Supporting Information

Šmarda et al. 10.1073/pnas.1321152111

SI Methods

Plant Material. Plants were collected in the wild or sourced from various institutional or personal collections (Dataset S1, Table S1). Wherever possible, diploid species were selected for study. However, diploids were not always available, especially for some of the more exotic monocot families, where obtaining any material to study is difficult and ploidy-level data are frequently lacking. Whenever possible, herbarium vouchers were prepared and stored in the Herbarium of the Department of Botany and Zoology, Masaryk University. Alternatively, photographic vouchers were prepared, and these images are available on request.

Genomic DNA Base Composition and Genome Size Measurements. Genomic parameters were measured by flow cytometry using two different fluorochromes: the DNA intercalating propidium iodide (PI; measuring the absolute 2C genome size) and the AT-selective DAPI (measuring the AT fraction of the genome). Details of the method, based on the work by Barow and Meister (1), are described in the works by Šmarda et al. (2, 3). In the species with large amounts of secondary compounds, samples for PI measurements were prepared in Tris $MgCl_2$ buffer (4) separately from the samples measured with DAPI. Analyses were made on four flow cytometers (PA-I, CyFlow SL, and two CyFlow ML flow cytometers; Partec GmbH). Measurements of each species were always done with an internal standard (Dataset S1, Table S5) and repeated at least three times on different days. At least 5,000 nuclei were analyzed in each measurement, with peaks accepted generally only if their coefficient of variance was below 3.5 (measurements with DAPI) or 5 (measurements with PI). Genomic DNA base compositions (GC contents) were calculated from the flow cytometry data using the Excel spreadsheet in the work by Šmarda et al. (2), which is available at http://sci.muni.cz/botany/systemgr/download/Festuca/ATGCFlow.xls. Measurements made on different samples of the same species with the same ploidy level were averaged before statistical analyses. Altogether, 249 samples comprising 239 species (including two species with two distinct ploidy levels) were measured, and the results are presented in Dataset S1, Tables S1 and S2.

The genome sizes and GC contents of the internal standards were derived from comparison with the fully sequenced cultivar of *Oryza sativa* Nipponbare (2C = 777.64 Mbp, GC = 43.6%, genome cover = 95%) (5), the species with the most complete genomic sequence for an angiosperm to date. The genomic parameters of internal standards were calculated directly from measurements with rice (for *Solanum lycopersicum* and *Raphanus sativus*) or derived from our measurements with *Solanum lycopersicum* (for *Carex acutiformis* and *Bellis perennis*) (Dataset S1, Table S5). In the case of peak overlaps between the standard and the sample, internal standardization with one of the already measured samples was used as an alternative (Dataset S1, Table S1). Genomic parameters of standards with higher genome sizes were derived from the standard sample ratios with measurements using the respective dyes given in the work by Doležel et al. (6) (Dataset S1, Table S5). Using this approach together with data for sample standard ratios given in the work by Doležel et al. (6), human cells are estimated to have a 2C genome size of 6,055.03 Mbp (6.19 pg). This estimate is very close to the predicted 2C DNA content of the human genome by the International Human Genome Sequencing Consortium (6,153.33 Mbp, 6.29 pg) (7) as well as estimates of human 2C genome size achieved using alternative biochemical methods (8). The genome size estimates reported here are, therefore, ∼12% lower compared with other genome size estimates, which were determined from standards with genome sizes that were derived from an overestimated genome size of human of 2C = 7.0 pg (cf. 8).

Phylogenetic Tree of Measured Species. The phylogenetic tree of the studied taxa was obtained by pruning the recent large-scale dated angiosperm phylogeny of Zanne et al. (9) (ZAP) (Fig. 1, *SI Methods*, and Figs. S1–S3). ZAP contains >30,000 angiosperm species and includes ∼70% of the species studied here. For the remaining 30% of species, we identified the closest relatives of the species studied here from ZAP and used them as surrogates to obtain insights into their phylogenetic position. In doing so, we always tried to find the closest possible relative to the measured species in ZAP (Dataset S1, Table S2), although in many cases, simple random selection would result in the same topology and branch lengths on the tree [e.g., having two species belonging to two different genera that are not in the ZAP, we may select randomly another one species from the same genera to show the phylogenetic relationships of the genera, age of their common ancestor (i.e., their divergence times), and corresponding branch lengths in an ultrametric tree]. In rare cases, if there was a complete absence of very close relatives of a measured species or even a whole genus in ZAP, the species were added manually to the tree according to information from other relevant phylogenetic literature (Dataset S1, Table S2). In such cases, a newly created node was added arbitrarily in the middle of the corresponding branch. Only for *Agapanthus* and *Tetraria* was phylogenetic information so scarce that we were forced to treat relationships of taxa within these genera as polytomies.

Because the branching pattern and dating of Poaceae in ZAP was too different from the accepted views on the evolution of this group by the grass phylogenetic community, the ZAP tree for spikelet grasses was replaced with a recent 531-species phylogeny of grasses taken from the Grass Phylogeny Working Group II (10). The controversy with ZAP was the nonmonophyly of the two traditional grass clades [i.e., Bambusoideae+Ehrhartoideae+Pooideae (BEP) and Panicoideae+Arundinoideae+Chloridoideae+Centothecoideae+Micrairoideae+Aristidoideae+Danthonioideae (PACCMAD)] and the dating of their divergence, which is much younger in ZAP than indicated by the evidence from numerous fossils and assumed by the modern grass literature (11–14) (given perhaps by an absence of reliable calibrations points for grasses in ZAP). For the purposes of our study, the original undated Poaceae tree by the Grass Phylogeny Working Group II (10) was dated using the likelihood procedure of the chronos function of the R package ape (15). The node ages obtained were then applied to 23 nodes in the pruned spikelet grasses tree (BEP and PACCMAD clades). Of four tested dating models (relaxed, correlated, and discrete with either one or two rate categories), the discrete model with a single rate category (strict clock model) was found to describe the data in the best way (Φ information criterion, ΦIC = 1,205); this latter model was selected for the final calibration. Two calibration points widely accepted by the grass phylogenetic community were selected: appearance (crown node) of spikelet grasses 55–68 Mya documented from fossils (13) [the upper age limit was selected so as not to conflict with the dating of the divergence of Restionaceae and Poaceae in ZAP (9)] (Dataset S1, Table S2) and Panicoideae and Chloridoideae divergence 14–55 Mya (cf. 14). After replacing the node of spikelet grasses in ZAP with that by the Grass Phylogeny Working Group II, the node ages of taxa (and corresponding branch lengths) in the ZAP, situ-

ated between the newly dated crown node of grasses (55 Mya) and the Poaceae and Restionaceae ancestor (68 Mya), were narrowed into this time interval in proportion to their original dating between these two nodes in ZAP. The final monocots phylogeny is provided in Dataset S3.

Life History Traits. ***Natural distribution ranges of species (variable continent).*** Data on the natural distribution ranges of the studied species were compiled mainly from the World Checklist of Selected Plant Families (http://apps.kew.org/wcsp/) and occasionally from data extracted from local floras. For statistical analyses, the occurrence of each species was scored as presence/absence using the following geographic regions.

Eurasia: Eurasian continent (including Macaronesia and the Indonesian archipelago and excluding New Guinea)

Africa: African continent and Madagascar

N_America: North American continent north of Mexico

S_C_America: South and Central America (including Mexico and south)

Australia: Australian continent and Tasmania

Oceania: Including New Zealand, New Guinea, and the archipelagos in the Pacific Ocean

Species that occurred in two or more continents were assigned as cosmopolitan.

The extent of the distribution area of species (variable distribution). This variable was scored in a 3° ordinal scale.

Local: Species with localities within an area of ~500 × 500 km

Common: Species with localities within an area of ~5,000 × 5,000 km

Widespread: Species localities over an area of ~5,000 × 5,000 km

The scoring of species was done by combining Global Biodiversity Information Facility (GBIF) data with information from local flora accounts. In the statistical analyses, this variable was tested as either nominal or numeric.

General ecology and habitat preferences (variable biome). Because of the absence of detailed ecological data available for the majority of studied species, the general ecology of a species was categorized based on its affinity to a particular biome or azonal habitat type (such as marine environment, wetlands, or aquatic bodies). The following categories were used.

Temperate forests: Temperate forests and woodlands

Sclerophyllous forests: Warm temperate and subtropical evergreen sclerophyllous forests

Tropical rainforests: Closed canopy forests of the high-precipitation and all-year warm tropical zone

Mediterranean: Temperate sclerophyllous woodlands and scrub in regions characterized by Mediterranean type and seasonal climates

Grasslands: Natural C3 grass-dominated grasslands characterized by cold winter climate

Savannas: C4 grass-dominated grasslands, often with scattered presence of trees and shrubs of subtropical and tropical regions

Semideserts and deserts: Vegetation of (extremely) arid ecosystems characterized by low (and often, poorly predictable) precipitation patterns

Arctic and alpine: Grass- and low scrub-dominated vegetation developing at high altitudes (above the tree line) and high latitudes

Freshwater wetlands: Freshwater swampy habitats and associated water bodies characterized by at least temporary high water supply (temporary flooding and year-round availability of groundwater)

Marine: Marginal regions of the sea supporting sea grass meadows

The ecological data for each species and its affinity to a particular biome or azonal wetland type were sourced from available ecological literature and expert opinion. Only the dominant (in some cases, up to two different categories) ecology is featured. Species recorded in two different categories were deemed unspecific.

In addition to biome, two other ecological variables were also recorded.

Open. Cases where a species tends to dominate in open, treeless habitats characterized by increased stress from UV radiation with low and irregular availability of water and large fluctuations in daily temperature.

Moisture. Refers to the overall water availability of the habitat and the tolerance of plants to water deficit. This trait was scored using a 4° quasiordinal scale.

(*i*) plants capable of growing in arid and semiarid habitats with <300–400 mm mean annual precipitation; (*ii*) plants with optimum growth in mesic habitats with >400 mm mean annual precipitation; (*iii*) plants growing in wetlands and swampy habitats with permanent high levels of underground water or frequent flooding; (*iv*) plants always growing in water. In the statistical analyses, this variable was tested as either nominal or numeric.

Life forms. Three principal life forms were distinguished.

Annual: Short-lived herbs and grasses able to complete the lifecycle within one growing season or 1 y

Woody: Shrub or tree species having secondary growth (i.e., palms and bamboos but not bananas)

Perennial: Perennial herbs not fitting any of the previous categories

They were treated separately as presence/absence variables or one summarizing nominal variable life form.

Within the perennial category, three subcategories related to geophytism and nutrient storage syndromes were scored.

Geophytes_wide: A broad definition of geophytes that includes any plant with survival in unfavorable conditions that is promoted by the formation of underground storage organs (e.g., rhizome geophytes)

Geophytes_narrow: Geophytes in a narrow sense (i.e., tuber or bulb geophytes only)

Bulb geophytes: Bulb-forming geophytes that are usually related to the ability to temporally separate mitosis and cell division from cell and body expansion (16)

The world checklist of selected plant families (http://apps.kew.org/wcsp/home.do) was used as the source of the data to score these subcategories.

Pollination (variable pollination). Four different pollination modes were considered.

Anemogamous: Wind pollinated

Entomogamous: Insect pollinated

Zoogamous: Pollinated by vertebrates (i.e., birds, bats, and rats)

Hydrogamous: Water plants where pollen transport is by water

Only the dominant modes of pollination were scored for each species. These data were sourced from local floras and specialized scientific papers or supplied by experts.

Pollen morphology (variable desiccation-sensitive pollen). The data on pollen morphology and corresponding desiccation sensitivity were based on personal observations of pollen morphology and information sourced from palynological papers and databases assuming presence (orthodox pollen) or absence (desiccation-sensitive pollen) of morphological devices allowing pollen volume changes caused by dehydration before presentation and rehydration on the stigma (figure 1 in ref. 17 and figures 3 and 4 in ref. 18). This variable was treated as binary, with only clear cases scored as desiccation-sensitive; otherwise, species were scored as non-desiccation-sensitive (orthodox).

Chromosome numbers and ploidy-level data. Chromosome numbers for measured plants were extracted from available literature, namely the index to plant chromosome numbers (www.tropicos.org/Project/IPCN), the work by Fedorov (19), and the karyological literature listed in Dataset S1, Table S1. In case of controversy in the provided counts, various sources of information were considered to decide which one count most probably refers to the measured plants. These sources of data were (*i*) the origin of sampled and counted plants, (*ii*) the total number of records, and (*iii*) the comparison of the measured genome size with the ploidy-level data of the same or related species in the Plant DNA C-values database (http://data.kew.org/cvalues/; release 6.0, December of 2012). For 16 species, chromosomes were counted directly using the rapid squash method on root tips from cultivated plants or geminating seeds. Root tips were pretreated at room temperature in a saturated water solution of p-dichlorbenzene for 2 h and then fixed in a cold mixture of ethanol: acetic acid (3:1) for 24 h. The fixed material was treated immediately. The root tips were macerated in a mixture of ethanol: hydrochloric acid (1:1) for 2 min at room temperature. Temporary slides were made by squashing the cut and stained meristems in lactopropionic orcein (cf. 20).

Ploidy levels were derived from chromosome number data using the basic chromosome number estimates by Raven (21) or basic chromosome numbers opportunistically searched in available taxonomic databases and literature (not indicated in detail).

Chromosomal Structure Data (Holocentrism). The presence of a holocentric chromosome structure was based on the survey by Bureš et al. (22). We classified all species of Cyperaceae and Juncaceae as holocentric. We also scored *Prionium serratum* (Thurniaceae), which is sister to the Cyperaceae and Juncaceae clade, as holocentric, although this decision is based solely on its unusually small genome size and low GC content shared with Cyperaceae and Juncaceae, because there are currently no cytological data available for this family.

Calculation of Climate–Niche Parameters. Species distribution data were obtained from the GBIF portal (www.gbif.org; last accessed in March of 2011) and the South African National Floristic Database (23), which contains data for the southern African species (www.sanbi.org). All data were checked to remove entries for plants cultivated in botanical gardens and discard data considered to be unreliable because of (*i*) probable species misidentifications, (*ii*) species growth in nonnative conditions (i.e., artificial plantations), or (*iii*) suspected errors in the GBIF database (e.g., where a species entry includes an outlying geographical coordinate with no additional supporting data). When evident, we corrected errors arising from interchanges of latitude and longitude data and replaced some rounded coordinates placed in the sea to the nearest point on the coast (e.g., data from distribution mapping, where only centers of the quadrants are reported). For species with only limited georeferenced data available, the geographical coordinates of known localities were determined using Google Earth (www.google.com/earth/index.html).

Nineteen bioclimatic (BioClim) variables and altitude were extracted from the WorldClim global database of interpolated climate layers at 1 × 1-km spatial resolution (24) (www.worldclim.org/). The BioClim variables are derived from the monthly averages of mean, minimum, and maximum temperature and monthly total precipitations. A quarter is a period of any 3 consecutive months. The temperature variables are indicated in degrees Celsius, which are multiplied by 10. The meanings of 19 used BioClim variables are listed below.

BioClim 1: Annual mean temperature. This value is the mean of all of the monthly mean temperatures.

BioClim 2: Mean diurnal range. The mean of all of the monthly diurnal (daily) temperature ranges. Plants experiencing high diurnal temperature stress (such as in high-radiation, high-altitude, and desert environments) have to cope repeatedly with temperature stress on a daily basis and may be expected to have developed mechanisms to enable a fast response to changing temperatures.

BioClim 3: Isothermality. This value is an index calculated as the mean diurnal range (BioClim 2) multiplied by 100 and divided by the annual temperature range (BioClim 7). Isothermality informs about the extremity of the temperature regime in terms of comparing diurnal (daily) vs. annual (yearly) ranges. Organisms living in regions of high summer/winter or day/night oscillations are exposed to thermal stress. Those experiencing the diurnal stress (such as plants in high-radiation environments at high altitudes of tropical mountains or in deserts) have to cope with this stress repeatedly on a daily basis and have developed mechanisms to enable a fast response to changing temperature. Those organisms living in thermally seasonal (summer/winter) environments experience slow changes of temperature regimes, and therefore, their responses to temperature changes are expected to be slow.

BioClim 4: Temperature seasonality. This type of seasonality is expressed as the SD of the monthly mean temperatures. It informs whether a plant has to cope with long-term response to changing temperatures (e.g., enforcing winter tissue senescing or cold hardening). The temperature seasonality is generally high in winter cold temperate regions and low in the tropics.

BioClim 5: Average maximum temperature of warmest month. The highest of any monthly maximum temperature averages.

BioClim 6: Average minimum temperature of coldest month. The lowest of any monthly minimum temperature averages.

BioClim 7: Temperature annual range. The difference between the average maximum temperature of warmest month (BioClim 5) and the average minimum temperature of coldest month (BioClim 6).

BioClim 8: Mean temperature of wettest quarter. The wettest quarter of the year is determined (to the nearest month), and the mean temperature of this period is calculated.

BioClim 9: Mean temperature of driest quarter. The driest quarter of the year is determined (to the nearest month), and the mean temperature of this period is calculated.

BioClim 10: Mean temperature of warmest quarter. The warmest quarter of the year is determined (to the nearest month), and the mean temperature of this period is calculated.

BioClim 11: Mean temperature of coldest quarter. The coldest quarter of the year is determined (to the nearest month), and the mean temperature of this period is calculated.

BioClim 12: Annual precipitation. The sum of all of the monthly precipitation estimates.

BioClim 13: Precipitation of wettest month. This variable is self-explanatory.

BioClim 14: Precipitation of driest month. This variable is self-explanatory.

BioClim 15: Precipitation seasonality. Precipitation seasonality is the SD of the monthly precipitation estimates expressed as a percentage of the mean of those estimates (i.e., the annual mean).

It informs about temporal distribution of precipitation over a year. In regions experiencing drought vs. wet periods, the precipitation seasonality is higher; hence, plants living in these environments have developed traits to cope with temporal lack of water.

BioClim 16: Precipitation of wettest quarter. The wettest quarter of the year is determined (to the nearest month), and the total precipitation over this period is calculated.

BioClim 17: Precipitation of driest quarter. The driest quarter of the year is determined (to the nearest month), and the total precipitation over this period is calculated.

BioClim 18: Precipitation of warmest quarter. The warmest quarter of the year is determined (to the nearest month), and the total precipitation over this period is calculated.

BioClim 19: Precipitation of coldest quarter. The coldest quarter of the year is determined (to the nearest month), and the total precipitation over this period is calculated.

In addition to the median value of a climatic variable (indicating the general preference of a species), 10th, 25th, 75th, and 90th percentiles were also calculated to serve as surrogates for the extremes and limiting values of particular factors, which allow for partial control of the multitude of factors controlling species distribution. For example, when searching for differences in species sensitivity to temperature, two species may not differ in terms of mean annual temperatures observed across their distribution range, because this distribution is determined by multiple ecological and evolutionary factors (place of origin of their ancestor, presence of dispersal barriers, number and location of suitable habitats, absence of reliable distribution data from some regions, etc.) not considered in the model. However, maximum temperature stress at a few locations (where other factors could be considered to be less important) can reveal that only one of the species is tolerant to incidental freezing. Conventionally, this kind of multifactor control in data is analyzed using quantile regression (25, 26), and our analysis may be considered as one of its phylogenetic alternatives. Indeed, higher or lower quantiles of climatic data showed stronger explanatory power than their medians compared with GC content in the statistical analyses (Dataset S1, Tables S3 and S4).

Spatial Data Resampling Algorithm. Information on a species distribution is almost always spatially biased because of oversampling (collecting data from just one or a few often easily accessible areas) or undersampling, leading to the creation of gaps in data coverage in certain geographical regions. This effect may lead to distortions in the pattern inference and therefore, should be removed using a proper stratification approach so that a limited subset of data can be selected that reliably represents the overall distribution of a species. Because of the strong effect of local oversampling in many species in GBIF, a stratification algorithm was developed to control for this flaw using resampling of the species distribution data. The algorithm is based on heterogeneity-constrained random resampling, which was originally used for stratification of vegetation plots in vegetation databases (27). As an additional constraint, we defined the minimum distance between any current and a newly added locality (*min_dist*) that was set relative to the maximum distance between the two most distant localities in the nonstratified dataset.

The algorithm starts with the random selection of one location. In the next step, a new location is randomly selected from the remaining ones and added to the selection if its distance to any of the already selected locations in the subset is larger than *min_dist*. This procedure continues until all locations in the nonstratified dataset have been tested. This selection process is repeated (*no_cycles*) times. Mean pairwise distance and variance are calculated between all locations selected within each of the subsets. In the next step, the subsets are sorted by the decreasing mean value of all pairwise distances between selected locations and then sorted again by increasing variance of the location distances. The rank in mean distances indicates the extent to which the random selection includes outlying locations (a higher rank means that a greater number of outlying localities is selected), whereas the rank in variance refers to the regularity of spatial spread of localities (a minimum variance indicates regular spread, whereas a higher value indicates some spatial clustering). Ranks from both sorting runs are summed for each subset, and the subset with the lowest summed rank is selected as the most representative (i.e., the subset where locations are far from the others and regularly spatially spread) (Fig. S5). For stratification of localities in all species, we used consistently *no_cycles* = 1,000 and *min_dist* = 1/50.

Because of extreme differences between the number of locations for some species and the undesirable increase of computer time in the case of calculations with larger datasets, we prestratified the original species datasets if there were more than 2,000 locations (*no_locations*). We defined a latitude/longitude grid and selected a size of grid cells that would result in about 2,000 cells for the whole dataset. All locations occurring in one cell of the defined grid were considered to be duplicates, and only one of them was randomly selected for the next stratification. This approach allowed us to reduce the number of locations, while keeping the overall distribution pattern of the original nonstratified data unchanged. A script for performing the whole stratification procedure using R (28) is provided in Dataset S2.

Statistical Analyses. Ancestral state reconstructions of the GC content and genome size (log 2C values or log 1Cx values) were done by applying a generalized least squares approach using the ace module of the ape package (29) in R (28). Significant increases or decreases were detected by comparing the actual ancestral node values with a null model (i.e., with the random node values obtained with the same procedure calculated with randomly reshuffled tip labels on the phylogeny). The randomization was repeated 9,999 times. This procedure is implemented in the software Phylocom (http://phylodiversity.net/phylocom/) (30) or its R version [the package picante (31)], or it can be calculated for a phylogeny "phy" and a trait "trait" using our R script given in Dataset S4.

The presence of phylogenetic signal in GC content was tested by applying the Bloomberg's permutation test (cf. 32) based on 9,999 permutations. This test was calculated using the phylosignal function of the R package picante (31). The Pagel λ-coefficient of phylogenetic signal was calculated using the maximum likelihood procedure as captured in the fitContinuous function of the R package geiger (33). For the ancestral state and phylogenetic signal analyses, existing polytomies in *Agapanthus* and *Tetraria* were manually resolved, with close to zero (10^{-9}) branch lengths given to the newly created branches. The branching was designed to minimize divergence of the GC content along the resulting tree.

Tests for the regression relationships between the genomic parameters and both the life history traits and climatic niche data were performed using the phylogenetic generalized least squares procedure (pgls) using the pgls function of the caper package (34) in R. An optimum transformation of the initial Brownian motion-based covariance matrix was achieved by using a maximum likelihood procedure to search for the optimum multiplication parameter of internal branch lengths (λ) that allows simultaneous correction for the actual extent of the Brownian motion process and any artifact variation in measured parameters that may appear between closely related taxa because of measurement uncertainty. The searching was done initially for all possible models, including a single explanatory variable (e.g., GC ~ Holocentrics), and in the case of continuous variables, including both linear and quadratic forms [i.e., GC ~ log2C and GC ~ log2C + I(log2C^2), respectively]. Averaging all of the λ-values, we obtained a value close to 0.918 for the full tree and 0.882 for the reduced tree for testing the 1Cx monoploid genome size data. These values were always within the 5–95% confidence interval of optimum λ-values in all tested models, and they were used consistently as the fixed λ-parameters in all subsequent pgls analyses, allowing for full comparability of any tested explanatory

models (with either 2C or 1Cx values). The final multiple regression explanatory model was built up by manual forward selection of explanatory variables; in each step, the one with inclusion that caused the most significant decrease in the Akaike information criterion of the resulting model was selected (Table 1 and Dataset S1, Tables S3 and S4). By modeling, we did not consider character interactions and tested both linear and quadratic relationships in the case of continuous characters. The log-normally distributed data were transformed with decadic [$x' = \log(x)$; genome sizes] or natural [$x' = \ln(x + 10)$; rainfall BioClim data and altitude] logarithm to conform to the criteria of normality before any statistical analyses.

1. Barow M, Meister A (2002) Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. *Cytometry* 47(1):1–7.
2. Šmarda P, Bureš P, Horová L, Foggi B, Rossi G (2008) Genome size and GC content evolution of *Festuca*: Ancestral expansion and subsequent reduction. *Ann Bot (Lond)* 101(3):421–433.
3. Šmarda P, Bureš P, Šmerda J, Horová L (2012) Measurements of genomic GC content in plant genomes with flow cytometry: A test for reliability. *New Phytol* 193(2):513–521.
4. Pfosser M, Amon A, Lelley T, Heberle-Bors E (1995) Evaluation of sensitivity of flow cytometry in detecting aneuploidy in wheat using disomic and ditelosomic wheat-rye addition lines. *Cytometry* 21(4):387–393.
5. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436(7052):793–800.
6. 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(4): 625–631.
7. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431(7011):931–945.
8. Doležel J, Greilhuber J (2010) Nuclear genome size: Are we getting closer? *Cytometry A* 77(7):635–642.
9. Zanne AE, et al. (2014) Three keys to the radiation of angiosperms into freezing environments. *Nature* 506(7486):89–92.
10. Grass Phylogeny Working Group II (2012) New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytol* 193(2):304–312.
11. Stromberg CAE (2011) Evolution of grasses and grassland ecosystems. *Annu Rev Earth Planet Sci* 39:517–544.
12. Edwards EJ, et al.; C4 Grasses Consortium (2010) The origins of C_4 grasslands: Integrating evolutionary and ecosystem science. *Science* 328(5978):587–591.
13. Crepet WL, Feldman GD (1991) The earliest remains of grasses in the fossil record. *Am J Bot* 78(7):1010–1014.
14. Bouchenak-Kdelladi Y, Verboom GH, Savolainen V, Hodkinson TR (2010) Biogeography of the grasses (Poaceae): A phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Bot J Linn Soc* 162(4):543–557.
15. Paradis E (2013) Molecular dating of phylogenies by likelihood methods: A comparison of models and a new information criterion. *Mol Phylogenet Evol* 67(2): 436–444.
16. Grime JP, Mowforth MA (1982) Variation in genome size – an ecological interpretation. *Nature* 299(5879):151–153.
17. Franchi GG, Nepi M, Dafni A, Pacini E (2002) Partially hydrated pollen: Taxonomic distribution, ecological and evolutionary significance. *Plant Syst Evol* 234(1-4): 211–227.
18. Franchi GG, et al. (2011) Pollen and seed desiccation tolerance in relation to degree of developmental arrest, dispersal, and survival. *J Exp Bot* 62(15):5267–5281.
19. Fedorov AA, ed (1969) *Chromosome Numbers of Flowering Plants* (Komarov Botanical Institute, Academy of Sciences, Leningrad, Russia).
20. Rotreklová O, et al. (2011) Chromosome numbers of Carex. *Preslia* 83(1):25–58.
21. Raven PH (1975) The bases of angiosperm phylogeny: Cytology. *Ann Mo Bot Gard* 62(3):724–764.
22. Bureš P, Zedek F, Marková M (2013) Holocentric chromosomes. *Plant Genome Diversity*, eds Leitch IJ, Greilhuber J, Doležel J, Wendel J (Springer, Vienna), Vol 2, pp 187–208.
23. Magill RE, Russell GE, Morris JW, Gonsalves P (1983) PRECIS, the Botanical Research Institute herbarium data bank. *Bothalia* 14(3-4):481–495.
24. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25(15):1965–1978.
25. Koenker R, Hallock KF (2001) Quantile regression. *J Econ Perspect* 15(4):143–156.
26. Cade BS, Noon BR (2003) A gentle introduction to quantile regression for ecologist. *Front Ecol Environ* 1(8):412–420.
27. Lengyel A, Chytrý M, Tichý L (2011) Heterogeneity-constrained random resampling of phytosociological databases. *J Veg Sci* 22(1):175–183.
28. R Development Core Team (2012) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
29. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290.
30. Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: Software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* 24(18):2098–2100.
31. Kembel SW, et al. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26(11):1463–1464.
32. Münkemüller T, et al. (2012) How to measure and test phylogenetic signal. *Methods Ecol Evol* 3(4):743–756.
33. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: Investigating evolutionary radiations. *Bioinformatics* 24(1):129–131.
34. Orme D (2012) *The Caper Package: Comparative Analysis of Phylogenetics and Evolution in R*. Available at http://cran.r-project.org/web/packages/caper/vignettes/caper.pdf. Accessed March 23, 2013.

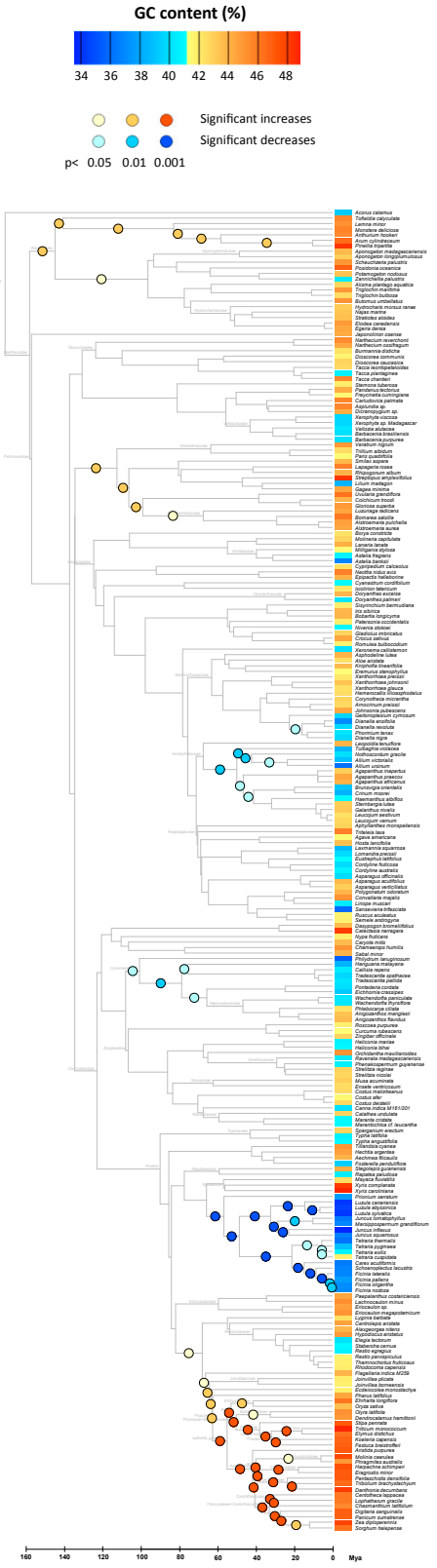

Fig. S1. Phylogenetic tree of monocots showing taxa names, selected node names, measured GC contents and branches where significant increases and decreases of GC contents appeared.

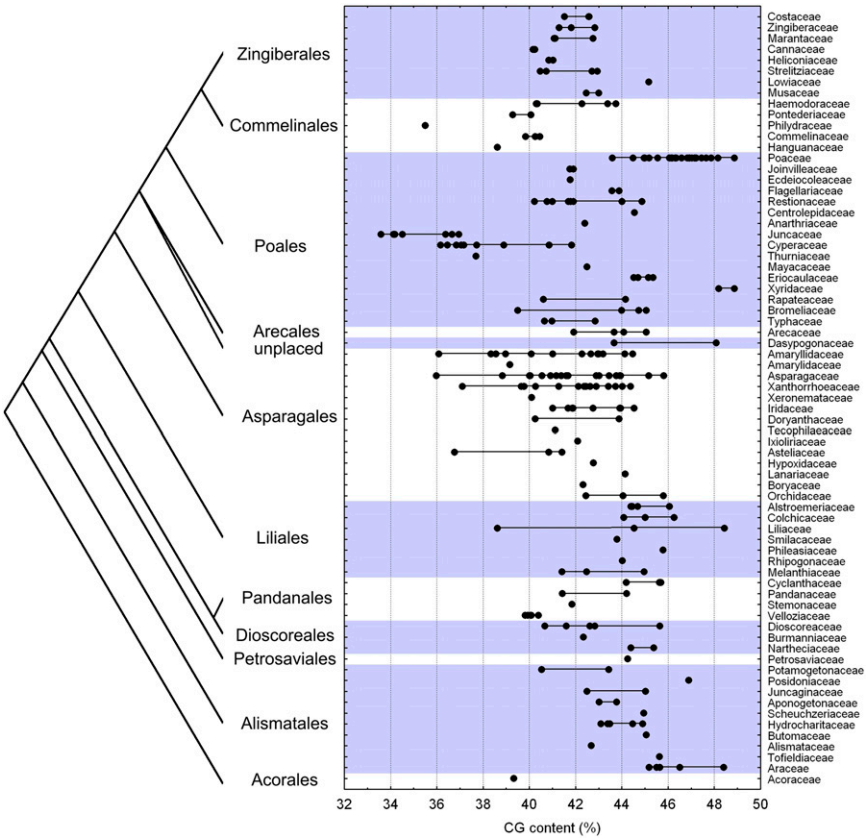

Fig. S2. Variation of GC content in monocot families. Points represent individual measured taxa. Different background colors highlight different orders or unplaced families.

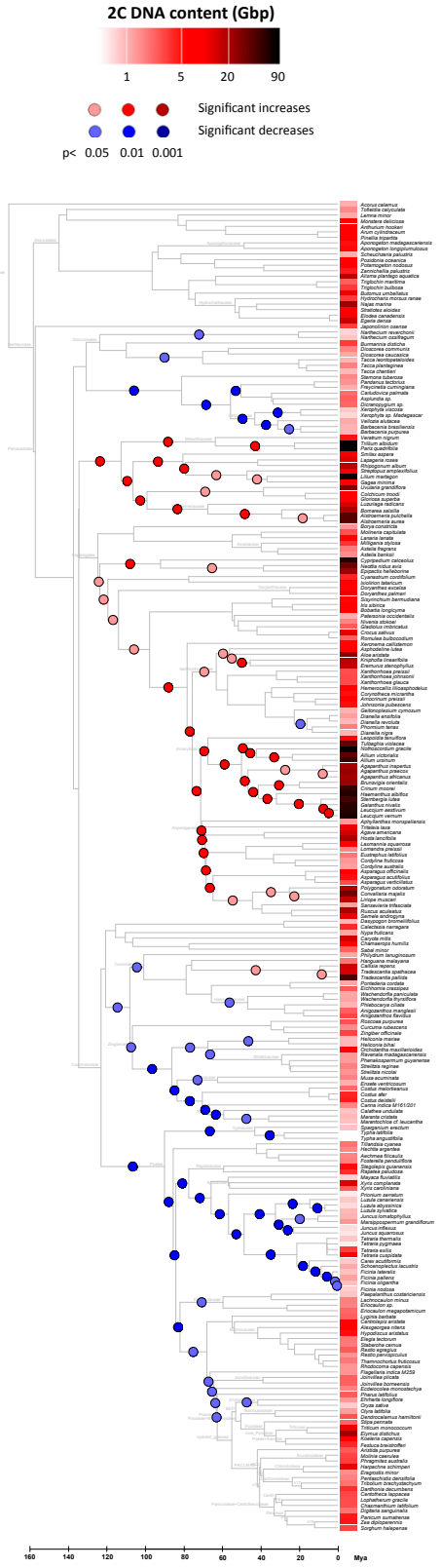

Fig. S3. Phylogenetic tree of monocots showing taxa names, selected node names, measured 2C DNA contents, and branches where significant increases and decreases of 2C DNA contents appeared.

Fig. S4. Phylogenetic tree of monocots showing taxa names, selected node names, measured 1Cx monoploid genome sizes, and branches where significant increases and decreases of 1Cx monoploid genome size appeared.

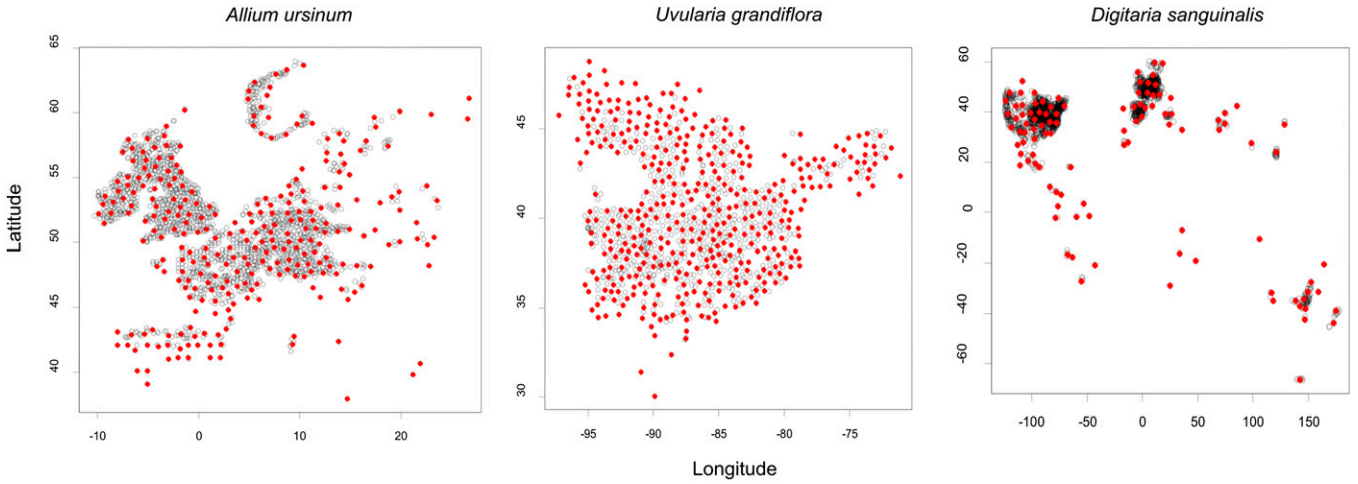

Fig. S5. Examples of the stratification algorithm function (*no_cycle s* = 1.000; *min_dist* = 1/50; *no_locations* = 2.000) on worldwide distribution data for three species with distinct distribution patterns and natural range sizes: *Allium ursinum* (Europe), *Uvularia grandiflora* (North America), and *Digitaria sanguinalis* (cosmopolitan). Empty circles, all locations available in GBIF; red circles, locations selected with the stratification algorithm and used for additional analyses.

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (TXT)
Dataset S3 (TXT)
Dataset S4 (TXT)