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

Hai-Yang Zhang[1], Xiaotao Lü[2], cunzheng wei[3], Jeff Powell[4], Xiaobo Wang[5], Dingliang Xing[6], and Xing-Guo Han[7]

[1]Western Sydney University Hawkesbury Institute for the Environment
[2]Institute of Applied Ecology, Chinese Academy of Sciences
[3]Institute of Botany Chinese Academy of Sciences
[4]Hawkesbury Institute for the Environnment
[5]Lanzhou University College of Pastoral Agriculture Science and Technology
[6]School of Ecological and Environmental Sciences, East China Normal University
[7]State Key Laboratory of Vegetation and Environmental Change

September 25, 2021

Abstract

Elucidating mechanisms underlying community assembly and biodiversity patterns is central to ecology and evolution. Genome size (GS, i.e. nuclear DNA content) determines species' capacity to tolerate environmental stress and therefore potentially drives 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 investigated their occurrences within plant communities across 52 sites spanning a 3200-km transect in the temperate grasslands of China. Using species distribution modelling, we found that environmental factors showed larger effects on β-diversity of large-GS than that of small-GS species and that communities with abundant resources had a greater representation of large-GS species. The latter finding was confirmed following analysis of data from a 10-yr resource (water, nitrogen, and phosphorus) manipulation experiment in which resource addition resulted in increased community weighted GS based on plant biomass estimates, suggesting that large-GS species are more sensitive to environmental resource limitation and explaining the greater environmental selection on β-diversity of large-GS species. These findings highlight the roles of GS in driving community assembly and predicting species responses to global change.

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; Kraft *et al.* 2011). 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 limitation and ecological drift (Laliberté*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; Blonder 2018). Traits that have been considered thus far are largely related to ecological strategy axes along which plant species vary in their abilities to acquire and allocate resources (Adler *et al.* 2014; Salguero-Gómez *et al.* 2016). As a fundamental trait that significantly varies across angiosperms (2400-fold) and precisely correlates with diverse phenotypic characters at cellular and organismal level, genome size (GS, i.e. nuclear DNA content), has received relatively little attention in the context of its role in community assembly. Genome size is generally constant within a species, which has functional consequences for species'

environmental tolerances, dispersal capacity, and interactions with other species (Knight & Ackerly 2002; Herben *et al.* 2012). The potential impact of plant genome size on community assembly processes is starting to be recognized (Greilhuber & Leitch 2013; Pellicer *et al.* 2018), but to date, its role in shaping β-diversity has not been tested empirically.

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 (Laliberté *et al.* 2014; van Breugel*et al.* 2019). Environmental filtering is hypothesized to differ for large- vs. small- GS species (Faizullah *et al.* 2021). First, the small-GS species grow faster due to 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 (Knight & Beaulieu 2008; Hessen *et al.* 2010). 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 & Leitch 2008; Hessen *et al.* 2010). Indeed, a recent study showed that large-GS species became more dominant under conditions with higher nutrient availability (Guignard *et al.* 2016). Thus, we expect that environmental filtering would have stronger effects on large-GS species than that for small-GS species.

To assess the roles that genome size plays in plant community assembly, we used data from 520 plant communities in 52 sites (10 plant communities per site) along a 3200-km transect in the temperate grasslands of northern China (Fig. 1). We measured plant genome size [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 genome size, for details see Greilhuber*et al.* (2005)] for 161 herbaceous species occurring along the transect (Fig. 1; Table S1-2). Generalized dissimilarity models [GDMs (Ferrier *et al.* 2007; Fitzpatrick *et al.* 2013)] were used to quantify the effects of genome size, environmental variation, and geographical distance on β-diversity along the gradient. We also estimated the importance of genome size as a continuous trait in driving species distribution with a joint species distribution model (Hierarchical Modelling of Species Communities [HMSC (Ovaskainen*et al.* 2017)]. HMSC can simultaneously model the environmental responses of multiple species accounting for shared habitats and evolutionary histories (i.e., phylogenetic correlations) and calculated the community weighted mean genome size along the environmental gradient (Tikhonov *et al.* 2020). We hypothesized that environmental filtering would play a more important role in driving the β-diversity for large-GS than that for small-GS species. Finally, to confirm the findings of our observational transect study, we also analysed data from a 10-yr field experiment manipulating resource availability to confirm that large-GS species and their higher resource requirements would be favored over small-GS species after reduction of resource limitation.

Materials and Methods

Transect study

Study sites

The transect study was conducted across a 3200-km scale of northern China's grasslands (Wang *et al.* 2017), extending from the Xinjiang Uygur Autonomous Region in the west to eastern Inner Mongolia (83.45° E to 120.36° E, 42.89° N to 49.19° N; Fig. 1). There are four vegetation types along this transect, including alpine-, desert-, typical- and meadow-steppe from west to east. Information on dominant species observed at each study site is supplied in Supplementary Dataset 1. Species richness decreased when moving 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 in 2012, during the period with peak aboveground

biomass. 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 represent the local natural vegetation (Wang *et al.* 2014). At each site, ten 1 m x 1 m quadrats were established. For each quadrat, we clipped the aboveground tissues of living plants, sorted these by species, and stored in paper bags. Community data from 10 quadrats were pooled together to represent the local species pool for each site. Soil samples from each quadrat were collected with five soil cores (2.5 cm diameter x 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 centre of the quadrat. Soils from the five soil cores were combined for each quadrat and sieved through a 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 and altitudes were recorded using a handheld GPS (eTrex Venture, Garmin, Olathe, Kansas, USA). Climate attributes, including mean annual temperature (MAT) and precipitation (MAP) of each sampling site were obtained from the global climate data set WorldClim 2.0 (1-km spatial resolution) (Fick & 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 & Riley 1962). Thus, our environmental predictors included MAP, MAT, altitude, soil N, and soil P.

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 (our study sites contained mainly herbaceous vegetation); of these we were able to obtain measurements of genome size 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 Supplementary Dataset 1 and 2).

For each species, three to twelve individuals were selected and leaf samples from each individual plant were measured for the genome size and the averaged value was used. For each individual, leaves from at least three fresh samples were analysed by flow cytometer (BD LSRFortessa, USA) (Doležel *et al.* 2007). Samples were chopped with a sharp razor blade for ca. 15-25s in a petri dish containing 1mL ice-cold nuclear LB01 isolation buffer (15 mmol/L Tris, 20 mmol/L EDTA-Na2, 0.5 mmol Spermine tetra-hydrochloride, 80 mmol/L KCl, 20 mmol/L NaCl, 0.1%volume percentage Triton X-100, 2 mmol/L DTT, pH=7.5). The homogenate was gently sucked up by pipette and filtered through a 50 μm nylon mesh. A volume of 50 μL of RNase (1 mg mL^{-1}; Sigma, St Louis, MO, USA) and 50 μL of Propidium iodide (PI, 1 mg mL^{-1}; CyStain PI Absolute P, Sysmex Partec GmbH Görlitz, Germany) were then added and mixed by gentle shaking. The mixture was incubated for 4 min in the dark at 4 °C and loaded onto the flow cytometer to measure genome size. The pg of DNA per nucleus in each peak is estimated by comparing measurements to a known internal reference standard. We had five species as our primary reference standard and two species as our secondary reference standard. For the primary reference standard, seeds were obtained from Centre of Plant Structural and Functional Genomics of the Institute of Experimental Botany, Czech Academy of Sciences, in 2017, including *Solanum lycopersicum* cv. Stupicke (1C = 0.98 pg) (Doležel *et al.* 1992),*Glycine max cv.* Polanká (1C = 1.25 pg) (Doležel *et al.*1994), *Zea mays* CE-777 (1C = 2.72 pg) (Lysak & Dolezel 1998),*Pisum sativum* cv. Ctirad (1C = 4.55 pg) (Doležel *et al.*1998), *Secale cereale* cv. Dankovske (1C = 8.10 pg) (Doležel*et al.* 1998). Based

on the primary reference standard, genome size of two local species: *Kerria japonica* (L.) DC (1C = 0.50 pg), *Hosta plantaginea* (Lam.) Aschers. (1C = 11.33 pg) were further calibrated against *S. lycopersicum* and *V. faba* and chosen as a secondary reference standard. The mean, standard deviation, and individual measurements of the studied species and the used internal standards are reported in Table S1. Note that we originally measured genome size for 169 species but we excluded eight species with high coefficients of variation in peak quantification (e.g. > 5 %) given their unprecise measurement (Doležel *et al.* 2007).

Among the 161 species sampled for genome size measurements, 23 species occurred more than once across the resampling sites, where we could check intraspecific variation of ploidy levels (Table S2). The resampling sites refer to the sites that have been investigated during the transect study but were revisited for measuring genome sizes. Most of these species showed quite stable 1C DNA content values across the transect, except *Artemisia frigida* and *Agropyro cristatum* . Previous studies had reported the 1C value for diploid *A. frigida*(2.63 pg) and *A. cristatum* (6.49 pg) (Garcia *et al.* 2008; Said *et al.* 2018), with their polyploid species being identified in nature (Wan *et al.* 2011; Zhao *et al.* 2017). In our case, the measured 1C values indicated that *A. frigida* and*A. cristatum* occurred more often as diploid (1C = 3.25 and 6.84 pg) in meadow steppe while as tetraploid (1C = 5.26 and 13.30 pg) in the other three grassland types. We report the diploid and tetraploid 1C values for those two species according to their occurring grassland type (Table S1) and used the mean value in our data analysis to handle the intraspecific ploidy variation.

For the 161 species, we assessed their contributions to biomass and species richness of the total plant community at each site. Sites were included in further analyses if, for that specific site, the species that had genome size values contributed to more than 80% of biomass and richness. Thus, for the final analysis, we used data for plant communities containing 161 species from 52 sites along the transect. Overall, the plant 1C values varied 260-fold from the smallest*Astragalus scaberrimus* (0.12 pg) to the largest *Allium ramosum* (31.5 pg), with a median and mean of 1.59 and 3.19 pg, respectively (Fig. S1; Table S1).

Phylogenetic trees

We constructed the plant phylogenetic tree using an updated version of the mega-phylogeny published by (Zanne *et al.* 2014) as the backbone. For those genera and species in our dataset that were absent from the meta-phylogeny, we used S.PhyloMaker (available at https://github.com/jinyizju/) to add them to their respective families (in the case of genera) and genera (in the case of species) in the mega-phylogeny under Scenario 3 (Qian & Jin 2016). The polytomies were resolved by the multi2di function in the ape package in R. Finally, our phylogenetic tree had 161 species (following APG IV, Fig. S1). We estimated phylogenetic signals for genome size by employing Blomberg's K and Pagel's λ tests, with the significance being estimated with randomization and likelihood ratio tests in the package phylosignal in R (Keck *et al.* 2016) (Fig. S1).

Data analysis

To match with the site-level plant community data, the corresponding soil property data were averaged for each site. Principal components analysis was applied to the bioclimatic variables for the sites and the five studied environmental variables (i.e. MAP, MAT, altitude, soil N, and soil P) to show how study sites can be characterized by these environmental variables (Fig. S2).

Beta diversity partitioning

Spatial variation in plant community composition was estimated from two components, species replacement (turnover component, β_{turnover}) and changes in species richness (nestedness component, $\beta_{\text{nestedness}}$), which together contribute to the total β-diversity (β_{total}). To test for effects on β-diversity that are independent of species richness differences, we first used the Simpson index of dissimilarity to examine how β_{turnover} changed along the environmental gradient. We also calculated the β_{total} with the Sørensen index to examine changes of the total β-diversity. We calculated the two β-diversity components via the betapart package in R (Baselga & Leprieur 2015). All these analyses were based on species presence/absence data.

Generalized Dissimilarity Modelling (GDM)

We used the GDM approach to analyse β-diversity patterns along environmental gradients, which is widely used to identify important environmental drivers for β-diversity and to test the independent significance of these drivers (using permutation tests). The advantage of GDMs is that they allow nonlinear relationships between dissimilarity and distance (Ferrier *et al.* 2007; Fitzpatrick *et al.*2013). The environmental matrix in our study included habitat variables (MAT, MAP, altitude, soil N, and soil P) and geographical distances (i.e. spatial distance from latitude and longitude) between sites.

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 maximum height of the line shows the relative importance of the studied predictor in explaining the variation of β-diversity in the model. 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 environment 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 environment 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 heterogeneity (linked to niche differentiation processes) in determining variability in β-diversity (Chase & Myers 2011). GDMs were fitted to the β-diversity for turnover component (β_{turnover}) and total (β_{total}), separately, using the gdm function in the gdm library (Manion *et al.* 2017). The results were generally similar for β_{turnover} and β_{total} (Fig. 2-3 vs. Fig. S3-4) and therefore we only reported the results for β_{turnover} in the man text for simplicity.

Modelling species responses with HMSC

We applied HMSC from the Hmsc R package (Tikhonov *et al.* 2020) to fit a joint species distribution model to plant community data simultaneously including information on traits (genome size), environmental covariates, and phylogenetic relationships in a single model (Pollock *et al.* 2012; Warton *et al.* 2015). We included MAP, MAT, altitude, soil N, and soil P as fixed effects (i.e., predictors) and used the sampling site as the sampling unit and the random effect. HMSC estimates species responses (slope parameters) to environmental covariates and uses these responses as the species' functional niche to estimate the strength and direction in which these niches moderate multiple species responses to environmental filtering. To account for the phylogenetic correlations among all the species, we included a phylogenetic correlation matrix (construction details of the phylogenetic trees see above) in the model's covariance structure. HMSC model can indicate the strength of the phylogenetic signal based on parameter ρ (varying from 0 to 1, a higher value indicating stronger phylogenetic signal) (Ovaskainen *et al.* 2017).

We fitted the model to the plant community (with a probit-link for the presence/absence data) with Bayesian inference, using the posterior sampling scheme (Ovaskainen *et al.* 2016). We used four Markov Chain Monte Carlo (MCMC) chains, each of which consisted of 15,000 iterations, out of which we discarded the first 5,000 as the burn-in and thinned the remaining by 10 to yield in total 1000 posterior samples per chain. We assessed the convergence of the MCMC chains by examining the distribution of the potential scale reduction factor over the parameters that measure the responses of the species to the fixed effects included in the model. The MCMC convergence of the HMSC models was satisfactory as the potential scale reduction factors (psrf) are close to one (the psrf of the ?-parameters that measure species response to environmental covariates (Ovaskainen *et al.* 2017) were on average 1.16). We examined the explanatory and predictive powers of the HMSC models through species-specific Tjur's R^2 values (Tjur 2009). Twofold cross-validation was performed to assess the predictive power of the model.

Based on the output of HMSC, we first characterized species' responses to each of the environmental variables as a mean parameter and its 95% posterior probability. Second, we predicted the community weighted mean genome size based on 161 species considered in the model as responses to the included environmental covariates. Note that genome size was log-transformed. Third, we used HMSC to capture the species-to-species associations (positive or negative co-occurrences) with latent variables. We considered statistical

support for these associations based on 90% posterior probabilities. Association patterns were classified as positive or negative, with this result used to estimate the frequency of two taxa co-existing (or not) compared to the frequency expected based on shared habitats or random occurrences.

Resources addition experiment

To further test whether large-GS species would be favored over small-GS species by increasing available resources, we obtained data from a long-term field resources manipulation experiment that was conducted in the typical steppe of Duolun, Xilingol in northern China (116. 3°E, 42.0°N). 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); because 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 $Ca(H_2PO_4)_2$, leading to 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 growing seasons every year from a 0.3 m × 0.3 m quadrat randomly chosen from within every plot. Biomass was sorted by species and was oven-dried at 65 °C for 48h to constant weight. From May to July 2017, genome size was measured for each species following the same procedure as in the transect study. Community weighted mean (CWM) genome size was calculated based on biomass for each plot. A linear mixed model was applied on CWM genome size, with resources treatment and experimental year as fixed effect while plot as a random effect. All analyses were conducted in R software 3.6.2 (R Core Team 2019).

Results and discussion

Γενομε σιζε ανδ β-διερσιτψ

The ranges of genome size for the species that occurred at each study site increased from west to east in our transect, with the maximum genome size for each study site also increasing from west to east (Fig. 1). Precipitation, soil nitrogen, and soil phosphorus (Fig S2) were all relatively higher at the eastern than that from the western end, which is consistent with our hypothesis that species with larger genomes would be more frequent where resource availability was greater.

Pairwise species turnover among communities ($\beta_{turnover}$) showed a positive correlation with increasing environmental dissimilarity (Fig. 2a), in line with many previous findings (Robroek*et al.* 2017; Mori *et al.* 2018). Plant species were then divided into two subsets based on whether their genome was greater or less than 2.5 pg (the median genome size for angiosperms globally (Leitch & Leitch 2013)). Consistent with our hypothesis, environmental variables explained more variation in $\beta_{turnover}$ for the large-GS than that for the small-GS species (Fig. 1b, d-h; Fig. 2a), while geographic distance explained more variation in $\beta_{turnover}$ for the small-GS than the large-GS species (Fig. 2b-c). For each specific environmental predictor, altitude, soil N, soil P, and MAP [marginally] explained more variation in $\beta_{turnover}$ for the large-GS species.

These results indicate that the relative importance (or effects) of environmental heterogeneity and geographic distance (which can indicate, in part, dispersal limitation) on plant β-diversity depend greatly on genome size. To check the sensitivity of our approach for separating species into large- and small-GS groups based on a single threshold (2.5 pg/1C), we repeated the GDM analysis after dividing the species based on the median value of our studied species (1.59 pg/1C). We found similar results, thus we report the results based on 2.5-threshold here in the main text and that of the 1.59-threshold in Fig S5. In addition, for variation partitioning, we also assessed the sensitivity of our results to four different thresholds in genome size [Fig. 3; median and mean values for the present study (1.59 and 3.19 pg) or for global angiosperms (2.5 and 5.90 pg) (Leitch & Leitch 2013)]. We observed qualitatively similar results, with the environment explaining the most variation for the large-GS species and geographic distance explaining the most variation for the small-GS species (Fig. 3a-c). The thresholds differed mainly in how variation was partitioned for species

classified to the large genome size category, with environment explaining more variation and geographic distance explaining less variation as the threshold increased, indicating that the species with the largest genomes were most strongly impacted by environmental filtering (Fig. 3).

We found spatial distances explained less variation in β-diversity for the large-GS species. In an attempt to link genome size and dispersal capacity among plant species, we also included the primary dispersal mode for the dominant species that we observed (Dataset S1). This includes genera that contain multiple species with large genomes, such as *Allium* and *Agropyron* , both of which tend to have an animal dispersal mode in seed dispersal. Previous studies have shown that genome size is positively correlated with seed mass (Beaulieu*et al.* 2007). Thus, lower dispersal limitation for the large-GS species might be due to their production of larger seeds and having a better chance of attracting dispersers. While confirmation requires further work, the associations that we have uncovered make the study of genome size a potentially powerful tool for understanding dispersal patterns in driving plant community assembly.

Genome size and species distribution: HMSC model

The HMSC model showed a good fit to the data, with a mean Tjur R^2 of 61% for explanatory power and 14% for predictive power. We observed both positive and negative associations with species occurrences for each variable (Fig. 4a). In our HMSC model, parameter ρ, the strength of the phylogenetic signal, was 0.1 in median, with the 95% confidence interval between 0.0 and 0.3 (overlap with zero), suggesting that species niches (i.e. their responses to the environmental covariates) showed no phylogenetic correlations. We then calculated the community weighted mean (CWM) genome size against environmental covariates by using predicted occurrences from the HMSC model (Fig. 4b). This is because using CWMs based on raw occurrence data in regressions can be subject to type I error (Miller *et al.*2019). The model approach to analyse trait–environment associations with community data maximizes power and information contained in the data (Brown *et al.* 2014; Miller *et al.* 2019). CWM of genome size was greater in communities from those sites with higher resource availability, supporting our hypothesis that alleviating resource (water and nutrient) limitation would favour large-GS species. Altitude was a poor predictor for CWM of genome size while MAT was negatively correlated with CWM of genome size. Previous studies also found that low temperature favours large-GS plants because the large-GS species gain an advantage in frost resistance (Grime & Mowforth 1982; MacGillivray & Grime 1995).

10-years water and nutrient addition in a field experiment

Results from the field manipulation experiment showed that the CWM of genome size increased significantly after resource addition ($F = 20.72, P < 0.001$; Fig. 5; Table S3). This effect was generally consistent across different years, with the interactive effect of resources addition and experimental year being marginally non-significant ($F = 2.02$, $P = 0.052$). These results were in line with our model prediction from HMSC, i.e., enhanced resources generally have positive effects on CWM genome size by favoring large-GS species and suggest that large-GS plants are stronger competitors than small-GS plants when resource availability increases. 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 and evolutionary scales. Taken together, these results confirm that increased resource availability favors large-GS plants and provide direction for a new dimension of ecological studies in the face of global change.

Water is the most limiting resource in temperate grasslands (Bai*et al.* 2004), here we showed that increased precipitation boosted the abundance of the large-GS plants within the community. Genome size is positively correlated with cell size while negatively correlated with tissue elasticity (Castro-Jimenez *et al.* 1989). Increased elastic tissue and smaller cell size are important for turgor maintenance in plants under drought stress, suggesting that natural selection should favor plants with smaller genomes in xeric environments (Castro-Jimenez*et al.* 1989). Similarly, a positive correlation was found between 2C DNA content of Californian angiosperms and annual precipitation (Knight & Ackerly 2002). In addition to water availability, soil nutrient availability can also constrain the performance of large-GS species. Previous studies found that

the large-GS species became more dominant after nutrient addition (Šmarda *et al.* 2013; Guignard*et al.* 2016), in line with our results from the transect study and the 10-yr resource manipulation experiment. Our work combines spatial investigation and temporal monitoring to show that the enhancement of resource availability can shift plant dominance in grassland communities by favoring large-GS species.

Genome size and species co-occurrence

Biotic interactions among plant species in the community that share a common set of environmental tolerances, in whole or in part, can also be an important driver of variation in species composition (Cornell & Lawton 1992; Tilman 1994; Stephens *et al.* 2020). We evaluated the direction and magnitude of these species' interactions via the residual species-to-species association matrices derived from the HMSC analysis (Fig. S6). We found that the frequencies of both positive and negative interactions (90% posterior probability) were greater when both members of the interacting pair had large genomes, compared with when one or both members possessed a small genome. Such difference became larger as the threshold used to differentiate large and small genome size increased (Fig. S6). In other words, both facilitation and competition were more likely to occur among the large-GS plants than with the small-GS plants. An outcome of these interactions could be that facilitation among the large-GS plants could lead to communities dominated by those species in resource-rich environments, while competition among those species under resource-limiting conditions may be responsible for the loss of the large-GS plants. Regardless, our results indicate that the importance of biotic filtering in structuring β-diversity was higher for the large-GS species.

Genome size and polyploids

Variation in genome size arises as a consequence of the amplification/deletion of transposable elements (Lee & Kim 2014) or polyploidy (i.e., whole-genome duplication), the latter often being accompanied by interspecific hybridization (allopolyploidy) (Leitch*et al.* 2008). Polyploidy usually affects species characteristics in environmental tolerance, adaptation, and plant-microbe interactions (Van de Peer *et al.* 2021). However, our current study lacks ploidy information for most species and additional studies are needed to determine whether observed effects of genome size in driving β-diversity arise from polyploidization or changes of transposable elements (Sheth*et al.* 2020).

Conclusion

One of the central questions in macroecology and biogeography is to understand why species composition differs among different sites. Our results indicate genome size can be used to bridge the ecological and evolutionary processes together for a deeper understanding of how species composition evolves over the biome scale (Segraves 2017; Pellicer *et al.* 2018). This opens a new avenue to understand how genome size contributes to community shifts along gradients of environmental change (including human-induced climate change such as shifting temperature and rainfall patterns) and to gather more mechanistic and predictive insights into community assembly processes. Given that climate models predict higher temperature and increased aridity in the temperate steppe (Day *et al.* 2018), species with large genomes may be more threatened by global change, and should therefore receive more attention in conservation efforts.**References**

Adler, P.B., Salguero-Gómez, R., Compagnoni, A., Hsu, J.S., Ray-Mukherjee, J., Mbeau-Ache, C. *et al.* (2014). Functional traits explain variation in plant life history strategies. *Proc. Natl. Acad. Sci. U.S.A.* , 111, 740-745.

Bai, Y., Han, X., Wu, J., Chen, Z. & Li, L. (2004). Ecosystem stability and compensatory effects in the Inner Mongolia grassland. *Nature* , 431, 181-184.

Baselga, A. & Leprieur, F. (2015). Comparing methods to separate components of beta diversity. *Methods Ecol. Evol.* , 6, 1069-1079.

Beaulieu, J.M., Moles, A.T., Leitch, I.J., Bennett, M.D., Dickie, J.B. & Knight, C.A. (2007). Correlated evolution of genome size and seed mass. *New Phytol.* , 173, 422-437.

Blonder, B. (2018). Hypervolume concepts in niche-and trait-based ecology. *Ecography* , 41, 1441-1455.

Brown, A.M., Warton, D.I., Andrew, N.R., Binns, M., Cassis, G. & Gibb, H. (2014). The fourth-corner solution–using predictive models to understand how species traits interact with the environment.*Methods Ecol. Evol.* , 5, 344-352.

Castro-Jimenez, Y., Newton, R., Price, H. & Halliwell, R. (1989). Drought stress responses of Microseris species differing in nuclear DNA content. *Am. J. Bot.* , 76, 789-795.

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

Chase, J.M. & Myers, J.A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* , 366, 2351-2363.

Cornell, H.V. & Lawton, J.H. (1992). Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *J. Anim. Ecol.* , 61, 1-12.

Day, J.A., Fung, I. & Liu, W. (2018). Changing character of rainfall in eastern China, 1951–2007. *Proc. Natl. Acad. Sci. U.S.A.* , 115, 2016-2021.

Doležel, J., Doleželová, M. & Novák, F. (1994). Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biol. Plant.* , 36, 351-357.

Doležel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysák, M., Nardi, L. *et al.* (1998). Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Ann. Bot.* , 82, 17-26.

Doležel, J., Greilhuber, J. & Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* , 2, 2233.

Doležel, J., Sgorbati, S. & Lucretti, S. (1992). Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plant* , 85, 625-631.

Faizullah, L., Morton, J.A., Hersch-Green, E.I., Walczyk, A.M., Leitch, A.R. & Leitch, I.J. (2021). Exploring environmental selection on genome size in angiosperms. *Trends Plant Sci.* , https://doi.org/10.1016/j.tplants.2021.1006.1001.

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

Fick, S.E. & Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* , 37, 4302-4315.

Fitzpatrick, M.C., Sanders, N.J., Normand, S., Svenning, J.-C., Ferrier, S., Gove, A.D. *et al.* (2013). Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. *Proc. Royal Soc. B.* , 280, 20131201.

Garcia, S., Canela, M.A., Garnatje, T., McArthur, E.D., Pellicer, J., Sanderson, S.C. *et al.* (2008). Evolutionary and ecological implications of genome size in the North American endemic sagebrushes and allies (*Artemisia* , Asteraceae). *Biol. J. Linn. Soc.* , 94, 631-649.

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

Greilhuber, J. & Leitch, I.J. (2013). Genome size and the phenotype. In: *Plant Genome Diversity, Volume 2, Physical Structure, Behaviour and Evolution of Plant Genomes* (eds. Leitch, IJ, Greilhuber, J, Doležel, J & Wendel, JF). Springer, pp. 323-344.

Grime, J. & Mowforth, M. (1982). Variation in genome size—an ecological interpretation. *Nature* , 299, 151.

Guignard, M.S., Nichols, R.A., Knell, R.J., Macdonald, A., Romila, C.A., Trimmer, M. *et al.* (2016). Genome size and ploidy influence angiosperm species' biomass under nitrogen and phosphorus limitation. *New Phytol.* , 210, 1195-1206.

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

Hessen, D.O., Jeyasingh, P.D., Neiman, M. & Weider, L.J. (2010). Genome streamlining and the elemental costs of growth. *Trends Ecol. Evol.* , 25, 75-80.

Keck, F., Rimet, F., Bouchez, A. & Franc, A. (2016). phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol. Evol.* , 6, 2774-2780.

Knight, C.A. & Ackerly, D.D. (2002). Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecol. Lett.* , 5, 66-76.

Knight, C.A. & Beaulieu, J.M. (2008). Genome size scaling through phenotype space. *Ann. Bot.* , 101, 759-766.

Knight, C.A., Molinari, N.A. & Petrov, D.A. (2005). The large genome constraint hypothesis: evolution, ecology and phenotype. *Ann. Bot.* , 95, 177-190.

Kraft, N.J.B., Comita, L.S., Chase, J.M., Sanders, N.J., Swenson, N.G., Crist, T.O. *et al.* (2011). Disentangling the drivers of beta diversity along latitudinal and elevational gradients. *Science* , 333, 1755-1758.

Laliberté, E., Zemunik, G. & Turner, B.L. (2014). Environmental filtering explains variation in plant diversity along resource gradients. *Science* , 345, 1602-1605.

Lee, S.-I. & Kim, N.-S. (2014). Transposable elements and genome size variations in plants. *Genomics Inform* , 12, 87-97.

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

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

Leitch, I.J. & Leitch, A.R. (2013). Genome size diversity and evolution in land plants. In: *Plant Genome Diversity Volume 2* . Springer, pp. 307-322.

Lysak, M.A. & Dolezel, J. (1998). Estimation of nuclear DNA content in *Sesleria* (Poaceae). *Caryologia* , 51, 123-132.

MacGillivray, C. & Grime, J. (1995). Genome size predicts frost resistance in British herbaceous plants: implications for rates of vegetation response to global warming. *Funct. Ecol.* , 9, 320-325.

Manion, G., Lisk, M., Ferrier, S., Lugilde, K.M., Fitzpatrick, M.C., Fitzpatrick, M.M.C. *et al.* (2017). Package 'gdm'. *A toolkit with functions to fit, plot, and summarize Generalized Dissimilarity Models: CRAN Repository, R* .

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* , 21, 178-185.

Miller, J.E., Damschen, E.I. & Ives, A.R. (2019). Functional traits and community composition: A comparison among community-weighted means, weighted correlations, and multilevel models. *Methods Ecol. Evol.* , 10, 415-425.

Mori, A.S., Isbell, F. & Seidl, R. (2018). β-diversity, community assembly, and ecosystem functioning. *Trends Ecol. Evol.* , 33, 549-564.

Murphy, J. & Riley, J.P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Anal. Chim. Acta* , 27, 31-36.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods Ecol. Evol.* , 7, 549-555.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D. *et al.* (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* , 20, 561-576.

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

Pollock, L.J., Morris, W.K. & Vesk, P.A. (2012). The role of functional traits in species distributions revealed through a hierarchical model.*Ecography* , 35, 716-725.

Qian, H. & Jin, Y. (2016). An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *J Plant Ecol* , 9, 233-239.

R Core Team (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.*

Robroek, B.J., Jassey, V.E., Payne, R.J., Martí, M., Bragazza, L., Bleeker, A. *et al.* (2017). Taxonomic and functional turnover are decoupled in European peat bogs. *Nat. Commun.* , 8, 1-9.

Said, M., Hřibová, E., Danilova, T.V., Karafiátová, M., Čížková, J., Friebe, B. *et al.* (2018). The *Agropyron cristatum*karyotype, chromosome structure and cross-genome homoeology as revealed by fluorescence in situ hybridization with tandem repeats and wheat single-gene probes. *Theor. Appl. Genet.* , 131, 2213-2227.

Salguero-Gómez, R., Jones, O.R., Jongejans, E., Blomberg, S.P., Hodgson, D.J., Mbeau-Ache, C. *et al.* (2016). Fast–slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proc. Natl. Acad. Sci. U.S.A.* , 113, 230-235.

Segraves, K.A. (2017). The effects of genome duplications in a community context. *New Phytol.* , 215, 57-69.

Sheth, S.N., Morueta-Holme, N. & Angert, A.L. (2020). Determinants of geographic range size in plants. *New Phytol.* , 226, 650-665.

Šmarda, P., Hejcman, M., Březinová, A., Horová, L., Steigerová, H., Zedek, F. *et al.* (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 Phytol.* , 200, 911-921.

Stephens, C.R., Gonzalez-Salazar, C., Villalobos, M. & Marquet, P. (2020). Can Ecological Interactions be Inferred from Spatial Data?*Biodivers. inform.* , 15, 11-54.

Tikhonov, G., Opedal, O.H., Abrego, N., Lehikoinen, A., de Jonge, M.M., Oksanen, J. *et al.* (2020). Joint species distribution modelling with the r-package Hmsc. *Methods Ecol. Evol.* , 11, 442-447.

Tilman, D. (1994). Competition and biodiversity in spatially structured habitats. *Ecology* , 75, 2-16.

Tjur, T. (2009). Coefficients of determination in logistic regression models—A new proposal: The coefficient of discrimination. *Am Stat* , 63, 366-372.

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

Van de Peer, Y., Ashman, T.-L., Soltis, P.S. & Soltis, D.E. (2021). Polyploidy: an evolutionary and ecological force in stressful times.*Plant Cell* , 33, 11-26.

Wan, T., Sun, Q., Cai, P., Meng, X., Yi, W. & Wang, W. (2011). Observation of chromosome karyotypes of *Artemisia frigida* in different ecological areas of Inner Mongolia. *Acta Botanica Boreali-Occidentalia Sinica* , 31, 456-461.

Wang, C., Wang, X., Liu, D., Wu, H., Lu, X., Fang, Y. *et al.*(2014). Aridity threshold in controlling ecosystem nitrogen cycling in arid and semi-arid grasslands. *Nat. Commun.* , 5, 1-8.

Wang, X.-B., Lu, X.-T., Yao, J., Wang, Z.-W., Deng, Y., Cheng, W.-X.*et al.* (2017). Habitat-specific patterns and drivers of bacterial β-diversity in China's drylands. *ISME J.* , 11, 1345-1358.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. *et al.* (2015). So many variables: joint modeling in community ecology. *Trends Ecol. Evol.* , 30, 766-779.

Xu, Z., Wan, S., Ren, H., Han, X., Li, M.-H., Cheng, W. *et al.*(2012). Effects of water and nitrogen addition on species turnover in temperate grasslands in northern China. *PloS one* , 7, e39762.

Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G. *et al.* (2014). Three keys to the radiation of angiosperms into freezing environments. *Nature* , 506, 89-92.

Zhao, Y., Xie, J., Dou, Q., Wang, J. & Zhang, Z. (2017). Diversification of the P genome among *Agropyron Gaertn* .(Poaceae) species detected by FISH. *Comp. Cytogenet.* , 11, 495-509.

Acknowledgements

We are grateful to all the team members in 2012-Transect-Investigation in the northern grassland of China. We thank Otso Ovaskainen, Nerea Abrego, Øystein Opedal, and others for organizing the workshop:*Multivariate modelling in ecology and joint species distribution models* . This work was supported by National Key Research and Development Program of China (2016YFC0500703) and Natural Science Foundation of China (31822006).

Figure legends

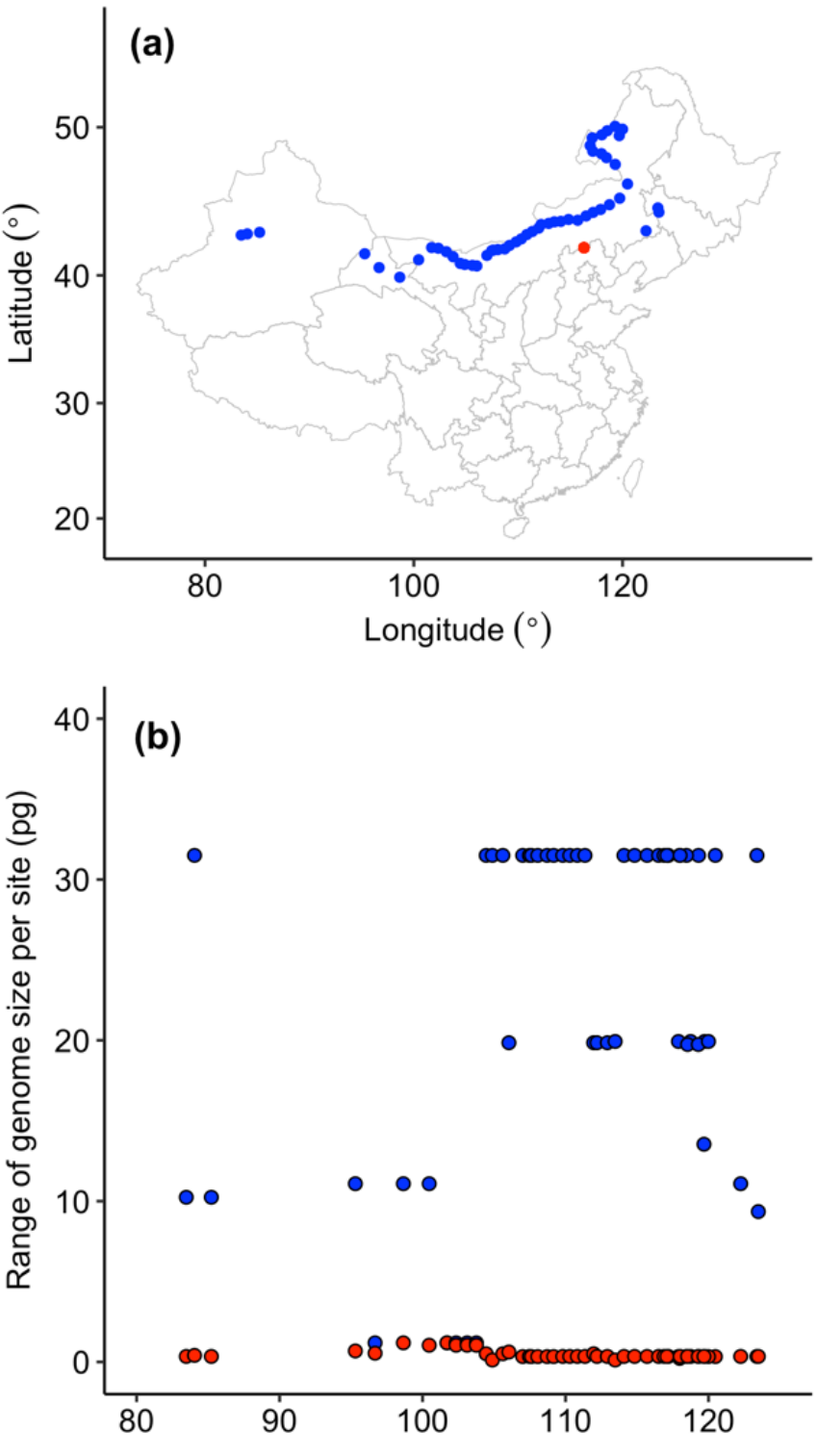

Fig. 1. Distribution and species genome size range for each of the transect sites. (a) Geographic distribution of the 3200-km (52 study sites, blue points) along the climatic gradient in the grasslands of northern China. Duolun station, where we conducted our 10-yr resources manipulation experiment, was indicated by red point. (b) Range of genome size for the species that occurred for each study site along the transect. Blue points indicate the species that have the maximum genome size while red points indicate the minimum one.

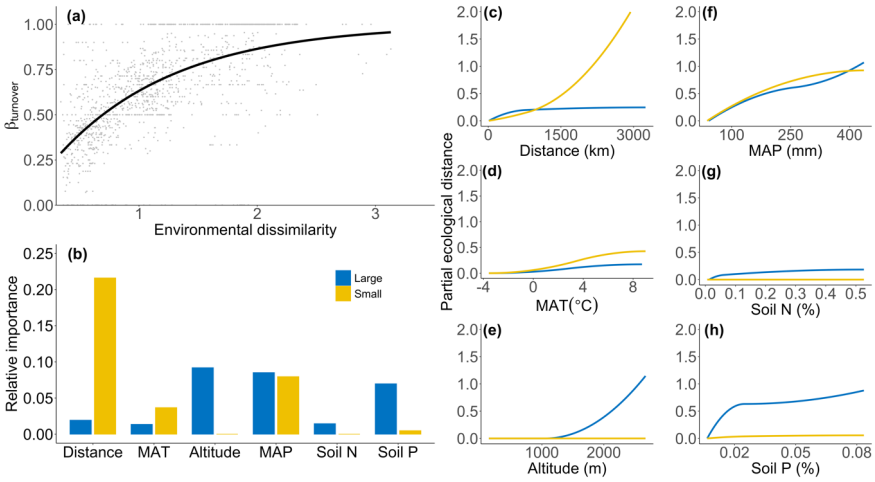

Fig. 2 Παιρωισε ςομποσιτιοναλ δισσιμιλαριτψ αμονγ πλαντ ςομμυνιτιες (τυρνοερ ςομπονεντ, $\beta_{\text{τυρνοερ}}$, βασεδ ον Σιμπσον ινδεξ) αλονγ τηε 3200-κμ ενιρονμενταλ γραδιεντ. These results were generated from the generalized dissimilarity model (GDM, for details see methods) based on presence/absence data. (a) Relationship between β_{turnover} and environmental dissimilarity based on all the environmental variables (MAP, MAT, altitude, soil N, and soil P). (b) Relative importance of variables for explaining variation in β_{turnover}. (c-h) Partial ecological distances (i.e. effects on β_{turnover}) showing the individual effects of each variable on β_{turnover} for species with large (blue) and small (yellow) genomes. The separation of plants into groups by genome size was based on the global median genome size value of terrestrial plants (2.5 pg/1C value) in the Kew database (see*Materials and Methods*). Results for separation into groups based on the median value of our studied species (1.59 pg/1C value) are supplied in **Fig. S3** . Locations on each line associated with larger values on the y-axis indicate an increased likelihood of observing that genome-size group at that value along the x-axis, and higher maxima in curves indicate larger effects associated with that variable overall.

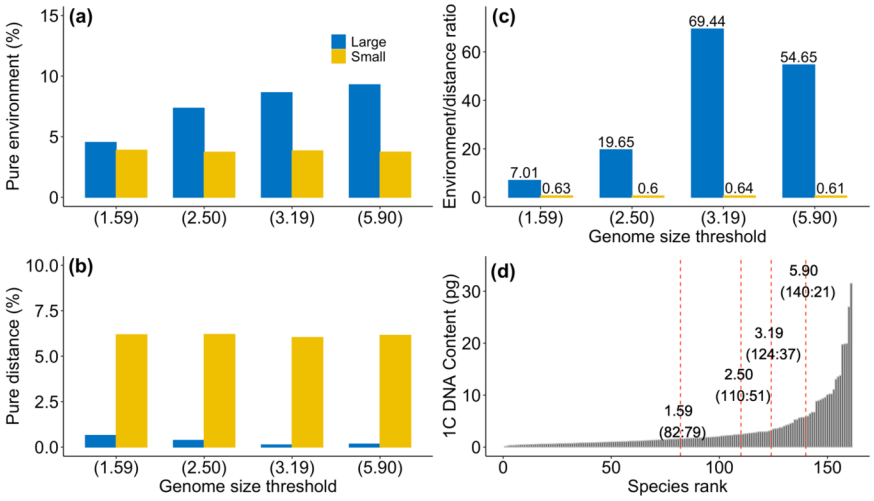

Fig. 3 Τηε προπορτιον οφ αριανσε (% αριατιον) ιν σομποσιτιοναλ δισσιμιλαριτψ αμονγ πλαντ σομμυνιτιες ($\beta_{TURNOVER}$, βασεδ ον Σιμπσον ινδεξ) ιν εαση συβσετ (λαργε/σμαλλ γενομε σιζε) ας εξπλαινεδ βψ ενιρονμενταλ αριαβλες (ινσλυδινγ MAΠ, MAT, αλτιτυδε, σοιλ N, ανδ σοιλ Π) ανδ γεογραπηιςαλ διστανσε υσινγ διφφερεντ γενομε σιζε τηρεσηολδς. Plant communities were separated into the large- and small- genome size (1C DNA content, pg) subsets by either the median (1.59 pg) or mean (3.19 pg) genome size for the 161 species of the present study or from the median (2.50 pg) or mean (5.90 pg) genome size of the global terrestrial plants in the Kew database (see *Materials and Methods*). Generalized dissimilarity models (GDMs) were fitted using presence/absence data for the large- (blue) and small- (yellow) genome size subsets. (a) Pure environmental variation without a spatial component represents the effect of environmental filtering. (b) Pure distance variation without an environmental component can represent the effect of dispersal limitation if all important environmental drivers are included in the environmental component. (c) The ratio of variation explained by pure environmental variation versus geographic distance was calculated, with estimates also shown numerically on top of each bar. (d) Measured genome size for the 161 species 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.

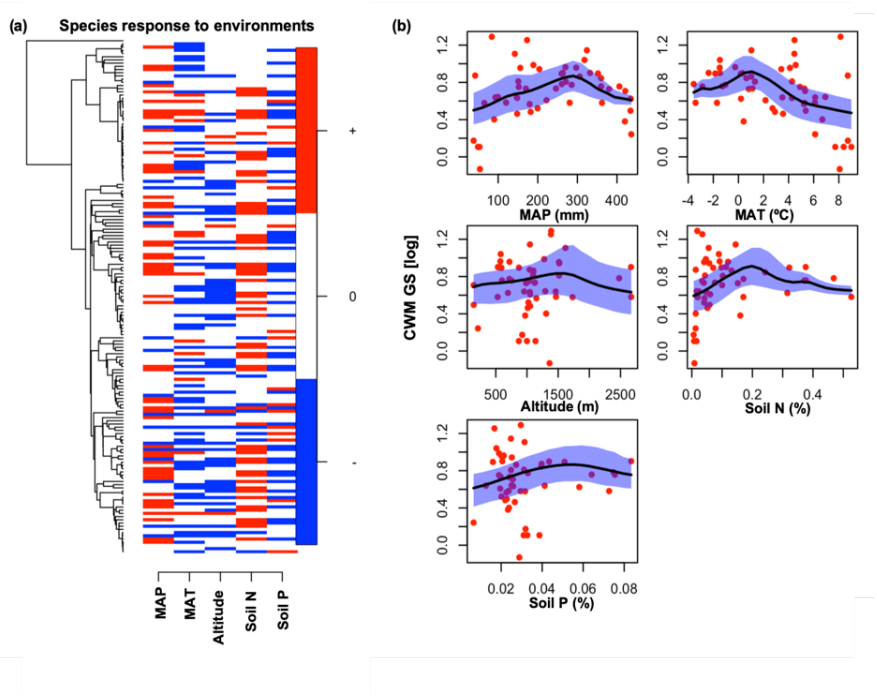

Fig. 4 Results of HMSC models showing effects of environmental filtering based on species presence/absence data. (a) HSMC-based estimates of species responses to the environmental covariates (MAP, MAT, altitude, soil N and P), i.e β parameter. Plant species were listed according to their phylogenetic relationships (a high-resolution tree with detailed genome size (GS, 1C DNA content, pg) was supplied in Supplementary Fig. S1). The figure legend in panel (a) indicates positive (red) and negative (blue) responses to the environmental covariates, based on credible intervals with posterior probability at least 95%. (b) Model-based community weighted means (CWMs) of genome size (log-transformed) and its response to the environmental covariates. The CWM of genome size was calculated based on the prediction from the HMSC model.

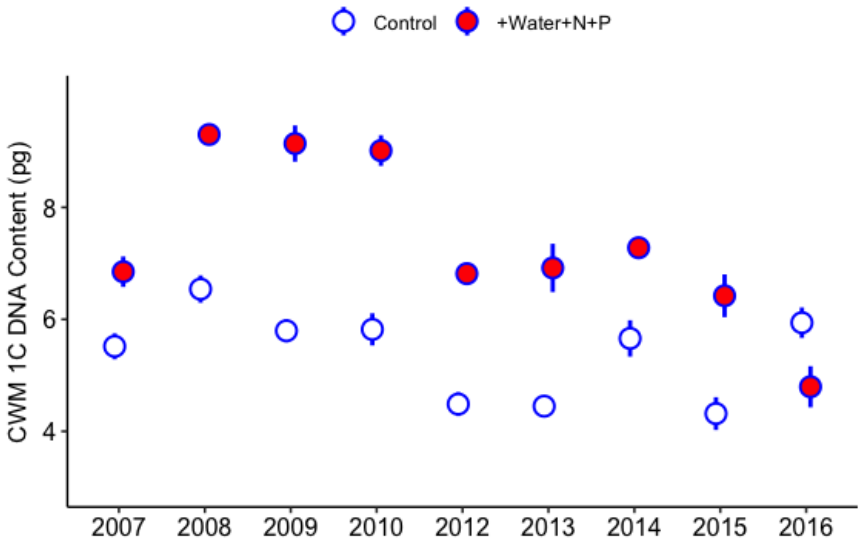

Fig. 5. Comparison of community weighted means of genome size (1C DNA content; pg) for plots without (open circles; control) and with (filled circles) additional water, nitrogen and phosphorus (+Water+N+P) in a temperate steppe from 2007 to 2016 (except in 2011). Error bars indicate ± one standard error (n = 7).