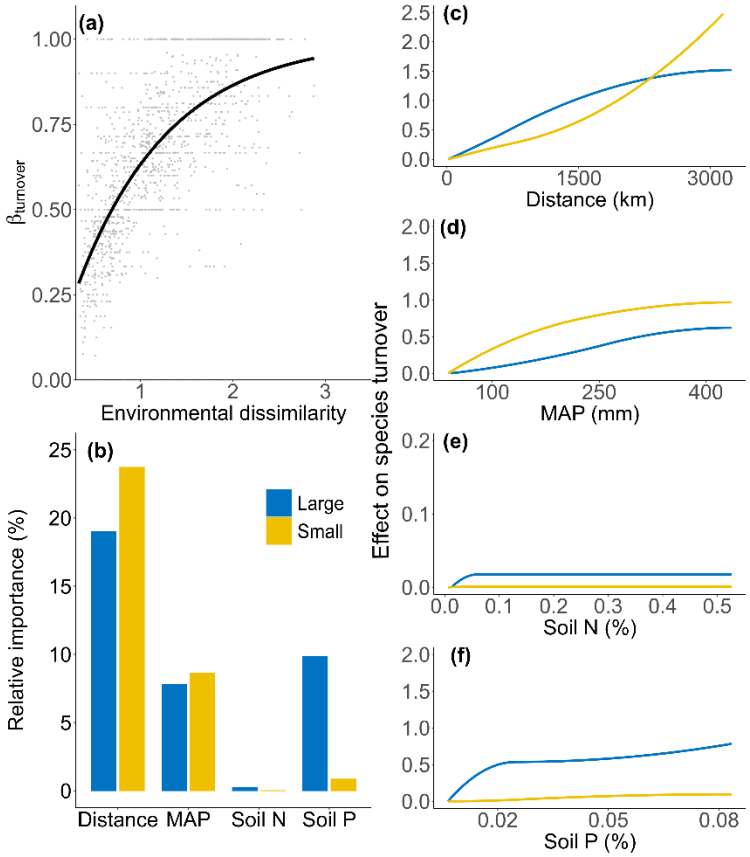

Figure 3

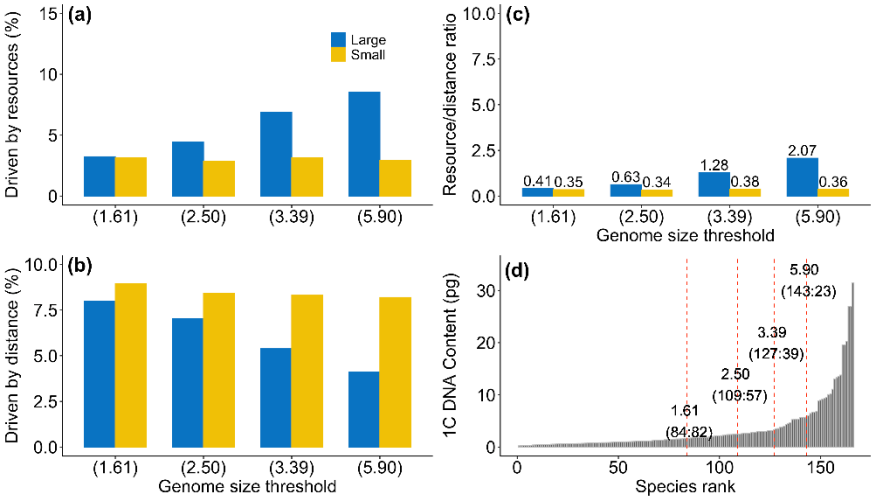

Figure 4

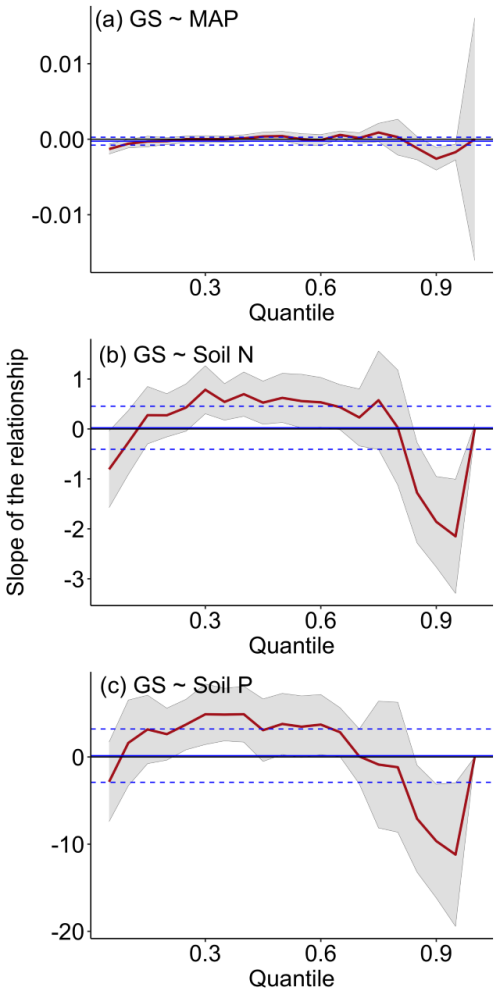

Figure 5