Article

Seedling root system adaptation to water availability during maize domestication and global expansion

Peng Yu [1,2,21] ✉, Chunhui Li[3,21], Meng Li [4,21], Xiaoming He[1,2,21],
Danning Wang [1,2], Hongjie Li[1,2], Caroline Marcon[1], Yu Li[3], Sergio Perez-Limón[4],
Xinping Chen [5], Manuel Delgado-Baquerizo [6,7], Robert Koller[8], Ralf Metzner[8],
Dagmar van Dusschoten[8], Daniel Pflugfelder[8], Ljudmilla Borisjuk[9],
Iaroslav Plutenko [9], Audrey Mahon[10], Marcio F. R. Resende Jr.[10],
Silvio Salvi [11], Asegidew Akale[12], Mohanned Abdalla[12], Mutez Ali Ahmed[12],
Felix Maximilian Bauer [13], Andrea Schnepf [13], Guillaume Lobet[13,14],
Adrien Heymans [14], Kiran Suresh [15], Lukas Schreiber[15],
Chloee M. McLaughlin [16], Chunjian Li[17], Manfred Mayer[18],
Chris-Carolin Schön [18], Vivian Bernau [19], Nicolaus von Wirén [20],
Ruairidh J. H. Sawers [4] ✉, Tianyu Wang [3] ✉ & Frank Hochholdinger [1] ✉

The maize root system has been reshaped by indirect selection during global adaptation to new agricultural environments. In this study, we characterized the root systems of more than 9,000 global maize accessions and its wild relatives, defining the geographical signature and genomic basis of variation in seminal root number. We demonstrate that seminal root number has increased during maize domestication followed by a decrease in response to limited water availability in locally adapted varieties. By combining environmental and phenotypic association analyses with linkage mapping, we identified genes linking environmental variation and seminal root number. Functional characterization of the transcription factor *ZmHb77* and in silico root modeling provides evidence that reshaping root system architecture by reducing the number of seminal roots and promoting lateral root density is beneficial for the resilience of maize seedlings to drought.

The spread of crops from their ancestral habitats and expansion of cultivation was accompanied by substantial phenotypic changes, driven by a combination of direct farmer selection and environmental adaptation[1]. Maize (*Zea mays* ssp. *mays*) was initially domesticated in southwest Mexico approximately 9,000 years ago from the wild lowland teosinte *Zea mays* ssp. *parviglumis*, with subsequent admixture with the highland teosinte *Zea mays* ssp. *mexicana* contributing substantially to modern populations[2,3]. Following domestication from *Zea mays* spp. *parviglumis*, maize spread to the highlands of Mexico

and South America[4] (Fig. 1a). Subsequent adaptation to temperate climates allowed the expansion of maize from the tropics to diverse environments around the globe[5,6]. Root system function is instrumental in colonizing new habitats[7] and acquiring resources; in particular, water and nutrients in natural soils of different geographical origins[8]. During domestication and diversification, the maize root system has become more complex by acquiring the capacity to form seminal roots, a feature largely absent in the maize progenitor teosinte[9,10]. In maize seedlings, the number of seminal roots determines the overall structure of the
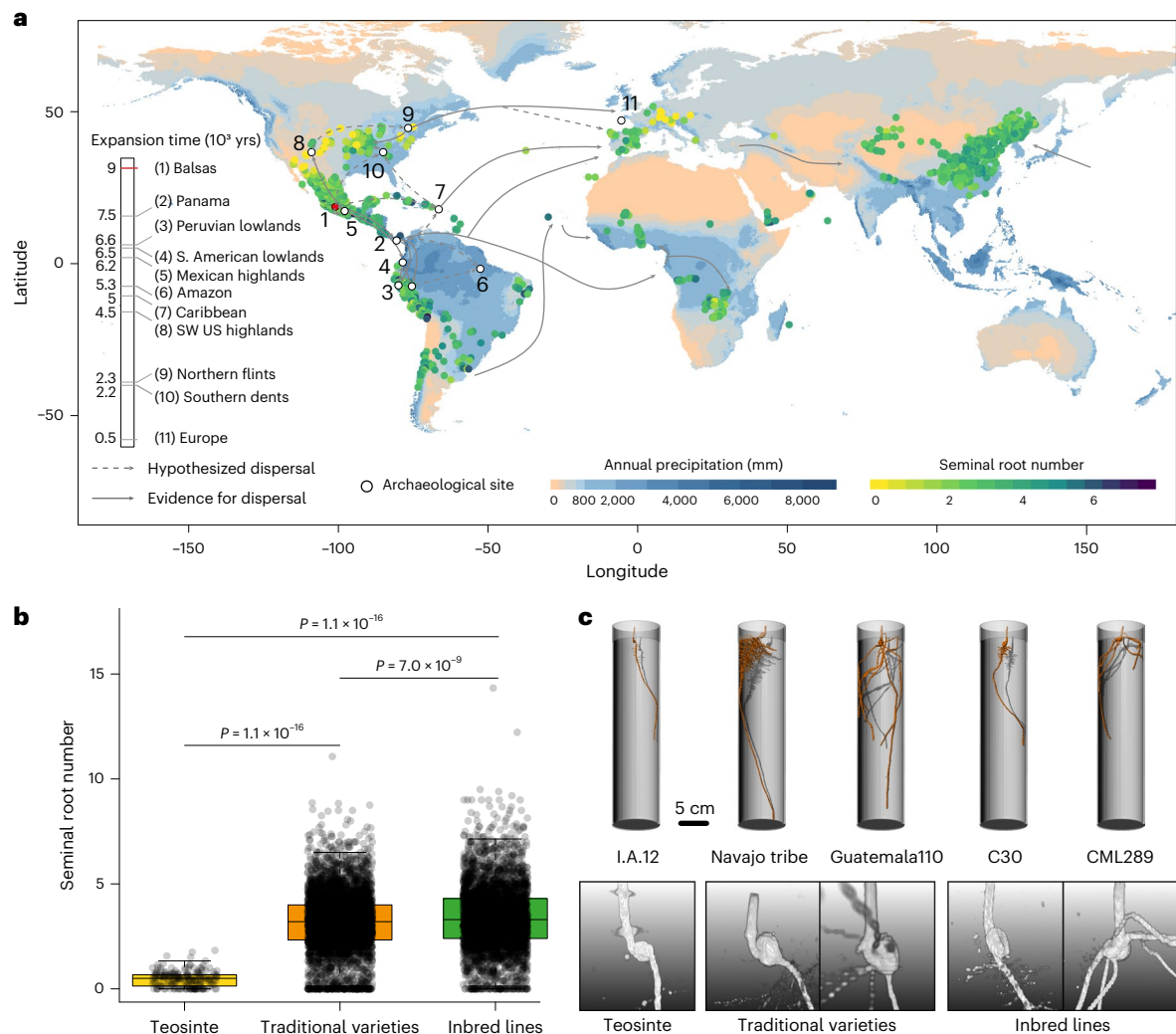
Fig. 1 | Maize evolutionary history resolves global organization of SRN.
a, Geographical variability of SRN in traditional varieties of maize (*n* = 2,424).
SRN was determined in globally collected traditional varieties of indicated
geographical origin. Domestication and expansion times[4] for maize populations
are indicated accordingly. The global map of average annual precipitation
between 1991 and 2020 is derived from NOAA's Climate.gov (https://www.
climate.gov) and WorldClim prec 30 s (https://www.worldclim.org/data/
worldclim21.html#google_vignette). The map was produced in R (v.4.2.2) with
the packages 'maps' and 'raster'. **b**, Seminal root differentiation across the genus
Zea including teosinte (*n* = 173), traditional varieties (*n* = 4,868) and modern
inbred lines (*n* = 4,049). Each dot indicates the average SRN of each investigated

accession. *n* = 10 biologically independent seedlings per inbred line and *n* = 20
biologically independent seedlings per traditional variety. Boxes span from the
first to the third quartile, lines represent the median and whiskers include data
within the 1.5× interquartile range of the lower and upper quartiles. Data points
outside of whiskers represent outliers. Significant differences among groups
are indicated by exact values (one-way ANOVA, Tukey's HSD). **c**, Reconstruction
of root system architecture and initiation sites of seminal roots by non-invasive
magnetic resonance imaging (MRI) in natural soil. Teosinte, I.A.12 (Ames 21793);
traditional varieties: Navajo tribe (PI 311229), Guatemala 110 (PI 490825); modern
inbred lines: C30 (Ames 26815), CML289 (Ames 32336).

root system and thereby the depth and soil volume that the roots can
explore[11–13]. Seminal roots are formed endogenously in the embryo
between 22 and 40 days after pollination[14]. They are beneficial for nitro-
gen and phosphorus acquisition during maize seedling development[13]
and can persist and remain functional during the whole life cycle of
the maize plant[14]. However, the question of how the maize root system
adapted its form and function during domestication and global expan-
sion remains elusive. Understanding the genetic basis, environmental
drivers and the potential adaptive value of seminal root number (SRN)
variation to changing environments is essential to develop crops that
are resilient to future climatic challenges.

Results

Variation of SRN follows maize domestication

We investigated the environmental and genetic factors driving diver-
sity in SRN in the genus *Zea*. We quantified SRN in a set of >9,000

Zea accessions representing worldwide diversity across the major
maize-cultivating regions in the Americas, Europe, Asia and Africa. Our
collection included 173 wild teosinte accessions, 4,868 traditional varie-
ties (also known as landraces) and 4,049 modern inbred lines. Under
controlled conditions, maize varieties produced up to 11 seminal roots
in traditional varieties (Supplementary Table 1) and up to 14 seminal
roots in modern inbred lines (Supplementary Table 2). Overall, maize
varieties formed on average 3.3 seminal roots, while teosinte acces-
sions (Fig. 1b,c) did not produce any seminal roots in 23% of the acces-
sions (*n* = 173) (Supplementary Fig. 1a and Supplementary Table 3).
Interestingly, although SRN was low across all teosinte accessions,
highland teosinte (*Zea mays* ssp. *mexicana*) produced significantly
more seminal roots than the lowland teosinte *Zea mays* spp. *parvi-
glumis* (Supplementary Fig. 1b). These data are consistent with the
previously advanced hypothesis that seminal root formation in *Zea* is
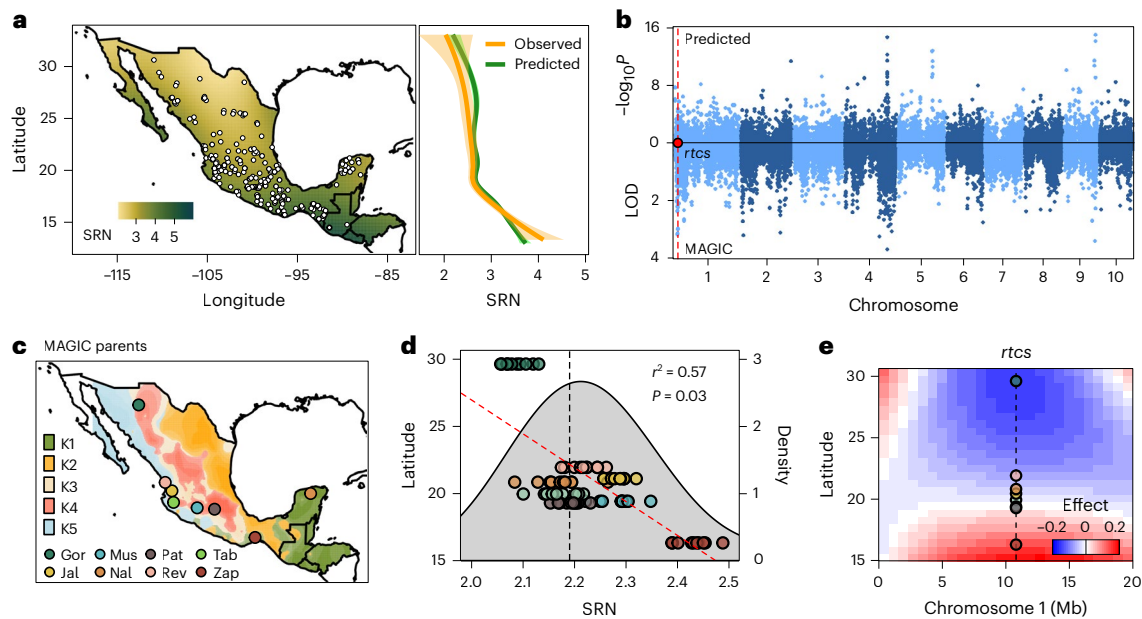a domestication trait[10,15].

Fig. 2 | Geographical and genomic signatures of SRN variation in Mexico.
a, SRN decreases along a latitudinal gradient from south to north in Mexico. White dots indicate the locations of sampled native maize traditional varieties. Green saturation indicates increasing predicted SRN based on a random forest model. The marginal plot depicts the calculated means of observed and predicted SRN over latitude. The map was produced in R (v.4.2.2) with the packages 'raster' and 'ggplot2'. **b**, Genomic loci of SRN variation. Miami plot shows GWA of SNP predicted trait values above the *x* axis and with measured traits from a MAGIC population below the axis. Alternating colors indicate the ten maize chromosomes. Significant *P* values were controlled by a one-sided *F*-test (null hypothesis). The *rtcs* gene was labeled accordingly. **c**, Source of the eight founders of the MAGIC population. Shading on the map corresponds to an ancestry coefficient (K = 5) based on a broader genotyped panel. LOD, logarithm of odds. MAGIC founder traditional variety abbreviations: Gor, Gordo; Mus, Mushito; Pat, Palomero Toluqueno; Tab, Tabloniclo; Jal, Jala;

Nal, Nal Tel; Rev, Reventador; Zap, Zap Chico. The ancestry coefficients were calculated using the 'tess3r' package and the map was produced using packages 'maps' and 'raster' in R (v.4.2.2). **d**, Genome-wide effects in the MAGIC population support a latitudinal dependency of SRN. Dots show predicted SRN for each of the eight founder parents based on a genome-wide model generated from the derived MAGIC families. Multiple points for each founder indicate the result of dropping each chromosome in turn from the model. The trend line and correlation are based on the complete model with a linear fit of latitude on predicted SRN by two-sided statistics (null hypothesis) using all ten chromosomes. The gray curve shows the frequency density of the whole population with the vertical line at the mean. **e**, MAGIC founder allele effects in a 20 Mb window around *rtcs*. Polynomial fit of marker effects against source latitude for the eight alleles segregating in the MAGIC population. The vertical dashed line indicates the position of *rtcs*.

Recently, it has been suggested that the increase in seed size during domestication was a prerequisite for seminal root formation[13]. In our study, SRN was only weakly correlated with seed size or the proportion of the embryo to the whole seed area across 2,429 modern inbred lines (Supplementary Fig. 2) and showed no relationship with embryo volume in a panel of diverse US traditional varieties (Supplementary Fig. 3). Thus, it is likely that the formation of seminal roots is independent of the process of seed selection during breeding. However, we cannot rule out the possibility that other factors from the seed may have an effect on the SRN. To further investigate the relationship between seed traits and SRN, we evaluated an additional collection of 663 modern US inbred lines (Supplementary Table 4) and 975 globally distributed traditional varieties (Supplementary Table 5). We found that sweet corn, flour corn and popcorn varieties characterized by specific carbohydrate compositions formed fewer seminal roots than other varieties in modern inbred lines (Supplementary Fig. 4a) or traditional varieties (Supplementary Fig. 4b). Analysis of near-isogenic lines with mutants that alter the sugar composition of the endosperm (*sugary1*, *shrunken2*) demonstrated that seminal root formation is independent of the amount of carbohydrates available during seed development (Supplementary Fig. 5). Therefore, we reasoned that the increase in SRN was part of domestication during the global expansion of maize but was independent of seed traits in maize, which have been strongly modified by human selection and breeding. It should be noted that sowing depth might have a potential effect on the variation of SRN because modern varieties are usually planted closer to the soil surface than teosinte.

Geographical and genomic signals of seminal root variation
To determine whether and how SRN varies with the environment, we applied machine learning to investigate the most important climatic and soil factors associated with SRN across 1,484 georeferenced traditional varieties sourced from diverse climatic and soil conditions (Supplementary Table 6). Traditional varieties that originated from arid regions had fewer seminal roots than those of other origins (Supplementary Fig. 6a). Using random forest modeling, we found that mean diurnal temperature range (Pearson's $r = -0.36$, $P < 0.001$), temperature seasonality (Pearson's $r = -0.29$, $P < 0.001$) and precipitation seasonality (Pearson's $r = -0.07$, $P = 0.010$) were the best environmental predictors of SRN followed by soil organic carbon (Pearson's $r = 0.11$, $P < 0.001$) and sand content (Pearson's $r = -0.16$, $P < 0.001$) (Supplementary Fig. 6b,c). High mean diurnal temperature range and precipitation seasonality are important meteorological indicators associated with extreme climates such as deserts. Importantly, we further showed that paleoclimatic levels of precipitation in the mid-Holocene (ca. 6,000 years ago) was a significant predictor of SRN (Supplementary Fig. 6d; Pearson's $r = 0.30$, $P < 0.001$), highlighting the importance of rainfall level in the evolutionary patterns of the maize root system. To better understand the relationship between SRN and the environment, we combined selected environmental variables into a second predictive random forest model. Focusing specifically on Mexican maize, we identified a broad trend of decreasing SRN with increasing latitude (Fig. 2a). We next used our trained model to predict SRN for an additional panel of 1,781 previously genotyped and georeferenced Mexican varieties[5] (Fig. 2a). Using the available genotypes and our predicted SRN values, we performed a

genome-wide association study (GWAS; Fig. 2b), identifying genomic loci linked to the combinations of environmental variables that themselves described SRN variation in our training set.

To phenotypically map SRN in Mexican maize, we generated and evaluated an eight-parent Multi-parent Advanced Generation Inter-Cross (MAGIC) population from founders that spanned the previously observed latitudinal cline in SRN (Fig. 2c). Comparison of the results of predicted trait GWAS and MAGIC mapping identified several shared genomic regions, including a locus on chromosome 1 linked to the previously described gene *rootless concerning crown and seminal roots*[16] (*rtcs*; Fig. 2b). The MAGIC population partially breaks down the population structure that can confound studies of local adaptation. On this basis, we used the MAGIC families to generate a genome-wide predictive model for SRN and then applied this model to the eight founder haplotypes. Interestingly, our model recovered the latitudinal trend in SRN that we have observed in our broader sampling (Fig. 2d). This result was robust to the removal of any single chromosome from the model, indicating that effects throughout the genome were contributing to the clinal trend, consistent with persistent directional selection and local adaptation. We examined the region of the genome around *rtcs* more closely by modeling separate allele effects for each of the eight founders, recovering evidence of an allelic series with effects ranging from positive to negative following the founder source from south to north (Fig. 2e). Thus, our ecological and genomic models suggest SRN variation is probably shaped by indirect selection for adaptation to new environments.

Northern Flint alleles drive seminal root differentiation

Previous population genetic analyses have described the expansion of maize out of northwestern Mexico and its subsequent adaptation to the dry environment of the southwestern US (Arizona and New Mexico)[17,18]. In our study, accessions sampled from the southwestern US had remarkably low SRNs (Fig. 3a). In fact, more than 57% (53 out of 92) of southwestern US accessions completely lacked seminal roots (Fig. 3b and Supplementary Table 7). Such seminal root defective phenotypes from the southwestern US were more drastic than those of the investigated teosinte lines (Supplementary Fig. 1a and Supplementary Table 1). Interestingly, we observed such low SRNs exclusively in the United States, Canada and some European countries (Supplementary Table 1), which associates with a higher share of Northern Flint, a group derived from the US Southwest[19–21]. Using a maximum-likelihood estimation, we evaluated the effect of the Northern Flint germplasm on SRN across our sampling. We found that the proportion of alleles derived from the Northern Flint germplasm negatively correlated with SRN in both the US (Fig. 3c) and modern European inbred lines (Fig. 3d).

SRN was not significantly correlated with proportions of germplasm derived from Tropical highlands, Tropical lowlands or Southern dent (Supplementary Fig. 7). We next genotyped 778 geographically diverse US traditional varieties and confirmed that the proportion of introgressed Northern Flint germplasm correlated negatively with SRN (Fig. 3e). We also evaluated a collection of introgression lines carrying genomic regions of the typical Northern Flint traditional variety Gaspé Flint[22]. The introgression lines with a higher share of the Northern Flint genome formed fewer seminal roots than the other panels evaluated (Supplementary Fig. 8). Overall, these phenotypic and genetic analyses indicate that alleles derived from the Northern Flint germplasm of southwestern US origin are an important factor determining SRN during the local adaptation of maize to different environments.

Seminal root variation contributes to root functional traits

To determine the potential adaptive importance of SRN across different environments, we used in silico root models to determine the impact of SRN in the context of whole root system architecture using 218 representative US maize traditional varieties (Supplementary Table 8). We first evaluated root architectural and morphological traits using a rhizobox system[23] to parameterize the structural–functional model *CPlantBox*[24]. The simulations illustrate that SRN negatively correlates with seedling primary root length and lateral root density along the primary root throughout the whole root system (Fig. 4a). We found that variation in SRN impacts seedling vigor by modulating the overall root system conductance (K_{rs}) (Fig. 4b). To explore whether changes in SRN will reshape root system architecture under realistic soil conditions, we used magnetic resonance imaging and positron emission tomography (MRI–PET) to compare the maize seminal rootless mutant *rtcs* to an isogenic wild-type line that produced an average of three seminal roots (Fig. 4c). In the absence of seminal roots, the *rtcs* mutant produced an increased number of lateral roots. Water uptake in young maize has previously been shown to be dominated by lateral roots[25], suggesting that reducing SRN to favor lateral root production may have an adaptive advantage for seedling establishment in water-limited conditions. We further characterized a specific southwestern US traditional variety (Navajo tribe) that we had identified to produce no seminal roots but a significantly enhanced number of lateral roots (Fig. 1c and Supplementary Fig. 9). Thus, variation in SRN might drive the overall dimension and branching of the whole root system, which will potentially determine the plant's capacity to capture water. We next used the *CPlantBox* realizations for each of the 218 traditional varieties to determine their standard uptake fraction and demonstrated that the potentially relative contribution of lateral roots to total root water uptake decreases with increasing SRN (Fig. 4d). Based on these

Fig. 3 | Variation in SRN coincides with proportional origin from Northern Flint maize sources. a, Patterns of water availability and seminal root differentiation across the USA. The annual average precipitation (1991–2020) map is derived from NOAA's Climate.gov (https://www.climate.cov) and WorldClim prec 30 s (https://www.worldclim.org/data/worldclim21.html#google_vignette). The map was produced in R (v.4.2.2) with the packages 'maps' and 'raster'. The size of the pie charts indicates the number of sampled traditional variety accessions while colored areas denote the proportions of SRN classes. **b**, Violin plots show SRN variation in traditional varieties originating from different geographical regions of the US. *n* = 20 biologically independent seedlings per traditional variety. The traditional variety accessions were contributed by the North Central Regional Plant Introduction Station and the International Center for Maize and Wheat Improvement. The geographical information of groups of traditional varieties derives from the narrative information of the US National Plant Germplasm System (https://npgsweb.ars-grin.gov/gringlobal). Sample sizes are Southwest US (*n* = 134), West US (*n* = 95), Midwest US (*n* = 348), Northeast US (*n* = 73), Southeast US (*n* = 259). A two-sided test of significance with Bonferroni correction was performed to adjust the *P* value for the multiple independent tests among regional pools.

Boxes span from the first to the third quartiles, center dots represent median values and whiskers extend 1.5× the interquartile range of the lower and upper quartiles. **c–d**, Pearson's correlation between SRN and the proportion of Northern Flint sources in the US Ames panel (**c**) and the European collection (**d**). Estimates of historical sources for individual Ames modern inbred lines and modern European inbred lines are extracted from published studies[43,44]. Here, the proportion of Northern Flint sources was correlated with SRN across modern maize inbred lines. The *P* value denotes the probability at which the correlation coefficient (solid line) is zero (null hypothesis) at a 95% confidence interval (shaded area). SS, stiff-stalk; NSS, non-stiff stalk; TS, tropical/sub-tropical; Mixed, mixture of these different germplasms. **e**, Pearson's correlation between SRN and the proportion of Northern Flint germplasm sources in US traditional varieties. The reference Northern Flint-sourced traditional varieties were defined accordingly[19]. Scatterplots show combined SRN data of traditional varieties from different geographical origins with best fit (solid line) and 95% confidence interval (gray shading) for linear regression ($P = 5.4 \times 10^{-109}$, *n* = 778). Different colors of dots correspond to different geographical origins of investigated traditional varieties.

modeling results, variation in SRN might determine the overall absorptive surface by impacting lateral root formation.

We selected 66 representative traditional varieties (Supplementary Table 8) from the panel of 218 and experimentally measured transpiration rates in wet soil, finding no significant difference between groups (Supplementary Fig. 10). We then used a soil–plant hydraulic model and determined that the stress onset limit (that is, the point at which a small increase in transpiration provokes a large drop in leaf water potential at a given soil water potential) occurred at less negative leaf water potential in the traditional varieties with lower SRNs (Fig. 4e).

Indeed, maize traditional varieties with one seminal root require higher water flow rates per unit root length than traditional varieties with five seminal roots, which induces a local drop in soil water potential and exhibits an earlier stomatal closure[26,27]. This allows the varieties to sustain similar transpiration rates. We propose that such adaptive stomatal behavior leading to lower transpiration is beneficial for seedling maize that is subject to water stress. In addition, salt-simulated drought conditions tend to increase the lignin accumulation along the tip of the primary root (Supplementary Fig. 11). Interestingly, traditional varieties with fewer seminal roots tend to respond more dramatically than
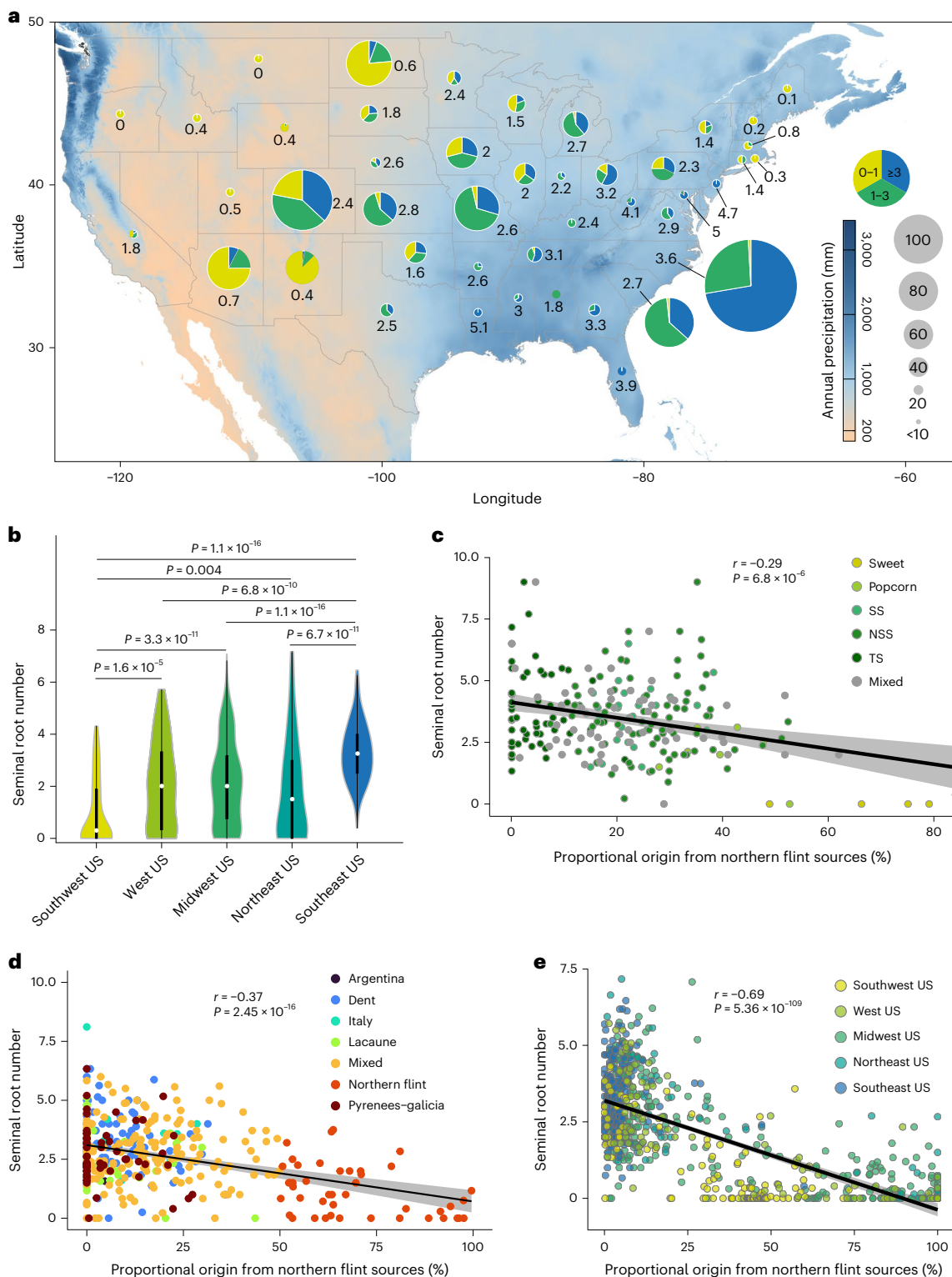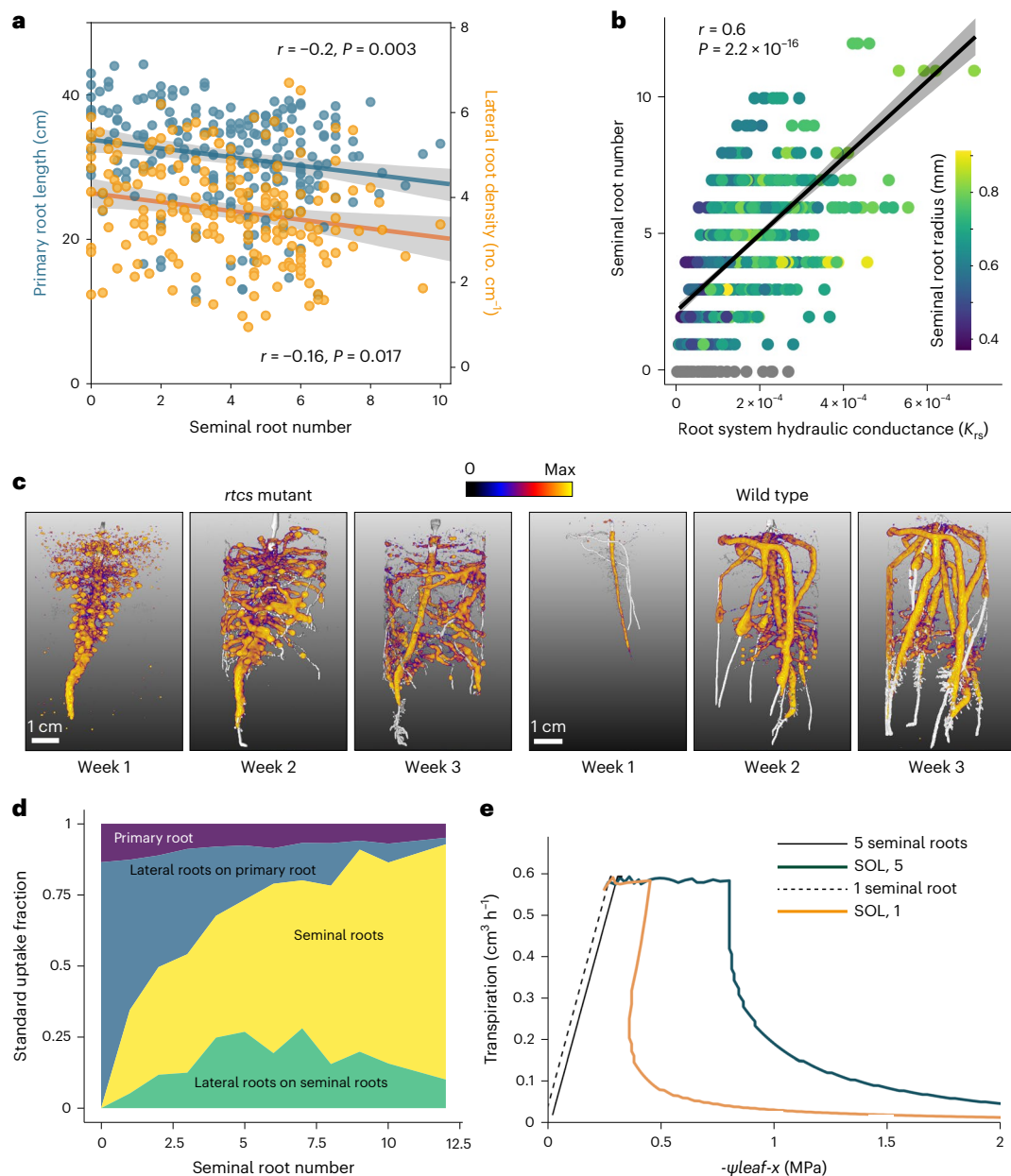
Fig. 4 | Variation in SRN drives overall root architectural and hydraulic properties. a, SRN is negatively correlated with rooting depth of the primary root and lateral root density in different maize traditional variety accessions grown in a rhizobox system. **b**, Seminal root variation affects overall root hydraulic properties. K_{rs} is based on 2D images of root systems grown in the rhizobox and simulated root architecture by structural–functional modeling. In both **a** and **b**, scatterplots show combined seminal root data of traditional varieties grown in the rhizobox and Pearson's regression with best fit (solid line) two-sided alternative hypothesis and 95% confidence interval (shaded area) ($n = 218$). **c**, Seminal root defects of the *rtcs* mutant cause highly branched lateral roots emerging from the primary root. Reconstruction of root architecture and carbon allocation by MRI combined with PET. Intensity of carbon deposition by radiolabeled ^{11}C is visualized by color code. Note that when ^{11}C was supplied to leaves for the first time, the first two seminal roots were already formed.

As MRI images were taken after the PET images, growing root tips are not in the same position. **d**, Standard uptake fraction of seminal roots and lateral roots as a function of SRN. For each SRN, the average proportion of water uptake per root type is expressed as a ratio relative to overall water uptake. The relative contribution to water uptake is considered separately for the primary root, lateral roots initiated from the primary root, total seminal roots and lateral roots initiated from seminal roots. Note that some of the traditional varieties with lower SRNs already formed very short crown roots but their contribution to water uptake is not considered. **e**, Simulation of transpiration rates of representative traditional varieties ($n = 76$) from a subset of 218 traditional varieties. A maize traditional variety with one seminal root requires larger gradients in soil water potential than a traditional variety with five seminal roots to sustain the same transpiration rate. Hence, stress onset limit (SOL) occurs at a lower negative leaf water potential for plants with lower SRN.

those with more seminal roots, especially under water stress conditions (Supplementary Fig. 11), which facilitates root penetration through dry soil. Thus, seminal root variation might contribute to the optimization of root architectural, hydraulic and physiological changes for improved plant tolerance to limited water availability.

ZmHb77 regulates root system architecture and drought resilience

To understand the genetic basis of variation in SRN in inbred maize, we performed GWAS using an association panel of 1,604 diverse modern inbred lines that mainly originated from the US, China and Europe and

that encompass the maize heterotic groups used in the US and China[28]. We observed substantial variation in SRN, with values ranging from 0–12 and an average of three (Supplementary Table 3). We detected a total of 160 associated single-nucleotide polymorphisms (SNPs) ($P = 1.0 \times 10^{-5}$), corresponding to 160 candidate genes underlying SRN (Fig. 5a and Supplementary Table 9). Among these candidate genes, we identified *rtcs*, which is known to regulate SRN in maize[16]. We next screened for novel mutants of these candidate genes in the *BonnMu* reverse genetics resource of maize[29] and identified transposon insertions in five distinct genes that resulted in reduced SRN (Supplementary Fig. 12 and Supplementary Table 10). Among those five genes, one gene, *Zm00001d045398* on chromosome 9, was annotated as *Homeobox-transcription factor 77* (ref. 30) (*ZmHb77*). To further validate the function of *ZmHb77* in regulating root development, we generated two independent CRISPR–Cas9 knockout lines (KO1 and KO3) (Fig. 5b). Both mutant alleles KO1 and KO3 conditioned a significant reduction in SRN (Fig. 5c,d) coupled with an increase in lateral root density (Fig. 5e,f), suggesting that this gene has a role in reshaping seedling root architecture by regulating SRN and lateral root density. We then carried out a soil cultivation box experiment with mutant and wild-type plants under well-watered and drought conditions followed by re-watering. The mutants showed a significant advantage regarding growth and photosynthesis rate under both drought and re-watering conditions, although there were no visible differences under well-watered conditions (Fig. 5g–j). Interestingly, mutants with fewer seminal roots but more lateral roots were more tolerant to drought and had a higher survival rate than wild-type plants after re-watering, while we observed no differences between mutants and wild type under well-watered conditions (Fig. 5i,k). These results support the notion that *ZmHb77* controls SRN and that SRN-dependent root architectural traits—in particular, lateral root density—improve drought tolerance and the recovery from drought stress.

Natural variation of the *ZmHb77* allele and function

To explore the natural variation of *ZmHb77* in association with root architecture and drought tolerance, we first identified five non-synonymous SNPs located in *ZmHb77* that form four major haplotype groups of maize inbred lines (Supplementary Fig. 13). The inbred lines with Hap1 formed significantly more seminal roots than the other haplotypes (Supplementary Fig. 13). We next aligned our structural–functional model results to georeferenced locations across the US. Interestingly, root system hydraulic conductance showed a general gradient pattern from the southwest dry area to the temperate region of the US (Fig. 6a), suggesting that root hydraulic conductance might have adapted with water availability. We then extended our drought analysis to the different traditional varieties and verified that Northern Flint varieties ($n = 5$) with fewer seminal roots contribute to drought tolerance and resilience (Supplementary Fig. 14a) for significantly higher biomass (Supplementary Fig. 14b) and stomatal conductance (Supplementary Fig. 14c) after re-watering. We next performed the

haplotype analysis for the *ZmHb77* allele in the traditional varieties and identified 41 high-confidence haplotypes (C allele) and the same number of A allele haplotypes (Supplementary Table 11). In particular, C allele haplotypes displayed significantly fewer seminal roots but significantly higher drought tolerance than the A allele haplotypes (Fig. 6b).

To further identify potential isogenic lines carrying the *ZmHb77* allele and drought tolerance based on Northern Flint-sourced varieties, we evaluated the SRN, lateral root density and dry biomass under well-watered and drought conditions for the whole Gaspé Flint introgression library introgressed into B73 (refs. 22,31). We first demonstrated that GF111 (inbred line developed by repeated selfing and selected from Gaspé Flint) had a great advantage with respect to drought tolerance and resilience in comparison to the inbred line B73 (Fig. 6c). Next, we explored the whole introgression population ($n = 68$) and identified that the lines with a higher share of the GF111 genome showed significantly ($r^2 = 0.12$, $P = 0.0015$) less seminal roots, but significantly ($r^2 = 0.38$, $P = 7.3 \times 10^{-9}$) higher lateral root density (Fig. 6d). At the same time, these genotypes provided drought tolerance as measured by the drought index of the dry biomass. Specifically, we identified four introgression lines (GF111^ZmHb77) with *ZmHb77* alleles from GF111 and another four lines (B73^ZmHb77) from B73, respectively. The GF111^ZmHb77 lines formed fewer seminal roots but a significantly higher lateral root density than the B73^ZmHb77 lines (Fig. 6d). We then performed an RNA sequencing experiment to explore the gene expression pattern in the embryo and root stele tissue. Interestingly, *ZmHb77* is, in general, lowly expressed in the embryo tissue but highly expressed in the root stele, where the lateral roots initiate (Fig. 6e), suggesting that the major function of *ZmHb77* is linked with lateral root formation. Based on the specific expression pattern of *ZmHb77* between the GF111^ZmHb77 and B73^ZmHb77 lines, *ZmHb77* might function in the promotion of seminal root formation but inhibition of lateral root density in maize seedlings (Fig. 6e). In particular, GF111^ZmHb77 lines displayed a strong drought tolerance as highlighted by a higher photosynthetic rate and stomal conductance (Fig. 6f and Supplementary Fig. 15). Indeed, less inhibition of *ZmHb77* on lateral root formation was demonstrated in the GF111^ZmHb77 lines under drought followed by re-watering (Fig. 6g). Interestingly, drought tolerance in maize driven by root architectural changes can be independently validated by the *rtcs* mutant and its wild type (Supplementary Fig. 16). Finally, we summarized our finding as a schematic model (Fig. 6h) in which *ZmHb77* acts as a central modulator contributing to the promotion of seminal root formation but inhibition of lateral root density in maize seedlings. Such root architectural and functional plasticity provides maize seedlings with great potential to balance external water constraints.

Discussion

Plant root system architecture has a critical role in the adaptation to environmental constraints[7,32]. However, to date, little is known about how the formation and function of root systems evolved in space and time during the domestication of agricultural crops, and it has

---

Fig. 5 | Functional characterization of *ZmHb77* controlling root traits and drought tolerance. a, Manhattan plot from GWA mapping of SRN in 1,604 diverse modern inbred lines. The dashed horizontal line represents the suggestive threshold ($P = 1.0 \times 10^{-5}$). The known gene *rtcs* and five novel candidate genes controlling SRN are indicated by arrows. **b**, Sequence of *ZmHb77* and the target sites of mutation by CRISPR–Cas9. PAM, protospacer-adjacent motif. **c–f**, CRISPR-knockout (KO1 and KO3) plants of *ZmHb77* display lower SRN (**c,d**) but higher lateral root density than the wild type (WT) (**e,f**). Root phenotyping was performed for 2-week-old maize plants grown in germination paper rolls. SRN was counted and lateral root density was obtained from the number of lateral roots per cm of primary root. For **d**, $n = 13$ biologically independent seedlings for WT and KO1, and $n = 10$ biologically independent seedlings for KO3. For **f**, $n = 15$ biologically independent seedlings for WT, $n = 11$ biologically independent seedlings for KO1, and $n = 10$ biologically independent seedlings

for KO3. Data are presented as mean values. Error bars, s.e.m. **g–i**, Comparison of drought tolerance between WT and plants of the two *ZmHb77* CRISPR-knockout lines grown under well-watered (**g**), drought (**h**) and drought followed by re-watering (**i**). In **g–i**, scale bar, 5 cm. **j**, Photosynthetic rate of mutants and wild type under well-watered and drought conditions. $n = 7$ biologically independent replicates. **k**, Survival rate of WT and *ZmHb77* knockout lines after exposure to drought stress. Wild-type and mutant seeds were precultured under well-watered conditions until the three-leaf stage and then either adequately supplied with water or not watered for another 12 days. Drought-stressed plants were re-watered and the survival rate was recorded after 7 days. Three biologically independent replicates were performed and each replicate included 12 individual plants. Significant differences between WT and KO lines are indicated by exact P values (one-sided Student's t-test). ns, not significant.
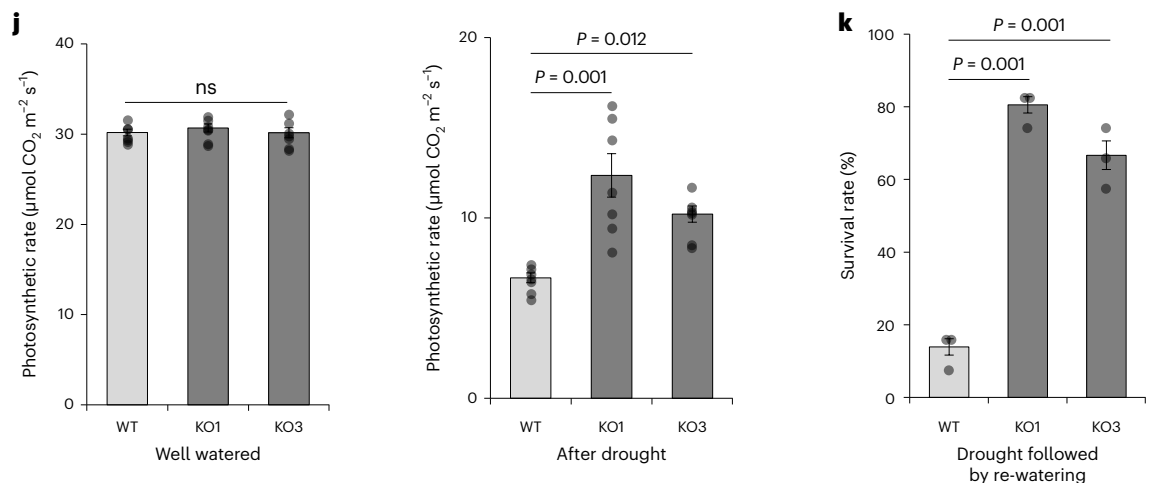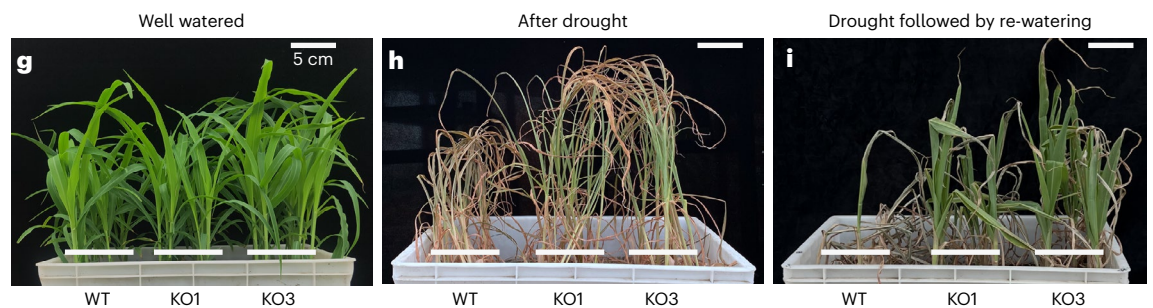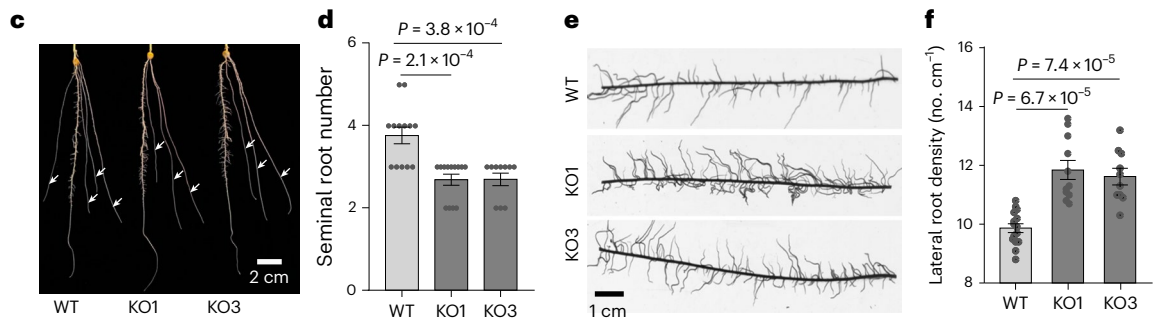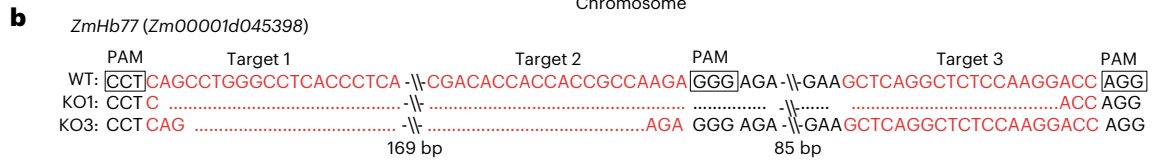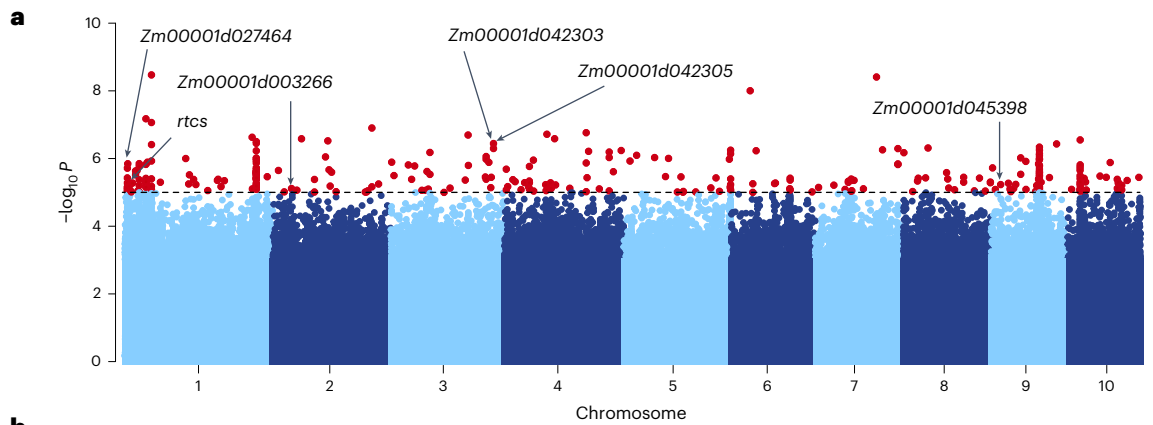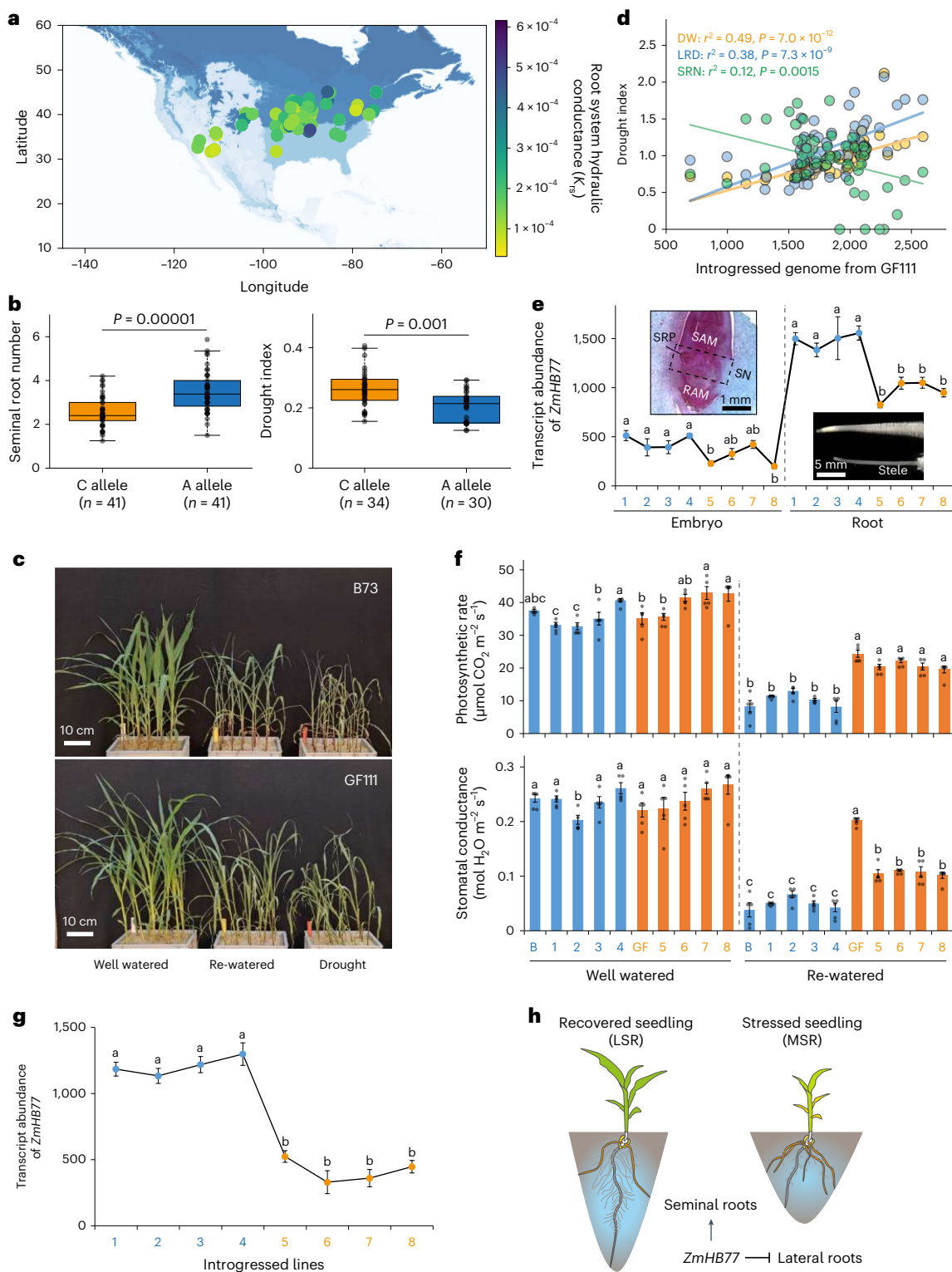
remained unclear to what extent root trait adaptation was required to introduce maize to new environments and what role root traits had in maize domestication. Using the global diversity of the genus *Zea*, our study demonstrates that SRN varies between domesticated maize traditional varieties and modern inbred lines compared with their wild teosinte progenitors and suggests that variation in SRN might have had an overriding role during the process of maize domestication[10] (Fig. 1). In traditional maize varieties, the demographically distinct groups sweet corn, flour corn and popcorn sourced from southwestern US have shown the fewest SRNs (Supplementary Fig. 4). Independent lines

of evidence indicate that adapted alleles, derived from Northern Flint maize contribute to the variation of SRN in both modern inbred lines and traditional varieties (Fig. 3). Subsequent local adaptation of SRN is in line with the maize domestication history, in which Northern Flint originated from the Southwest US desert[17,18] and then expanded to the northern US and Europe[33]. We further applied ecological and genomic models and found a clinal trend in SRN across latitude and climatic factors (Fig. 2a and Supplementary Fig. 9). Recently, such an adaptive signature has been reported in the geographical adaptation of rice to local soil nitrogen availability[34]. Here, we provide evidence for *rtcs*, a

Fig. 6 | Natural variation of the *ZmHb77* allele and its contribution to root architecture and drought tolerance in maize seedlings. a, Geographical distribution of root system hydraulic conductance. Each data point corresponds to the structural–functional model outcome. The precipitation data was derived from Köppen–Geiger climate classification maps at 1 km resolution (https://doi.org/10.1038/sdata.2018.214). The map was produced with Python (v.3.9.13) with the packages 'rasterio', 'geopanda' and 'matplotlib'. **b**, Haplotype analysis for traditional maize varieties. SRNs were counted in the paper-roll system (see Methods). The drought experiment was performed in soil and some varieties were removed from analysis because of unsuccessful germination. Significant differences between different alleles are indicated by exact *P* values (two-sided Student's *t*-test). Boxes span from the first to the third quartile, lines represent the median and whiskers include data within the 1.5× interquartile range of the lower and upper quartiles. Data points outside of whiskers represent outliers. **c**, Seedling performance of B73 and GF111 (inbred line developed by repeated selfing and selected from Gaspé Flint) grown under well-watered, drought and drought followed by re-watering. **d**, Correlation between drought index and the proportion of introgressed genome from GF111. Scatterplots show best fit (solid line) for linear regression. Two-sided test of significance with Bonferroni correction was performed to adjust the *P* value for multiple independent tests. LRD, lateral root density; DW, dry weight. **e**, Tissue-specific expression of *ZmHb77* in the embryo and root stele between different introgressed lines from B73

(1–4) and GF111 (5–8) donors. Significant differences between introgressed lines within each tissue are indicated by different letters (one-way ANOVA, Tukey's HSD, *P* = 0.04). *n* = 4 biologically independent replicates. Data are presented as mean values. Error bars, s.e.m. SAM, shoot apical meristem; RAM, root apical meristem; SN, scutellar node; SRP, seminal root primordia. **f**, Photosynthetic rate and stomatal conductance of different introgression lines from B73 and GF111 donors under well-watered and re-watered conditions, respectively. Significant differences between different lines are indicated by different letters under well-watered and re-watered conditions, respectively (one-way ANOVA, Tukey's HSD, *P* = 0.001). Data are presented as mean values. Error bars, s.e.m. and *n* = 4 biologically independent replicates per genotype and treatment. **g**, Expression of *ZmHb77* in the root stele tissue after re-watering among different introgression lines. Significant differences between introgressed lines in the root stele are indicated by different letters (one-way ANOVA, Tukey's HSD, *P* = 0.001). For **e** and **g**, data are presented as mean values. Error bars, s.e.m.; *n* = 5 biologically independent replicates per genotype and treatment. Detailed information on introgressed lines (Supplementary Fig. 15) is listed with B73 donors (1, IL-100-5-8-5; 2, IL-10-7-1; 3, IL-130-2-8-2; 4, IL-13-7-4) and GF111 donors (5, IL-121-6-6-6; 6, IL-130-2-8-5; 7, IL-140-8-4-4; 8, IL-140-8-4-5). **h**, Working model of a potential function of *ZmHb77* on the formation of seminal roots and lateral roots in contribution to maize seedling drought tolerance. MSR, more seminal roots; LSR, less seminal roots.

known determinant of SRN, to associate with variation in SRN along geographical gradients (Fig. 2e), emphasizing the importance of landscape and environmental factors in driving root trait differentiation.

In the near future, climate change will increase the incidence of drought, imposing a major threat to crop production[35]. Improved adaptive capacity to withstand flash drought is required for crops to mitigate such negative impacts in agricultural systems[36]. To tolerate stress and optimize the uptake of water even with a transient drought period, crops need to adapt root properties. We detected enhanced lateral root branching in both traditional varieties (Fig. 1c) and the *rtcs* mutants when seminal roots were absent (Fig. 4c) as well as recovering a similar result through in silico modeling (Fig. 4a). At the seedling stage, traditional varieties with fewer seminal roots can substantially reduce the carbon cost for the seed, and thus enable the formation of highly dense and long lateral roots along the primary root (Fig. 4). Interestingly, we detected a significantly higher accumulation of lignin in the primary root tip of traditional varieties with few or no seminal roots under osmotic stress conditions (Supplementary Fig. 14). Such adaptive behavior with enhanced lateral root branching in contact with water[25] and primary root lignification for better penetration of hard and dry soil[37] improves plant tolerance to limited water availability, especially for the survival of seedlings after severe drought[38]. In this context, we identified the transcription factor *ZmHb77* that affects overall root architecture by increasing SRN but suppressing lateral root density (Fig. 5c–f). Deletion of *ZmHb77* ultimately enhances the survival of plants after recovery from drought (Fig. 5j,l). Indeed, domesticated wheat and barley have also been reported to form a larger number of seminal roots than their wild relatives[12,39]. Based on the global warming scenario and an increasing incidence of drought, it is necessary to consider reducing the number of seminal roots in favor of lateral root branching for more efficient acquisition of soil water in the modern cultivars. It is important to note that such architectural plasticity will have its major impact during the seedling stage before crown roots become established[9] and sustain water uptake at later developmental stages. Our systemic analyses indicate that SRN is an important driver for the formation and pattern of lateral roots along the primary root (Fig. 6), thereby determining the overall absorptive surface and foraging capacity of crop roots. Variation in SRN alters hydraulic properties and may bear genetic potential to modify root plasticity and deepen our understanding of how plant roots sense and adapt to fluctuating water availability by hydropatterning[40] or xerobranching[41]. Future studies need to address how SRN variation can optimize root development and

hydraulic architecture for enhanced resilience in cereals[42]. Our results not only reveal the past signature of domestication and adaptation of maize roots but highlight the genetic potential to improve climate resilience in future crops.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-01761-3.

References

1. Meyer, R. S. & Purugganan, M. D. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14**, 840–852 (2013).
2. Hake, S. & Ross-Ibarra, J. Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife* **4**, e05861 (2015).
3. Yang, N. et al. Two teosintes made modern maize. *Science* **382**, 1013 (2023).
4. Ross-Ibarra, J. & Piperno, D. Maize moving. *Figshare* https://doi.org/10.6084/m9.figshare.12781307.v1 (2020).
5. Romero Navarro, J. A. et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* **49**, 476–480 (2017).
6. Swarts, K. et al. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* **357**, 512–515 (2017).
7. Ma, Z. et al. Evolutionary history resolves global organization of root functional traits. *Nature* **555**, 94–97 (2018).
8. Eshel, A. & Beeckman, T. (eds) *Plant roots: the hidden half* (CRC Press, 2013).
9. Hochholdinger, F., Yu, P. & Marcon, C. Genetic control of root system development in maize. *Trends Plant Sci.* **23**, 79–88 (2018).
10. Lopez-Valdivia, I. et al. Gradual domestication of root traits in the earliest maize from Tehuacán. *Proc. Natl Acad. Sci. USA* **119**, e2110245119 (2022).
11. Yu, P., Gutjahr, C., Li, C. & Hochholdinger, F. Genetic control of lateral root formation in cereals. *Trends Plant Sci.* **21**, 951–961 (2016).

12. Golan, G., Hendel, E., Méndez Espitia, G. E., Schwartz, N. & Peleg, Z. Activation of seminal root primordia during wheat domestication reveals underlying mechanisms of plant resilience. *Plant Cell Environ.* **41**, 755–766 (2018).

13. Perkins, C. & Lynch, J. P. Increased seminal root number associated with domestication improves nitrogen and phosphorus acquisition in maize seedlings. *Ann. Bot.* **128**, 453–468 (2021).

14. Hochholdinger, F., Woll, K., Sauer, M. & Dembinsky, D. Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programmes. *Ann. Bot.* **93**, 359–368 (2004).

15. Burton, L., Brown, K. M. & Lynch, J. P. Phenotypic diversity of root anatomical and architectural traits in *Zea* species. *Crop Sci.* **53**, 1042–1055 (2013).

16. Taramino, G. et al. The maize (*Zea mays* L.) *RTCS* gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. *Plant J.* **50**, 649–659 (2007).

17. Merrill, W. L. et al. The diffusion of maize to the southwestern United States and its impact. *Proc. Natl Acad. Sci. USA* **106**, 21019–21026 (2009).

18. da Fonseca, R. R. et al. The origin and evolution of maize in the Southwestern United States. *Nat. Plants* **1**, 14003 (2015).

19. Doebley, J. F., Goodman, M. & Stuber, C. W. Exceptional genetic divergence of northern flint corn. *Am. J. Bot.* **73**, 64–69 (1986).

20. Rebourg, C. et al. Maize introduction into Europe: the history reviewed in the light of molecular data. *Theor. Appl. Genet.* **106**, 895–903 (2003).

21. Hu, Y. et al. Genome assembly and population genomic analysis provide insights into the evolution of modern sweet corn. *Nat. Commun.* **12**, 1227 (2021).

22. Salvi, S. et al. Genetic dissection of maize phenology using an intraspecific introgression library. *BMC Plant Biol.* **11**, 4 (2011).

23. Osthoff, A. et al. Transcriptomic reprogramming of barley seminal roots by combined water deficit and salt stress. *BMC Genom.* **20**, 325 (2019).

24. Zhou, X. et al. CPlantBox, a whole-plant modelling framework for the simulation of water- and carbon-related processes. *in silico Plants* **2**, diaa001 (2020).

25. Ahmed, M. A., Zarebanadkouki, M., Kaestner, A. & Carminati, A. Measurements of water uptake of maize roots: the key function of lateral roots. *Plant Soil* **398**, 59–77 (2016).

26. Abdalla, M. et al. Stomatal closure during water deficit is controlled by below-ground hydraulics. *Ann. Bot.* **129**, 161–170 (2022).

27. Cai, G., Ahmed, M. A., Abdalla, M. & Carminati, A. Root hydraulic phenotypes impacting water uptake in drying soils. *Plant Cell Environ.* **45**, 650–663 (2022).

28. Li, C. et al. Genomic insights into historical improvement of heterotic groups during modern hybrid maize breeding. *Nat. Plants* **8**, 750–763 (2022).

29. Marcon, C. et al. *BonnMu*: a sequence-indexed resource of transposon-induced maize mutations for functional genomics studies. *Plant Physiol.* **184**, 620–631 (2020).

30. Qiu, X. et al. Genome-wide identification of HD-ZIP transcription factors in maize and their regulatory roles in promoting drought tolerance. *Physiol. Mol. Biol. Plants* **28**, 425–437 (2022).

31. Salvi, S. et al. Registration of Gaspé Flint 1.1.1, a small-size early-flowering maize inbred line. *J. Plant Registrat.* **16**, 152–161 (2021).

32. Giehl, R. F. & von Wirén, N. Root nutrient foraging. *Plant Physiol.* **166**, 509–517 (2014).

33. Tenaillon, M. I. & Charcosset, A. A European perspective on maize history. *Biologies* **334**, 221–228 (2011).

34. Liu, Y. et al. Genomic basis of geographical adaptation to soil nitrogen in rice. *Nature* **590**, 600–605 (2021).

35. Jägermeyr, J. et al. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food* **2**, 873–885 (2021).

36. Yuan, X. et al. A global transition to flash droughts under climate change. *Science* **380**, 187–191 (2023).

37. Schneider, H. M. et al. Multiseriate cortical sclerenchyma enhance root penetration in compacted soils. *Proc. Natl Acad. Sci. USA* **118**, e2012087118 (2021).

38. Wang, X. et al. Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genetics* **48**, 1233–1241 (2016).

39. Grando, S. & Ceccarelli, S. Seminal root morphology and coleoptile length in wild (*Hordeum vulgare* ssp. *spontaneum*) and cultivated (*Hordeum vulgare* ssp. *vulgare*) barley. *Euphytica* **86**, 73–80 (1995).

40. Orosa-Puente, B. et al. Root branching toward water involves posttranslational modification of transcription factor ARF7. *Science* **362**, 1407–1410 (2018).

41. Mehra, P. et al. Hydraulic flux-responsive hormone redistribution determines root branching. *Science* **378**, 762–768 (2022).

42. Maurel, C. & Nacry, P. Root architecture and hydraulics converge for acclimation to changing water availability. *Nat. Plants* **6**, 744–749 (2020).

43. Liu, K. et al. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* **165**, 2117–2128 (2003).

44. Gouesnard, B. et al. Genotyping-by-sequencing highlights original diversity patterns within a European collection of 1191 maize flint lines, as compared to the maize USDA genebank. *Theor. Appl. Genet.* **130**, 2165–2189 (2017).

¹Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany. ²Emmy Noether Group Root Functional Biology, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany. ³State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, PR China. ⁴Department of Plant Science, The Pennsylvania State University, State College, PA, USA. ⁵College of Resources and Environment, and Academy of Agricultural Sciences, Southwest University (SWU), Chongqing, PR China. ⁶Laboratorio de Biodiversidad y Funcionamiento Ecosistémico. Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, Sevilla, Spain. ⁷Unidad Asociada CSIC-UPO (BioFun), Universidad Pablo de Olavide, Sevilla, Spain. ⁸Institute of Bio- and Geosciences, Plant Sciences (IBG-2), Forschungszentrum Juelich GmbH, Juelich, Germany. ⁹Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany. ¹⁰Horticultural Sciences Department, University of Florida, Gainesville, FL, USA. ¹¹Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy. ¹²Chair of Root-Soil Interactions, TUM School of Life Sciences,

Technical University of Munich, Freising, Germany. [13]Institute of Bio- and Geosciences, Agrosphere (IBG-3), Forschungszentrum Jülich GmbH, Jülich, Germany. [14]Earth and Life Institute, Université catholique de Louvain, UCLouvain, Belgium. [15]Institute of Cellular and Molecular Botany (IZMB), Department of Ecophysiology, University of Bonn, Bonn, Germany. [16]Intercollege Graduate Degree Program in Plant Biology, The Pennsylvania State University, State College, PA, USA. [17]Key Laboratory of Plant-Soil Interactions, College of Resources and Environmental Sciences, National Academy of Agriculture Green Development, Ministry of Education, China Agricultural University, Beijing, PR China. [18]Plant Breeding, TUM School of Life Sciences, Technical University of Munich, Freising, Germany. [19]North Central Regional Plant Introduction Station, USDA-Agriculture Research Service and Iowa State University, Ames, IA, USA. [20]Department of Physiology and Cell Biology, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany. [21]These authors contributed equally: Peng Yu, Chunhui Li, Meng Li, Xiaoming He. ✉e-mail: yupeng@uni-bonn.de; rjs6686@psu.edu; wangtianyu@caas.cn; hochholdinger@uni-bonn.de

Methods

Global investigation of maize germplasm

Global maize germplasm was collected for systematic root phenotyping. A global collection of 2,444 modern inbred lines held in Ames, Iowa (Ames Inbred Diversity panel), was contributed by the United States Department of Agriculture (https://www.grin-global.org). The European germplasm used in this study included the CornFed panel consisting of 429 inbred lines, 38 inbred lines from the University of Hohenheim (Germany) and 20 inbred lines from Centro Investigacións Agrarias Mabegondo (Spain). Furthermore, we surveyed 1,604 Chinese inbred lines donated by the Chinese Academy of Agricultural Sciences representing the dominant genotypes used in Chinese maize breeding[28]. Finally, 4,868 traditional variety accessions and 173 teosinte accessions were contributed by the US National Plant Germplasm System (https://npgsweb.ars-grin.gov/gringlobal), the International Center for Maize and Wheat Improvement (www.cimmyt.org) collection, the Chinese Academy of Agricultural Sciences and the Technical University of Munich.

High-throughput root phenotyping

A high-throughput paper-roll system was used for phenotyping the seedling root system[45]. To synchronize the growth rates of different genotypes, we grew maize for 10 days until all seminal roots were formed in all maize genotypes. Ten and twenty representative seeds were germinated for maize inbred lines and traditional varieties, respectively. Owing to seed limitation, 15 teosinte seeds for each accession were investigated accordingly. In brief, the kernels were placed in one line on germination paper (length, 38 cm; width, 25 cm; Anchor Paper Co., Saint Paul, USA) at about 2 cm from the top, with an interspace of 3–4 cm. The paper was then rolled to ensure that all kernels stayed in place, and rolls with kernels were placed in a 5 l beaker with 1 l of deionized water and then transferred into a phytochamber with a 16 h light, 26 °C and 8 h dark, 18 °C cycle. The number of seminal roots was defined and recorded in Supplementary Fig. 1. The types of seminal roots include the dorsal seminal roots initiated between the mesocotyl and the seed embryo and ventral seminal roots initiated from the scutellar node (Supplementary Fig. 1).

Non-invasive root and carbon imaging using MRI–PET

For the MRI and PET experiment, the *rtcs* mutant and its wild type B73 were grown in pots (20 cm height, 8 cm inner diameter) filled with a mixture of 83.3% quartz sand plus 16.7% loam. The loam was derived from a Haplic Phaeozem in 0–50 cm depth (Schladebach, Germany) and the sand was from WF33, Quarzwerke (Frechen, Germany). Soil and sand were homogenized, dried and sieved to 1 mm. The mixture was fertilized, homogeneously filled up to 18 cm pot height and compacted to a bulk density of 1.47 g cm^{-3}. Seeds were surface-sterilized in 10% H_2O_2 for 10 min, primed in a saturated $CaSO_4$ solution for 3 h and planted at 1 cm depth (days after planting, 0). The pot surface was covered with perforated plastic foil. Plants were grown in a climate chamber at the same conditions described for the MRI experiment. Plants were watered to a soil volumetric water content of 18% with deionized water every other day.

Plants were measured with PET at 7, 15 and 21 days after planting (wild type) or 11, 15 and 22 days after germination (*rtcs* mutant). The PET climate chamber was set to the same conditions as the growth chamber of the plants but illumination was provided by home-built LED lamps with photosynthetically active radiation of 350 µmol m^{-2} s^{-1} at canopy level. For labeling, the shoot was sealed airtight in a cuvette connected to a gas exchange and ^{11}C tracer application system[46] and placed inside the PET measurement volume. The system was used to measure gas exchange of the plant in an open mode and to circulate $^{11}CO_2$ with a radioactivity of about 100 MBq for 6 min through the plant cuvette in a closed loop. Afterward, the system was opened to the environment again with exhaust gas passing a CO_2 absorber (Soda Lime HC Atemkalk,

Medisize Deutschland) and used for monitoring plant gas exchange as before to ensure the plant was actively photosynthesizing. The gas flow inside the cuvette was similar to the conditions in the climate chamber.

The PET instrument used for these measurements is a custom-built device based on digital silicon-based photo multipliers (Philips Digital Photon Counting) with a field of view of 18 cm in diameter and 20 cm height[47,48]. Measurement duration was 140 min starting immediately after the 5 min labeling pulse of $^{11}CO_2$, and images were reconstructed with a nominal resolution of 0.9 × 0.9 × 1.0 mm^3.

Images were reconstructed with PRESTO (v.3.01)[49] and prepared for print medium by making a maximum intensity projection over time of the PET image, co-registering with MRI images, and normalizing view angle, contrast and brightness in all images of the same plant to those of the respective first measurement with MeVisLab (v.3.4; MeVis Medical Solutions AG).

Structural–functional model of root system architecture

To explore the effect of seminal root variation on the overall root architecture, we performed a four-step image analysis for 218 traditional US varieties grown in the rhizobox as described in Supplementary Fig. 17 (ref. [50]). First, a convolutional neural network model was trained with *RootPainter* (v.0.16) to automatically segment the root system from the background[51]. Second, the segmented root system was binarized, and residing segmentation errors were corrected by hand. Third, each root system was semi-automatically traced with *Root System Analyzer* (*RSA*; v.2.0)[52]. For this purpose, we manually defined the start and end points of the primary root and each seminal root. The lateral roots were automatically detected by *RSA*, and detection errors were manually corrected. As a last step, an RSML file was exported from *RSA* and migrated to *SmartRoot* (v.4.21), which we used to manually label seminal and crown roots[53,54]. *SmartRoot* was also used to finally extract all architectural parameters, including a full *CPlantBox* (v.2.0) parameter set for each root system[24,55]. According to the young age of the root systems, we calculated linear elongation rates for each root type. The parameter set was used to generate five realizations of virtual root systems with the stochastic *CPlantBox* model, each representing one of the root systems grown on filter paper. From the root diameter of the generated root systems and the allometry that we could find and anatomical traits according to previous studies[15,56–59], average root anatomies were created with the *GRANAR* (v.1.1) model[60] for each root type present on the generated root system. The radial hydraulic conductivity and axial hydraulic conductance (K_r and K_x) of these root anatomies were estimated with the model *MECHA* (v.2.1)[61] using the same subcellular hydraulic properties[60]. The K_r estimations match three scenarios accounting for three hydrophobic barriers set up per generated root anatomy; one with only an endodermal Casparian strip, a second with a fully suberized endodermis and a third with a fully suberized endodermis and an exodermal Casparian strip. Based on previous work[59], an arbitrary delay of 2 days between two hydrophobic barrier scenarios was made to achieve maturity of the roots at the apical unbranched zone. The K_x estimations match two scenarios accounting for the maturation of the meta-xylem vessels, which, in this case, mature 1 day after the suberization of the endodermis. Before that, only proto-xylem vessels were included for the axial water uptake. Here, we used the root diameters to distinguish between the different genotypes, as they are one of the major factors influencing the root hydraulic properties[62]. From the *CPlantBox* root system architectures and their respective root hydraulic conductance, the K_{rs} and the standard uptake fractions were determined with the model *MARSHAL* (v.1.1)[63] for each virtual root system.

Linking SRN to environmental conditions

We first compared the SRN of traditional varieties with geographical coordinates across main climates using a one-way PERMANOVA (permutational multivariate ANOVA) with a post-hoc test. Climate

types were classified using the Köppen–Geiger climate classification (v.1; http://koeppen-geiger.vu-wien.ac.at/present.htm). We then conducted a machine-learning random forest analysis to assess the most important environmental factors associated with SRN. We considered key environmental factors such as climate type, current and past (Mid-Holocene, about 6,000 years ago) temperature and precipitation conditions, current climatic seasonality, soil properties and topography. Current and past climatic conditions were retrieved from the Worldclim dataset (https://worldclim.org/; ~1 km resolution). We included past climatic conditions to account for the influence of climate during the domestication process. We focused on past mean annual temperature and precipitation because we have reliable information on the past for these variables. Soil properties were retrieved from https://soilgrids.org (v.2; 250 m resolution).

Environmental GWAS and adaptive alleles analysis

Phenotypic data. A collection of 681 *Zea mays* traditional varieties accessions from the Maize Center of Origin, Mexico, were used to associate variation in SRN to climatic and edaphic gradients.

Environmental data. For each georeferenced accession of traditional varieties, we compiled climatic and soil descriptors representative of the long-term averages of their point of origin, following methods used in previously published work[64] (Supplementary Table 12). All used databases are publicly available and have global coverage. Data were collected from WorldClim[65], the NCEP/NCAR Reanalysis Project[66], NASA Surface Radiation Budget (https://asdc.larc.nasa.gov/project/SRB), Climate Research Unit[67], SoilGrids[68], the Global Soil Dataset[69] and ArcGIS[70,71]. See Supplementary Table 12 for a description of all environmental variables.

Feature selection. Importantly, not all variables in the environmental dataset will describe variation in SRN. Feature selection before model building eliminates unimportant or redundant variables by identifying those with significant associations to an outcome variable, improving model accuracy. The feature selection method Boruta was employed to identify environmental aspects that describe variation in SRN. Aspects of the abiotic environment that significantly described variation in SRN were identified using the Boruta() function from the Boruta package (v.7.0.0) in R[72].

Random forest. Environmental variables identified by Boruta were used for random forest model construction. Random forest works under the expectation that a response variable can be described by several explanatory variables through the construction of decision trees. Thus, each random forest model is representative of the non-linear, unique combination of explanatory variables that describe variation in a response variable. Random forest models were built in R (v.4.2.2) using the randomForest() function under default parameters. A total of 5,000 trees were built and one-third the number of explanatory variables were tried at each split[73]. Random forest models were trained with 80% of the data. Model success was evaluated with Nash–Sutcliffe efficiency, Out-of-bag R^2 and the mean absolute error of the remaining validation set.

Using the constructed random forest models, SRNs were predicted for 1,781 genotyped traditional varieties in Mexico. These traditional varieties were genotyped as a part of the Seeds of Discovery project (SeeD).

GWAS. We conducted a genome-association analysis to measure the associations between SNPs of traditional variety genotypes and SRNs predicted by environmental variables. We used a linear model[74] to fit the genotypic data and predicted SRNs for these accessions of traditional varieties. The first five eigenvectors of the genetic relationship matrix were included in the model to control for population structure.

Gene-level analysis. The summary SNP statistics from GWAS was analyzed using MAGMA (Multi-marker Analysis of GenoMic Annotation) to aggregate the effect of multiple genetic markers simultaneously and determine their joint association with the phenotype at the gene level[75].

Population genetic structure analysis. We performed population structure analysis to estimate the ancestry coefficient of the studied traditional varieties using the R package tess3r[76,77]. We ran the function tess3 assuming ancestry populations (K) from 1–9, and selected K = 5 based on the cross-validation scores.

Genetic analysis for traditional maize varieties

We next genotyped over 3,000 traditional maize varieties covering the global diversity. In brief, genomic DNA was extracted from the leaves of bulked maize seedlings for each traditional variety accession. Quantification and qualification of DNA were checked by agarose gel electrophoresis and a Qubit DNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies). For library preparation, 1 µg genomic DNA was digested using restriction enzymes, and the obtained fragments were ligated with barcodes and amplified by PCR. Subsequently, DNA fragments from different samples were pooled, and the desired fragments of DNA were recovered by electrophoresis. The quality of the libraries was controlled using the Qubit2.0 kit. Agilent 2100 was used to check the insert size of the libraries after diluting the library to 1 ng µl^{-1}. Quantitative PCR was performed to detect the effective concentration of libraries (the effective concentration of library >2 nM) when the insert size was appropriate. All steps were taken to ensure the quality of the libraries. The constructed libraries were sequenced by an Illumina NovaSeq 6000 system with a sequencing strategy of paired-end 150 bp. The raw FASTQ sequencing data were qualified based on the Q20 and Phred quality scores (Q_{phred} < 1%). Raw reads were filtered to remove reads containing adapters or reads of low quality; for example, uncertain nucleotides constitute >10% of either read (N > 10%). Reads were discarded when low-quality nucleotides (base quality of <5) constituted more than 50% of the read. Consistent with the Illumina sequencing platform, for pair-end data, we required that the Q30 was >85% and the error rate was <0.1%. The effective sequencing data was aligned with the maize B73 reference genome v.4 through Burrows–Wheeler Aligner[77] software (parameters, mem -t 4 -k 32 -M). The duplicates were removed by SAMTOOLS (v.1.12)[78] (parameters, rmdup). SNPs were detected by SAMTOOLS (mpileup -m 2 -F 0.002 -d 1000) with the following standards: read number per SNP > 4, <1000; SNP quality >20. To calculate the introgression of the Northern Flint genome, we filtered low-quality SNPs (minor allele frequency of <0.05, missing rate of >50%), and a total of 682,044 SNPs were obtained. The missing genotypes were then imputed with Beagle 4 (v.5.2) with default parameters[79]. Population structure was conducted with ADMIXTURE tool (v.1.3.0)[80] with default parameters. We then investigated the correlation between the dosage of introgressed genome from Northern Flint and SRN. The reference Northern Flint-sourced traditional varieties were defined according to the published article[19].

GWA mapping for a maize resequencing population

A GWAS for SRN was performed using a mixed linear model in the software EMMAX (version emmax-intel-binary-20120210). A total of 18,169,560 high-quality SNPs with minor allele frequency of ≥0.01 and missing rate of ≤20% in the 1,604 inbred lines were employed in the GWAS[28]. To determine the genome-wide threshold, we conducted permutation tests by randomly shuffling the phenotypes for SRN. Then we applied GWAS on the permuted phenotypes by using the same model that was used for the real observed phenotypes. The most significant P value across the whole genome was recorded. This random process was repeated 100 times. The distribution of the most significant P values across the 100 replicates was used to determine the threshold, which was the P value corresponding to a 5% chance of

a type I error. Finally, the threshold was set as $P < 1.0 \times 10^{-6}$. However, under this threshold, only 29 SNPs significantly associated with SRN were identified. We conducted GWAS for SRN using the kinship matrix of pairwise genetic distances as the variance–covariance in the mixed linear model of GWAS. The loci associated with population structure and false negative loci were filtered out. In addition, numerous minor effect loci for this trait were hardly detected under the strict threshold. Collectively, we conservatively chose 1.0×10^{-5} as the suggestive threshold to determine association signals. Although some false positive signals might be detected, the suggestive threshold would greatly contribute to a deep mining of candidate genes for SRN. According to the GWAS association signals, we estimated the candidate regions by pairwise linkage disequilibrium correlations. The linkage disequilibrium blocks (average ~20 kb) around peak SNPs (above the suggestive threshold) were defined as the candidate-associated regions. Genes within the candidate-associated regions were selected as the candidate genes for SRN.

Genome editing and evaluation of drought tolerance

For gene editing of *ZmHb77*, the single guide RNA (sgRNA) construct was designed and introduced into sgRNA expression cassettes by overlapping PCR[81]. The three sgRNA expression cassettes of *ZmHb77* were then integrated into the pCPB-ZmUbi::hSpCas9 vector. The gene-editing construct of *ZmHb77* was introduced into *Agrobacterium* strain EHA105 and transformed into the immature embryo of the maize inbred line B104 through *Agrobacterium*-mediated transformation. The *ZmHb77* knockout mutants of the T_2 generation and wild-type plants were used to conduct phenotyping in the paper-roll culture for 1-week-old seedlings.

For the survival rate experiment, two knockout lines and wild-type plants were planted in soil (turf to vermiculite in a ratio of 1:1). Three biological replications were performed using a cultivation box (length × width × depth, 52 × 34 × 15 cm) and 12 plants for each line planted per cultivation box as one replication. Irrigation was stopped at the three-leaf stage and drought treatment was performed for about 12 days. The plants were then re-watered, and the survival rate was recorded after 7 days. An LI-6400 portable photosynthesis system (LI-COR) was used to obtain photosynthesis measurements and stomatal conductance on the latest fully expanded leaves between 09:00 and 11:00 h.

Evaluation of root and drought of introgression population

To understand the contribution of Northern Flint-sourced germplasm to the root phenotype and drought tolerance, we grew the whole introgression population (*n* = 68; Supplementary Table 13) and the founders Gaspé Flint and B73 under well-watered and drought conditions. Different introgression lines were sown in a cultivation box (length × width × depth, 30 × 20 × 12 cm) that was filled with 4 kg of field soil with a 50% vermiculite mixture. The drought treatment was applied to the soil-grown seedlings at the three-leaf stage by reducing the watering and holding the soil water capacity at 22%, determined by a multifunctional device COMBI 5000 (STEP Systems). Approximately 20 days after cultivation, shoot dry biomass, SRN and lateral root density were determined accordingly after harvest. The drought index for each parameter was calculated as the determined ratio under drought and well-watered conditions. Meanwhile, the fresh leaf tips were sampled for genomic DNA extractions and whole genome resequencing for all introgressed lines together with founders Gaspé Flint and B73. Owing to the heterozygosity of Gaspé Flint, we further sequenced a purified inbred GF111, which was developed by repeated selfing (eight cycles) and selection starting from the Northern Flint variety Gaspé Flint[31]. The correlation between the dosage of the introgression genome from GF111 and the drought index measured by different traits (dry biomass, SRN and lateral root density) were plotted across the whole panel of introgression lines accordingly.

RNA sequencing experiment

We next carried out an RNA sequencing experiment to explore the gene expression pattern in the embryo and root stele tissue using different introgression and non-introgression lines of *ZmHb77* in a paper-roll system. We first identified four (IL-121-6-6-6, IL-130-2-8-5, IL-140-8-4-4 and IL-140-8-4-5) introgression lines (GF111^ZmHb77^) with *ZmHb77* alleles from the GF111 genome and another four (IL-100-5-8-5, IL-10-7-1, IL-130-2-8-2 and IL-13-7-4) lines (B73^ZmHb77^) from the B73 genome, respectively. In brief, these eight lines were cultivated in a paper-roll system[45] for 3 days until the first primary root tip was visible. The endosperm of these lines was hand-dissected and removed. The central parts of the isolated embryos were specifically dissected at the scutellar node tissue where the seminal roots are initiated[9]. For each replicate, ten individual tissues were pooled for each line. These lines were further grown on the paper-roll system for an additional 4 days until the primary root reached around 10 cm in length, at which point the stele tissue was manually peeled off[82]. Ten stele tissues per genotype were defined as one biological replicate. RNA extraction from embryo and stele tissues, RNA sequencing and bioinformatic analysis was performed according to our recent publication[45]. We further grew these eight lines under well-watered and re-watered conditions to investigate their drought resilience. The whole experiment was carried out using the cultivation box system as described above. The photosynthesis rate and stomatal conductance were estimated and the stele tissues were dissected and sequenced for gene expression for these eight lines under re-watered conditions.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw seminal root phenotyping data, geographical coordinates and soil modeling data are provided in Supplementary Tables 1–8. All germplasm information that is the geographically diverse teosinte accessions, maize traditional varieties and inbred lines contributed by NCRPIS, CIMMYT and the Chinese maize seed germplasm bank at the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (China) used in this study are summarized in the Supplementary Tables 1–5. Geographical coordinates and elevation information of the collection sites for the traditional maize varieties were retrieved from the public database of the US National Plant Germplasm System (https://www.grin-global.org). Soil and climate data were collected from WorldClim (https://worldclim.org), the NCEP/NCAR Reanalysis Project (https://psl.noaa.gov/data/reanalysis/reanalysis.shtml), NASA SRB (https://asdc.larc.nasa.gov/project/SRB), Climate Research Unit (https://www.uea.ac.uk/groups-and-centres/climatic-research-unit), SoilGrids (https://soilgrids.org v.2) and the Global Soil Dataset (https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en). Maize genome resequencing data of the Gaspé Flint introgression panel and root RNA sequencing data were deposited in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the BioProject ID PRJNA1095206. Source data are provided with this paper.

Code availability

The customized scripts included in this study are available at GitHub (https://github.com/PengYuMaize/GlobalSeminalRoot) with https://doi.org/10.5281/zenodo.10985812 (ref. 83).

References

45. Yu, P. et al. Plant flavones enrich rhizosphere Oxalobacteraceae to improve maize performance under nitrogen deprivation. *Nat. Plants* **7**, 481–499 (2021).

46. Metzner, R. et al. *In vivo* imaging and quantification of carbon tracer dynamics in nodulated root systems of pea plants. *Plants* **11**, 632 (2022).

47. Streun, M. et al., PhenoPET: a dedicated PET scanner for plant research based on digital SiPMs (DPCs). In *IEEE Medical Imaging Conference* (IEEE, 2014).

48. Hinz, C. Accurate quantitative and dynamic PET imaging with the phenoPET Scanner for plant studies. PhD dissertation, University of Wuppertal (2021).

49. Scheins, J. J. et al. High performance volume-of-intersection projectors for 3D-PET image reconstruction based on polar symmetries and SIMD vectorisation. *Phys. Med. Biol.* **60**, 9349–9375 (2015).

50. Bauer, F. M. et al. Development and validation of a deep learning based automated minirhizotron image analysis pipeline. *Plant Phenoics* **2022**, 9758532 (2022).

51. Smith, A. G. et al. RootPainter: deep learning segmentation of biological images with corrective annotation. *New Phytol.* **236**, 774–791 (2022).

52. Leitner, D. et al. Recovering root system traits using image analysis exemplified by two-dimensional neutron radiography images of lupine. *Plant Physiol.* **164**, 24–35 (2014).

53. Lobet, G. et al. A novel image-analysis toolbox enabling quantitative analysis of root system architecture. *Plant Physiol.* **157**, 29–39 (2011).

54. Lobet, G. et al. Root system markup language: toward a unified root architecture description language. *Plant Physiol.* **167**, 617–627 (2015).

55. Schnepf, A. et al. CRootBox: a structural–functional modelling framework for root systems. *Ann. Bot.* **121**, 1033–1053 (2018).

56. Chimungu, J. G. et al. Reduced root cortical cell file number improves drought tolerance in maize. *Plant Physiol.* **166**, 1943–1955 (2014).

57. Gao, K. et al. A comprehensive analysis of root morphological changes and nitrogen allocation in maize in response to low nitrogen stress. *Plant Cell Environ.* **38**, 740–750 (2015).

58. Yang, J. T. et al. Genotypic variation and nitrogen stress effects on root anatomy in maize are node specific. *J. Exp. Bot.* **70**, 5311–5325 (2019).

59. Heymans, A. et al. Combining cross-section images and modeling tools to create high-resolution root system hydraulic atlases in *Zea mays*. *Plant Direct* **5**, e334 (2021).

60. Heymans, A. et al. GRANAR, a computational tool to better understand the functional importance of monocotyledon root anatomy. *Plant Physiol.* **182**, 707–720 (2020).

61. Couvreur, V. et al. Going with the flow: multiscale insights into the composite nature of water transport in roots. *Plant Physiol.* **178**, 1689–1703 (2018).

62. Heymans, A. *In silico* analysis of the influence of root hydraulic anatomy on maize (*Zea mays*) water uptake. PhD dissertation, Université Catholique de Louvain (2022).

63. Meunier, F. et al. MARSHAL, a novel tool for virtual phenotyping of maize root system hydraulic architectures. *in silico Plants* **2**, diz012 (2020).

64. Lasky, J. et al. Genome–environment associations in sorghum landraces predict adaptive traits. *Sci. Adv.* **1**, e1400218 (2015).

65. Zomer, R. J., Trabucco, A., Bossio, D. A. & Verchot, L. V. Climate change mitigation: a spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agric. Ecosyst. Environ.* **126**, 67–80 (2008).

66. Kalnay, E. et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–472 (1996).

67. New, M., Lister, D., Hulme, M. & Makin, I. A high-resolution data set of surface climate over global land areas. *Clim. Res.* **21**, 1–25 (2002).

68. Hengl, T. et al. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* **12**, e0169748 (2017).

69. Shangguan, W., Dai, Y., Duan, Q., Liu, B. & Yuan, H. A global soil data set for earth system modeling. *J. Adv. Model. Earth Syst.* **6**, 249–263 (2014).

70. Sanchez, P. A., Palm, C. A. & Buol, S. W. Fertility capability soil classification: a tool to help assess soil quality in the tropics. *Geoderma* **114**, 157–185 (2003).

71. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. *Science* **339**, 940–943 (2013).

72. Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Soft.* **36**, 1–13 (2010).

73. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).

74. Gates D. J. et al. Single-gene resolution of locally adaptive genetic variation in Mexican maize. Preprint at https://doi.org/10.1101/706739 (2019).

75. de Leeuw, C. A. et al. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

76. Caye, K. et al. TESS3: fast inference of spatial population structure and genome scans for selection. *Mol. Ecol. Resour.* **16**, 540–548 (2016).

77. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

78. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

79. Browning, B. L. & Browning, S. R. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).

80. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

81. Li, Q. et al. CRISPR/Cas9-mediated knockout and overexpression studies reveal a role of maize phytochrome C in regulating flowering time and plant height. *Plant Biotechnol. J.* **18**, 2520–2532 (2020).

82. Yu, P., Eggert, K., von Wirén, N., Li, C. & Hochholdinger, F. Cell type-specific gene expression analyses by RNA sequencing reveal local high nitrate-triggered lateral root initiation in shoot-borne roots of maize by modulating auxin-related cell cycle regulation. *Plant Physiol.* **169**, 690–704 (2015).

83. Yu, P. PengYuMaize/GlobalSeminalRoot. *Zenodo* https://doi.org/10.5281/zenodo.10985812 (2024).

Acknowledgements

Author contributions

Competing interests

Additional information

nature portfolio

Corresponding author(s): Peng Yu, Ruairidh Sawers, Tianyu Wang, Frank Hochholdinger (lead contact)

Last updated by author(s): Apr 3, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All raw seminal root phenotyping data, geographical coordinates and soil modelling data is provided in Supplementary Tables 1-8. All germplasm information i.e. The geographically diverse teosinte accessions, maize traditional varieties and inbred lines were contributed by the North Central Regional Plant Introduction Station (NCRPIS), the International Maize and Wheat Improvement Center (CIMMYT) and the Chinese maize seed germplasm bank at the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (China) used in this study are provided in the Supplementary Datasets 1-5. Geographical coordinates and elevation information of the collection sites for maize traditional varieties were retrieved from the public database of the U.S. National Plant Germplasm System (https://www.grin-global.org/). Soil and climate data were collected from WorldClim (https://worldclim.org), the NCEP/NCAR reanalysis project (https://psl.noaa.gov/data/reanalysis/reanalysis.shtml), NASA SRB (https://asdc.larc.nasa.gov/project/SRB), Climate Research Unit (CRU) (https://www.uea.ac.uk/groups-and-centres/climatic-research-unit), SoilGrids (https://soilgrids.org/ v2) and the Global Soil Dataset (GSD) (https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/).
Data analysis	MRI-PET images of root systems were reconstructed with PRESTO (v3.01) and normalizing view angle, contrast and brightness in all images with MeVisLab (version 3.4, MeVis Medical Solutions AG). Seed embryo image processing and volumetric analysis were performed using MATLAB (vR2019b) software and AMIRA software (Amira3D 2022.1). The area of embryo and endosperm were detected and quantified by LemnaTec analytical software (v1). Linear mixed models were fitted using software ASReml-R (v4.0). Feature selection for eGWAS was employed using Boruta::boruta() (v7.0.0). Random Forest models were built using RandomForest::randomForest() function under default parameters in R (v4.2.2). Boxplots generated by BoxPlotR (http://shiny.chemgrid.org/boxplotr/). Statistical analyses by R (v4.1.0). The rhizobox RGB images were analysed by RootPainter (v0.16), Root System Analyzer (v2.0), and SmartRoot (v4.21). We created virtual root system representation for each variety using the CPlantBox model (v2.0) and computed individual root and root system hydraulic parameters with the MECHA (v2.1), GRANAR (v1.1) and MARSHAL (v1.1) models. The GBS sequencing data was aligned with the maize B73 reference genome (v4) through Burrows-Wheeler Aligner (BWA) (https://bio-bwa.sourceforge.net/) software. Single nucleotide polymorphisms (SNPs) were

detected by SAMTOOLS (v1.12). The missing genotypes were then imputed with Beagle 4 (v5.2) with default parameters. Population structure was conducted with ADMIXTURE tool (v1.3.0) with default parameters. A genome-wide association study (GWAS) was performed using a mixed linear model in the software EMMAX (Version emmax-intel-binary-20120210). Maize genome resequencing data of Gaspé flint introgression panel and root RNA sequencing data were deposited in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the BioProject ID PRJNA1095206. Source data of main figures are provided with this paper.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Maize genome resequencing data of Gaspé flint introgression panel and root RNA sequencing data were deposited in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the BioProject ID PRJNA1095206. Source data of main figures are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No specific statistical methods were used to determine sample size for the experiments. Instead, we used to date the largest datasets of maize germplasm including of 173 teosinte accessions, 4868 different maize landraces cover the global domestication area of maize and 4049 inbred lines cover the global maize germplasm pools. All of these genotypes fully reflect maize genetic diversity, were chosen for our main study. All detailed information was provided with supplemental dataset and figures.
Data exclusions	We exclude the genotypes which did not germinate or had severe fungal contamination issue.
Replication	The global investigation of maize inbred lines were explored by 10 individual seeds and landraces were explored by 20 individual seeds. All the other experiments were performed once with at least three independent biological replicates (See Supplemental Figures).
Randomization	Seeds of different maize genotypes (inbred lines or landraces) were randomly selected and grown under same conditions.
Blinding	All seed and root samples for all maize genotypes were harvested blindly with only sample numbers without access to the genotype identities. Data analysis was blinded to genotypes and allowing them to choose each subject blindly. Data on imaging, morphological and physiological analyses were acquired by personnel without genotype/treatment information.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
☒	☐ Demonstrate how to render a vaccine ineffective
☒	☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐ Increase transmissibility of a pathogen
☒	☐ Alter the host range of a pathogen
☒	☐ Enable evasion of diagnostic/detection modalities
☒	☐ Enable the weaponization of a biological agent or toxin
☒	☐ Any other potentially harmful combination of experiments and agents