Local selection shaped the diversity of European maize landraces.

Margarita Takou[*], **Kerstin Schulz**[*,‡] **and Markus G Stetter**[*,‡,1]
[*]Dept. of Plant Sciences, University of Cologne, Cologne, Germany, [‡]Cluster of Excellence on Plant Sciences, University of Cologne, Cologne, Germany

The introduction of populations to novel environments can lead to a loss of genetic diversity and the accumulation of deleterious mutations due to selection and demographic changes. The introduction of maize to Europe represents a recent range expansion. We investigate how the introduction history shaped the genetic diversity and differentiation of European traditional maize populations and quantify the impact of its recent range expansion and consecutive breeding on the accumulation of genetic load. We use genome-wide genetic markers of almost 2,000 individuals from 38 landraces, 155 elite breeding lines and large DH lines of two landraces and find three main clusters, in line with previously suggested diffusion routes of maize within Europe. On the finer scale we find extensive population structure within European maize, with landraces being highly differentiated even over short geographic distances. Yet, diversity change does not follow the continuous pattern of range expansions. Landraces maintain high genetic diversity distinct between populations, which does not decrease along the possible expansion routes. Signals of positive selection in European landraces that overlap with selection in Asian maize suggest convergent selection during maize introductions. Consistent with the maintenance of high diversity, we find no evidence of genetic load accumulating along the maize introduction route in European maize. However, modern breeding likely purged highly deleterious alleles in elite germplasm. Our results reconstruct the history of maize in Europe and show that landraces have maintained high genetic diversity that could reduce genetic load in the European maize breeding pools.

KEYWORDS European maize, adaptation, range expansion, genetic diversity, population structure, genetic load

Introduction

Species distributions are the result of range expansion to new environments, such as the post glacial colonization of Northern Europe (Hewitt 2000). These dynamic population genetic processes have a strong influence on the genetic diversity of the expanding species (Excoffier *et al.* 2009), including a decrease in the genetic diversity due to increased drift and the subsequent genetic differentiation among the newly established and core populations (Austerlitz *et al.* 1997; Excoffier *et al.* 2009; Slatkin and Excoffier 2012; Wang *et al.* 2017; Takou *et al.* 2021). At the same time, deleterious mutations accumulate at the front of the expansion range creating the genetic load (de Pedro *et al.* 2021; González-Martínez *et al.* 2017; Peischl *et al.* 2013). As locally adapted populations move to novel environments (Colautti and Barrett 2013; Savolainen *et al.* 2013), the adaptive potential of populations at the range edges can be compromised (Excoffier *et al.* 2009). Rapid range expansions are expected to become more frequent as climate change alters the natural environments, which will push species to new suitable conditions outside their original range (Waldvogel *et al.* 2020). Crops have spread rapidly around the globe and likely experienced the effects of range expansion (Huang *et al.* 2022). They could have encountered novel selective pressures caused by the different climate, soil compositions, and ecological dynamics of the new regions, as well as the human-driven selection for agricultural traits (Purugganan and Fuller 2009; Purugganan 2019). This interplay between natural selection and human-driven intervention ultimately resulted in the appearance of local traditional crop varieties (Huang *et al.* 2022).

One of the most important crops worldwide is maize, which has shown great adaptability to locations outside its initial range with one of the broadest cultivated ranges of all crops today (Tenaillon and Charcosset 2011). Maize domestication started approximately 9,000 years ago in a small region in Mexico, where the wild grass species teosinte, (*Zea mays ssp parviglumis* and *Zea mays ssp mexicana*), have given rise to the crop we know today (Tenaillon and Charcosset 2011; Yang *et al.* 2017b, 2023). Over time, maize spread throughout the Americas, facilitated by human movement (Kistler *et al.* 2018). The early dispersion process was characterized by two routes: one northward to the USA and Canada, and another extending southward to South America and the Caribbean coast (Tenaillon and Charcosset 2011; Kistler *et al.* 2018). This human mediated range expansion has led to decreased genetic diversity, introgression events and local adaptation across the colonization route (Wang *et al.* 2017; Hufford *et al.* 2012), which creates the opportunity to study range expansion and the dynamics of selection and decreasing genetic diversity. Hence, crops such as maize might have experienced similar range expansion forces that are predicted for natural populations.

European maize landraces offer a compelling system to explore the recent introduction and expansion at potential range edges. While extensive research has been done on the evolution, domestication, and traditional use of primary American maize

[1] Dept. of Plant Sciences, University of Cologne, Cologne, Germany E-mail: m.stetter@uni-koeln.de

landraces (Hufford *et al.* 2012; Yang *et al.* 2023; Wang *et al.* 2017), the secondary range expansion and development of their European counterparts have been little investigated (Revilla Temiño *et al.* 2003; Tenaillon and Charcosset 2011). Unlike the gradual expansion outside of the domestication area, maize's arrival in Europe is characterized by an abrupt introduction to a novel environment around 500 years ago (Janick and Caneva 2005; Brandenburg *et al.* 2017). Moreover, only a relatively small number of plants were initially brought from the Americas, likely resulting in a strong bottleneck (Tenaillon and Charcosset 2011). After its first introduction into Europe from the Caribbeans through Spain, there is evidence for further diffusion routes via France and Italy to the rest of Europe, as well as a secondary introduction from North American to Central Europe (Brandenburg *et al.* 2017; Leng *et al.* 1962; Rebourg *et al.* 2003). Additionally, the varying local European environments likely required local adaptation, as has been observed in Mexican maize populations (Tittes *et al.* 2021). Historical accounts state that by the middle of the 16th century maize fields were present across all Europe (Finan 1948; Janick and Caneva 2005; Brandolini and Brandolini 2009), making European maize landraces an invaluable resource for understanding the rapid introduction to new environments.

Here, we study the genetic diversity across Europe and how it was potentially shaped by the multiple introductions of maize to the continent. Using a comprehensive dataset of traditional varieties spanning the range of European maize, we were able to show that European landraces are highly diverse but genetically differentiated. We detect two major clusters of landraces within Europe that might depict their introduction history. In addition we identify local clusters of similarity between traditional varieties, likely reconstructing historical trade routes. Selection scans in two landrace DH line libraries with almost 1,000 individuals reveal genomic regions that were likely under selection in individual landraces and suggest potential adaptation signatures within Europe. However, European maize does not show signatures of range edge decreased diversity and accumulated genetic load. Together, our results give insights into the history of the rise of European maize in the last centuries.

Materials and Methods

Dataset composition and preparation

All data used in the present study are publicly available (Unterseer *et al.* 2016; Mayer *et al.* 2017, 2020). In brief, all individuals used were genotyped at high density with the 600k Affymetrix® Axiom® Maize Array, which is comprised of 616,201 variants, of which 6,759 represent insertions/deletion (Unterseer *et al.* (2014, 2016)). Specifically, 941 DH lines derived from 2 European landraces were taken from Mayer *et al.* (2020), which were divided as follows: 501 DH lines of Kemater Landmais Gelb (KL) and 409 of Petkuser Ferdinand Rot (PE). 952 individuals from 35 European maize landraces from Mayer *et al.* (2017). Lastly, 3 landraces from Unterseer *et al.* (2016) and 155 elite lines were added from Unterseer *et al.* (2016). The final panel had in total 2,954 individuals (Table S1).

The merging of the three data sets (Unterseer *et al.* (2016); Mayer *et al.* (2017, 2020) was performed using custom python scripts and all data sets were formatted in the HapMap format using Tassel 5.0 GUI (command: *./start_tassel.pl -Xmx4g*) (Bradbury *et al.* (2007)). Post conversion the IUPAC nucleotide codes in the converted files were replaced with the corresponding nucleotide bases as stated in the user guide. For merging, the Unterseer *et al.* (2016) data sets were updated to AGPv4 maize reference genome. All insertions/deletions in the datasets were removed, which resulted in 6,759 and 6,752 markers removed from the Mayer and Unterseer data sets, respectively. For quality control, we checked if the alleles were correctly merged, based on the individual marker information provided by the manufacturer. Additionally, only markers that had quality of "PolyHighResolution", "MonoHighResolution" or "NoMinorHom" across all datasets were kept, resulting in 419,477 SNPs.

Population Structure and Genetic Diversity analysis

To investigate the underlying patterns of variation in this maize dataset a Principal Component Analysis (PCA) was conducted using the scikit allel package (Miles *et al.* (2023)) for python v3.7. We sub-sampled the dataset to exclude the DH lines and reduce the individuals to 1,253 individuals divided into 43 populations for the PCA. In order to do the PCA, using the scikit allel package, singletons were removed from the sub-sampled dataset and linkage disequilibrium (LD) filtering/pruning was done using 5 iterations, a window size of 500bp, and a step size of 200bp. The PCA was then performed using the function *allel.pca()* (Miles *et al.* (2023)) and plotted using the matplotlib package (Hunter (2007)). We further investigated the clustering of the data by using the software STRUCTURE Pritchard *et al.* 2000. We run the analysis for up to K=40 and used the cross validation error (cv) to decide the optimal number of clusters.

Using scikit allele package (Miles *et al.* (2023)), all pairwise F_{ST} values between all the 35 European landrace population pairs from Mayer *et al.* (2017)) were estimated based on segregating SNPs. We calculated Hudson's F_{ST} using the *allel.blockwise_hudson_fst()* function with a block size of 100,000 variants(Smaragdov and Kudinov (2020); Bhatia *et al.* (2013); Hudson *et al.* (1992). To visualize the population differentiation patterns, the calculated Hudson F_{ST} values were plotted with a hierarchical clustered heatmap, using the package ComplexHeatmap in R v4.1.2 (R Core Team (2021); Gu *et al.* (2016)).

We estimated mean expected heterozygosity (H_{EXP}) per population for all landraces to investigate the distribution of genetic diversity and potential patterns of diversity among European landraces across the diverse European landscape. We calculated the expected heterozygosity using the *allel.heterozygosity_expected()* function and estimated the mean using the *numpy.nanmean()* function (Harris *et al.* (2020)). The geographical map including the H_{EXP} values was created using the plotly package (Inc. (2015)). The sample locations (longitude and latitude values from Mayer *et al.* (2017)) were used as proxies for each population's distribution. We estimated the geographic distance in kilometers of each landrace, to the two populations that are approximately the closest to the expected entry points in Europe; the population Tuy (TU; 42.04N, -8.64W) marks the entry point in Spain and the population Barisis (BA; 49.57N, 3.328E) in Central Europe. For the populations Altreier (AL), Fleimstal (FL) and Sornay (SO), whose coordinate information was missing, we used the best approximation of their locations. The inbreeding coefficient of each individual was estimated using vcftools (Danecek *et al.* 2011).

Spatial gene-flow analysis of population genetic data and selection scans

The Fast Estimation of Effective Migration Surfaces (FEEMS, version 1.0.0) analysis was conducted to investigate non-homogeneous gene flow between the European landraces (Marcus *et al.* (2021)). SNPs with more than 20% missing data and

monomorphic sites were removed with bcftools (Li 2011) and the input files were prepared with plink2 Chang *et al.* (2015). We used the FEEMS imputation feature for missing data. Additionally, the script was run with the options `scale_snps=True` and `translated=False` and the provided smaller grid size file `grid_100.shp`. A custom smoothing regularization parameter (λ = 100) was estimated from a cross-validation procedure using their provided `cross-validation.ipynb`.

We searched for signals of selection across the genomes of the double haploid lines generated for Petkuser Ferdinard Rot (PE DH) and Kemater Landmais Gelb (KL DH) as those are naturally phased datasets with approximately 500 individuals each. We used selscan Szpiech and Hernandez 2014; Szpiech 2024 to estimate the intergrated haplotype score (iHS). For each SNP, we estimated an empirical p-value by dividing its rank by the number of SNPs. We classified SNPs as being under selection when their p-value was below 0.05 and their absolute iHS value was greater than 2. Then, we extracted all the genes present in 10kbp upstream and downstream regions of the selected SNPs and tested for Gene Ontology enrichment using the online database agriGO v2.2 (Tian *et al.* 2017).

Moreover, we estimated genetic load per individual for the landrace populations and elite pools to estimate the accumulation of deleterious mutations. We used the Genomic Evolutionary Rate Profiling (GERP) scores from Kistler *et al.* (2018). To estimate genetic load, for each individual belonging to one of the landraces or elite pools, we scored each site based on the number of derived alleles they have and then, summed up the scores across all sites. We excluded 100 individuals that had more than 1.69% of missing sites (95th quantile of the distribution), in order to exclude potential differences in load due to missing information. Additionally, we estimated the fixed and segregating load within the dent elite and flint breeding pool, as well as for the dent and flint landraces. Finally, we counted the number of highly deleterious alleles within each sample. We defined as 'highly deleterious' all sites with GERP score larger than 5. We compared the genetic load distributions between the different groups with a Kolmogorov-Smirnov test.

Results

Genetic diversity is similar across European maize.

Maize was likely introduced multiple times into Europe through different routes (Brandenburg *et al.* 2017). This introduction history and the consecutive expansion should be reflected in the population structure of European traditional varieties. We assessed the population structure within European maize of 1,253 individuals, which were divided into 38 traditional populations and modern elite inbreds of the two heterotic pools, flint and dent, representing a diverse sampling of the European maize. We perfomed a a PCA based on 491,477 genome-wide SNP markers. The PCA showed the integrity of landrace populations, as individuals from the same landrace grouped together (Figure 1a). The first two principal components explaining 7.1% of the variation, mostly separating the populations into a Central European cluster, and a Southern European cluster. Additionally, PC2, which explained 2.8% of variation, clearly separated the flint (Europe) and dent (US and Europe) elite lines. It has been shown before that the flint and dent pools are genetically differentiated, despite sharing haplotypes (Brown and Anderson 1947; Unterseer *et al.* 2016; Haberer *et al.* 2020). This reflects the long divergent germplasm pools of the two heterotic groups. Brown and Anderson (1947); Unterseer *et al.* (2016). Moreover, there are 2 landrace populations, Polnischer Landmais (PL) and Altreier (AL), which cluster with the dent instead of the flint heterotic group. The existence of multiple well defined populations within the European maize landraces was also supported by a STRUCTURE analysis of the individuals (Figure S2). The STRUCTURE supports the presence of 35 (cv = 0.38024) separate populations (Mayer *et al.* 2017), with varying levels of shared ancestry between the populations.

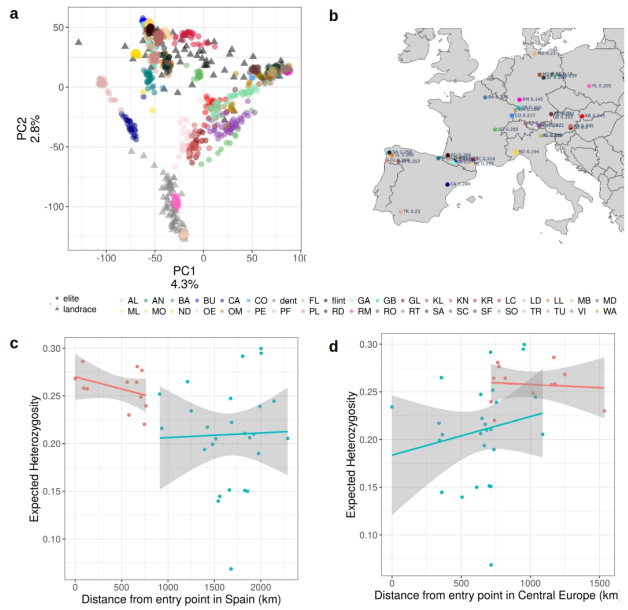

Figure 1 Population diversity of European maize. a. First two PCs of the PCA analysis on all 38 landraces and 2 modern elite pools. The color of the dots corresponds to a population, while the shape to whether a population is a landrace (circle) or elite breeding line (triangle). b. The geographic distribution of the landraces in Europe. For each population it's mean H_{Exp} is provided next to the label. The mean expected heterozygosity is plotted against the distance from the potential introduction points in Europe via c. Spain and d. Central Europe. The populations are colored based on their longitude and whether they are in East (blue) or West (orange).

We estimated genetic diversity using expected heterozygosity (H_{Exp}) per site within each population (Figure 1b). The median H_{Exp} was 0.232 and mean H_{Exp} was 0.224 across all populations (Figure S1a). The highest H_{Exp} was observed within the landrace Gleisdorfer (GL) and lowest for Fleimestal (FL), with H_{Exp} 0.299 and 0.068, respectively. The elite breeding lines had the highest inbreeding coefficients (median F_{IS}= 0.992), but only a few landraces had high inbreeding coefficients (Figure S1c). Surprisingly, the populations in Eastern Europe showed higher variation in inbreeding than in the West (Figure 1c).

In line with the theory of range expansion, we looked for patterns of decreased H_{Exp} across the geographic distance from potential introduction points to study the impact of maize range expansion within Europe. We estimated the Haversine geographical distance between the populations closest to the potential entry points of maize in Europe and the rest of the populations. We took the population Tuy (TU) in Spain as the closest population to the entry for introduction via Spain and the coordi-

nates of the population Barisis (BA) as the closest population to the second entry point in Central Europe as described in Tenaillon and Charcosset 2011. The H_{Exp} was significantly negatively correlated with the geographic distance from Tuy (p = 0.024, ρ = -0.366), which might reflect a pattern of declining genetic diversity along the range expansion route. However, the opposite pattern was observed for the Northern route from Barisis (p = 0.00099, ρ = 0.518). There was no correlation of H_{Exp} and distance within the South Eastern and Central European cluster of populations.

Populations and boundaries of gene flow across Europe.

The calculated F_{ST} values, which ranged from 0 to 0.531, with the mean and median of 0.255 and 0.257, respectively, support the presence of separate landrace populations (Figure 2a). The Polnischer Landmais (PL) population, which clustered with the dent elite lines in the PCA, has on average high F_{ST} values with the other populations (median of 0.352), corroborating the results of the PCA.

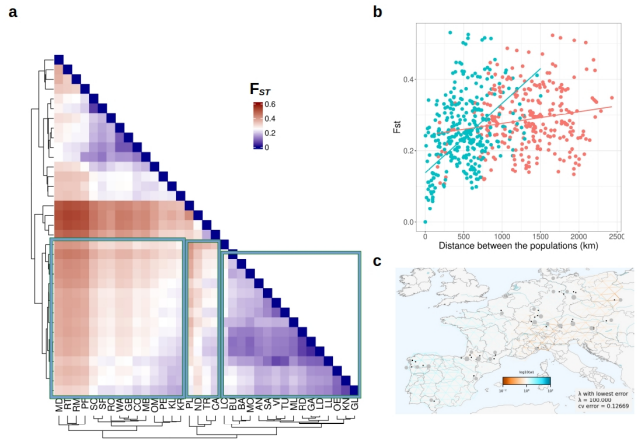

Figure 2 Population structure is prevalent between the European maize landraces. a. Pairwise F_{ST} values between all 38 landraces. The 3 clusters formed based on the F_{ST} values are highlighted with the green boxes. b. The F_{ST} values between each pair of populations are plotted against the distance between the populations. The pairs that are within the same cluster, either the South Eastern or Central European one, are colored in blue, while the pairs belonging to different clusters are shown in red. c. The effective migration rate within Europe is shown. The grid is colored blue when the estimated effective migration rate is lower than expected and red when it is higher than expected.

The three meta-population clusters, each of which includes landraces from South or from Central Europe are also definable by the F_{ST} values. On average, the F_{ST} values were significantly lower (Kolmogorov-Smirnov test p = 0.0031; Figure S3) between populations within the same cluster (mean of 0.267 and median of 0.257) than between populations between the two clusters (mean of 0.287 and median of 0.2794). A third smaller cluster of populations was observed within the Central European one, which consists of Polnischer landmais (PL; 51N, 18.5E), Nostrano dell Isola (ND; 44.61N, 8.24E), Tremesino (TR; 37.39N, -5.63W) and Castellote (CA; 40.79N, -0.39W). The genetic differentiation between populations increases with the distance between them (p < 2.2e-16, ρ = 0.409; Figure 2b). This pattern was stronger within the South Eastern European group (p = 9.665e-14, ρ = 0.721), than the Central European cluster of populations (p = 1.179e-13, ρ = 0.426), even though the median F_{ST} value between the South Eastern European populations was 0.136. In contrast, the Central European populations had a more defined population structure, with median F_{ST} of 0.253.The strongest genetic differentiation was observed between the landrace populations of Rheintaler St. Gallen (RT) and Nostrano dell Isola (ND; F_{ST} = 0.531), Rheintaler Monsheim (RM) and Nostrano dell Isola (ND; F_{ST} = 0.525), as well as between Rheintaler St. Gallen (RT) and Tremesino (TR; F_{ST} = 0.523) despite their geographic proximity to each other (approximately 324km and 559km between RT/ND and RM/ND, respectively). These examples demonstrate the complex local history of maize in Europe.

To further investigate the geographic patterns of diversity, we used the FEEMS analysis, which estimates effective migration surfaces (λ = 100; cv error = 0.126; Figure 2c). The effective migration rate was higher than neutral (log10(w) > 1) within the Iberian peninsula, with a barrier to migration present along the Pyrenees (log10(w) <1), reducing the gene flow between the Iberian peninsula and Central Europe (Figure 2c). Within the Central European cluster of populations, there were surfaces with lower effective migration (log10(w) <1). Those areas overlap with the locations of the Rheintaler Monsheim (RM), Nostrano dell Isola (ND) and Rheintaler St. Gallen (RT), which also have high F_{ST} values between them (Figure 2c).

Evidence for selection within two European maize landraces

The observed population structure within Europe could have been the result of selection pressures at the new environment. We scanned the genome of two double haploid (DH) line libraries derived from the Petkuser Ferdinand Roth (PE) and Kemater Landmais Gelb (KL) landraces (Mayer *et al.* 2020) for signals of recent positive selection. The iHS statistics measures the relative extended haplotype homozygosity between alleles and indicates signatures of recent selection. Within both Petkuser Ferdinand Roth (PE) DH and Kemater Landmais Gelb (KL) DH libraries, we detected SNPs with elevated absolute iHS value (Figure 3). In total, we found 53 and 3 significant markers for Kemater Landmais Gelb (KL) DH and Petkuser Ferdinand Roth (PE) DH respectively (p-value < 0.05; Table S2). In the Petkuser Ferdinand Roth (PE) DH, we found signatures of selection on chromosomes 4 and 10, while significant sites in the Kemater Landmais Gelb (KL) DH were distributed along all chromosomes. The 10kbp upstream and downstream of the significant markers in the Kemater Landmais Gelb (KL) DH included 66 genes. These were significantly enriched (corrected p < 0.05) for the Gene Ontology functions 'response to metal ion', 'single-organism cellular process', 'single organism process' and 'response to cadmium ion'. Moreover, the markers on chromosome 1 overlapped with previously identified selective sweep regions related to maize breeding in the US and China (Huang *et al.* 2022). Only nine genes were located within 10kbp distance upstream or downstream from the identified selection signal in the Petkuser Landmais Gelb (PE) DH (Table S3), but we did not find overlapping regions under selection in the two populations. Hence, this might reflect selection based on the local environment rather than the introduction to Europe or the general breeding process.

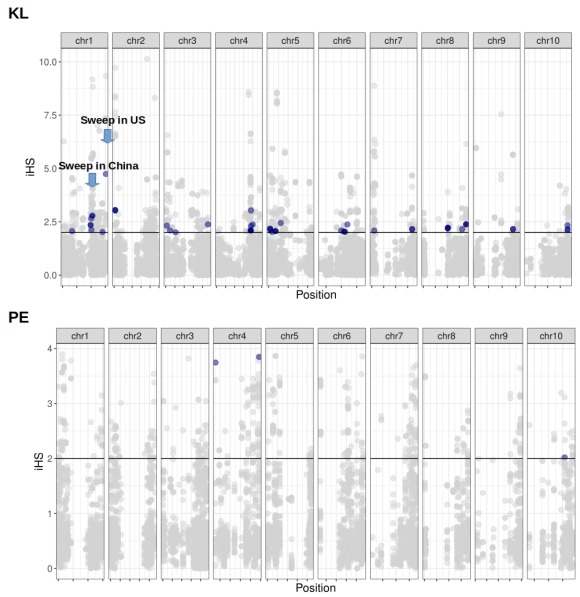

Figure 3 Recent selective sweeps in two European maize double haploid (DH) populations are rare. The absolute iHS values across the genome are plotted for the Kemater Landmais Gelb (KL) DH (top) and Petkuser Landmais Gelb (PE) DH (bottom) populations. In blue are highlighted the positions that are significant for being under positive selection (empirical p value cut-off of 0.05). In the Kemater Landmais Gelb population the two locations that have been previously identified as under selection in Chinese or US breeding lines as marked.

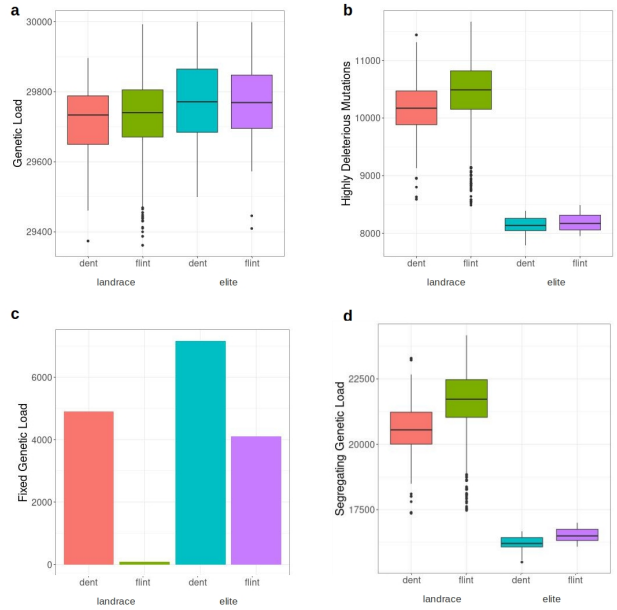

Figure 4 Purging of deleterious genetic load within the elite breeding lines. a. The genetic load estimated as the sum of the GERP scores for all sites within the dent germplasm landraces, flint germplasm landraces, flint germplasm elite lines and dent germplasm elite lines is given. b. The highly deleterious genetic load is estimated by counting the alleles with GERP score larger than 5 within each group. c. The fixed genetic load for each population is given as the sum of the GERP scores for those sites which have the derived allele fixed within each group of maize lines. d. The segregating genetic load is presented for each of the four groups as the sum of the GERP scores of the segregating sites.

Change of genetic load along the introduction routes and due to breeding

Range expansion has been previously associated with the accumulation of deleterious mutations and overall genetic load (de Pedro *et al.* 2021; Peischl *et al.* 2013). We estimated the genetic load per population, including the elite breeding pools, using the GERP scores from (Kistler *et al.* 2018). Estimates of genetic load per population differed significantly ($p < 2.2e-16$), with the population Rheintaler Monsheim (RM) having the highest median genetic load (sum of GERPs = 29,881.23), while the population with the least accumulated genetic load (sum of GERPs = 29,578.20) was Bugard (BU) (Figure S4). We further tested for correlation between accumulated genetic load and the distance from the two hypothetical introduction points in Southern and Central Europe and found significant correlations of $\rho = -0.16$ ($p = 1.18e-11$) and $\rho = -0.156$ ($p = 3.845e-11$), respectively.

Furthermore, we tested the differences in genetic load between landrace and elite groups, as inbreeding, drift and selection during breeding might have changed genetic load. We grouped both the landraces and the elite lines based on their germplasm origin of either dent or flint (Figure 4a). Genetic load was significantly higher in both elite pools than their counterpart landrace group, with p values of 0.0486 and 0.01346 for flint and dent germplasm respectively. The accumulating genetic load was also different between the dent landrace germplasm lines and the flint elite germplasm lines ($p = 0.0162$). Those differences were not observed for highly deleterious mutations. Both elite germplasm pools had a significantly ($p < 2.2e-16$) lower number of highly deleterious mutations than the landraces (Figure 4b), suggesting potential purging during breeding. In contrast, fixed load was significantly higher ($p < 2.2e-16$) in the elite lines than the landraces (Figure 2c), while the segregating load was significantly higher ($p < 2.2e-16$) in the landraces (Figure 2d). Those differences were observed irrespective of whether the lines belonged in the flint or dent germplasm.

Discussion

Genetic diversity is shaped by local factors rather than the introduction route.

Maize has been introduced to Europe (Janick and Caneva 2005) within the last 500 years and its range expanded rapidly throughout the European continent. Evidence exists for two introductory routes to Europe; one from the Caribbean via Southern Europe and one to Central Europe with material from North America (Tenaillon and Charcosset 2011; Brandenburg *et al.* 2017). Along this expansion route, local openly pollinating traditional varieties have formed and have been kept and traded by local farmers for centuries. Indeed, the population structure and relatedness between samples of these traditional varieties trace the history of maize in Europe. We used a large sample of European maize landraces to explore how the introduction and

expansion of maize in Europe has shaped genetic structure and diversity. The wide geographic sampling and representation of each landrace population enabled us to track the impact of colonization of novel environments in a more detailed scale. The analysis of genetic diversity within European landrace populations showed three major clusters across populations (Figure 1). One cluster, which includes the dent elite lines, show that 4 landraces are most likely derived from North American dent populations, while all other landraces clustered with the flint elite lines, confirming their flint ancestry and their involvement in the creation of European maize hybrid pools (Mayer *et al.* 2020; Tenaillon and Charcosset 2011; Brandenburg *et al.* 2017). The further separation into two groups within the flint cluster across Europe might represent the two introduction routes of maize into Europe (Tenaillon and Charcosset 2011).

Range expansion can have a strong influence on the genetic diversity of the expanding species and decrease genetic diversity and the increase genetic differentiation among the newly established and core populations (Excoffier *et al.* 2009). Despite the pattern of introduction routes and ancestry within Europe, we did not observe a pattern of decreasing genetic diversity along the potential introduction routes. We used the potential introduction locations of maize from Brandenburg *et al.* (2017) to investigate the impact of expansion along the separate routes, but did not find a pattern that would be expected for the colonization by a natural species (Zimmermann *et al.* 2014; Smith *et al.* 2020), which indicates the maintenance of genetic diversity across the continent. On the contrary, genetic diversity is maintained in similar levels across the studied maize populations, suggesting a stronger influence of local maintenance and potential adaptation. Within Europe, the observed pattern is different from what has been observed within the Americas, where a decrease in genetic diversity has been documented (Huang *et al.* 2022). American maize landraces have followed expansion route from Central Mexico towards North and South America, enabling tracking the changes in genetic composition along the route (Tenaillon and Charcosset 2011; Huang *et al.* 2022; Yang *et al.* 2023). In Europe, however, the pattern is less clear and suggests long range dispersal and a more rapid spread through the continent (Figure 1 and Figure 2). In potato and European bean, admixture and different admixture histories in the introduced range have been suggested as reason for maintainance of diversity in introduced ranges but often little population structure is observed within a smaller geographic region, such as the Nordic countries (Ortiz *et al.* 2023; Gutaker *et al.* 2019; Ames and Spooner 2008; Bellucci *et al.* 2023). Admixture between Central and Southern European maize populations might be prevalent (Bradbury *et al.* 2007; Mayer *et al.* 2017; Unterseer *et al.* 2016), however, we find that not only there is population structure dividing the landraces into multiple geographic distinct populations, but also there is an evident pattern of isolation by distance. We identify smaller local genetic clusters between landraces, including one along the Rhine and one into northern Germany that consisted of four populations (Figure 2). The F_{ST} values between those populations and their neighboring one is up to 0.5 (Figure 2). This sharp distinction likely represents historical trade routes and political and linguistic borders in the area during introduction of maize to Europe. Patterns of maintained genetic diversity and population structure in the short evolutionary times might be related to local environmental adaptation, but have also been connected to human activity in other crops (Bellucci *et al.* 2023). For instance, human activity has facilitated the geographic spread and genetic differentiation of rice(Courtois *et al.* 2012). Within Southern Europe, the landrace variability and genetic structure of tomato landraces, which were introduced from South America at a similar time as maize, are related to the fruit type (Corrado *et al.* 2014; García-Martínez *et al.* 2013), a highly coveted breeding trait.

Genomic regions selected during maize introduction

We used two very large populations of DH lines derived from diverging landraces to infer signals of recent selection in the landraces. The advantage of such a population is their natural phasing through the DH process, which removes problems with computational phasing in highly heterozygous outcrossing maize. We were able to detect signals of positive selection within the Central European Petkuser Ferdinard Rot (PE) landrace and Kemater Landmais Rot (KL) landrace. We found 53 and 3 potentially selected loci in each population but only a small overlap in selection signals, confirming the importance of local selection in the formation of these local landraces. In addition, we found a partial overlap of detected selection signals in one of the two populations with selective sweeps related to breeding efforts in American and Asian breeding maize lines (Huang *et al.* 2022), suggesting convergent selection targets in different regions of the world. Additional selection signals might be linked to the adaptation to the novel environments and subsequent breeding targets that maize encountered after its arrival in Europe.

Even though we have only tested selection in two populations, the traces of selection corroborate the results of population structure within European maize. Local adaptation can be achieved by changes without large changes in allele frequencies within populations (Le Corre and Kremer 2012), especially for polygenic traits such as the ones observed here. Regions of high genetic differentiation have been found to harbor genetic variation linked to adaptation to local climate in various plant species (De La Torre *et al.* 2019) and it can occur rapidly during range expansion (Colautti and Barrett 2013), even if there is there extensive gene flow (Hämälä 2018).

Local purging of genetic load due to inbreeding and selection

Finally, we estimated the genetic load within the European landrace populations. In maize, the deleterious mutations accumulate within complex phenotypes of interest, even though, in overall, they are maintained in low frequencies (Mezmouk and Ross-Ibarra 2014). However, accumulation of genetic load during range expansion has been observed in the Americas (Wang *et al.* 2017). Here, we observed that genetic load within the European landrace populations is variable but does not follow a strong pattern of accumulation in any cluster. Population subdivision, similar to the one observed here, could maintain the genetic load in similar levels (Glémin *et al.* 2003), though the observed migration between the different landraces could have also helped to alleviate the accumulation of load across the expansion route (Theodorou and Couvet 2006). Additionally, the introduction of maize in Europe has been relative recent and therefore the crop spread throughout the European continent relative fast. This fast range expansion could have limited genetic load accumulation (Gilbert *et al.* 2018), leading to the absence of a correlation between load and geographic distance from the possible introduction points in Europe. On the other hand, the adaptation to an environmental gradient can promote the accumulation of genetic load (Gilbert *et al.* 2017), maintaining differences between populations as we observe here.

When we compared the landraces to the elite breeding lines in our populations, we found that the number of highly deleterious mutations decreased in the elite lines, while fixed load was significantly higher (Figure 4b and 4c). These results suggest purging of highly deleterious alleles but fixation of mildly deleterious load, potentially due to inbreeding and decreased population size during selective breeding. Inbreeding can effectively remove the deleterious mutations particularly when they are recessive, making it more effective than selection alone (Glémin 2003), which might explain the higher load in landrace populations. Purging of highly deleterious mutations has been shown to decrease load in maize compared to teosinte, reducing inbreeding depression in maize compared to teosinte (Samayoa *et al.* 2021). Moreover, the need for inbred fitness in breeding programs would further favor purging during breeding and can lead to the removal of especially lethal mutations by inbreeding (Wang *et al.* 1999). Yet, the high genetic diversity observed within the landraces might enable masking of deleterious alleles but leads to the higher segregating genetic load. Masking of fixed mildly deleterious alleles in hybrids might contribute to heterosis in European maize hybrids (Yang *et al.* 2017a), but further reduction of fixed genetic load might enable further selective gain.

Taken together our results illustrate the complexity of the introduction of maize to Europe and the individual history of landraces within the continent and provide an example of local populations maintaining a broad genetic diversity.

Data Availability

All used data is publicly available from their respective publications of origin (Unterseer *et al.* 2016; Mayer *et al.* 2017, 2020). All scripts used in the analysis will be available on GitHub upon publication.

Author contribution

MGS conceived the study. MT and KS conducted the analysis. MT and KS prepared figures and tables. MT, MGS and KS wrote the manuscript. All authors discussed the results, edited, and approved the manuscript.

Acknowledgments

We acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC-2048/1 – project ID 390686111 and financial support by the Federal Ministry of Education and Research (BMBF, Germany) under the Plant Breeding Research for the Bioeconomy initiative (funding ID: 031B0195, project MAZE).

References

Ames, M. and D. M. Spooner, 2008 Dna from herbarium specimens settles a controversy about origins of the european potato. American journal of botany **95**: 252–257.

Austerlitz, F., B. Jung-Muller, B. Godelle, and P.-H. Gouyon, 1997 Evolution of coalescence times, genetic diversity and structure during colonization. Theoretical population biology **51**: 148–164.

Bellucci, E., A. Benazzo, C. Xu, E. Bitocchi, M. Rodriguez, *et al.*, 2023 Selection and adaptive introgression guided the complex evolutionary history of the european common bean. Nature communications **14**: 1908.

Bhatia, G., N. Patterson, S. Sankararaman, and A. L. Price, 2013 Estimating and interpreting fst: the impact of rare variants. Genome research **23**: 1514–1521.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, *et al.*, 2007 Tassel: software for association mapping of complex traits in diverse samples. Bioinformatics **23**: 2633–2635.

Brandenburg, J.-T., T. Mary-Huard, G. Rigaill, S. J. Hearne, H. Corti, *et al.*, 2017 Independent introductions and admixtures have contributed to adaptation of european maize and its american counterparts. PLoS genetics **13**: e1006666.

Brandolini, A. and A. Brandolini, 2009 Maize introduction, evolution and diffusion in italy. Maydica **54**: 233.

Brown, W. L. and E. Anderson, 1947 The northern flint corns. Annals of the Missouri Botanical Garden **34**: 1–29.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, *et al.*, 2015 Second-generation plink: rising to the challenge of larger and richer datasets. Gigascience **4**: s13742–015.

Colautti, R. I. and S. C. Barrett, 2013 Rapid adaptation to climate facilitates range expansion of an invasive plant. science **342**: 364–366.

Corrado, G., M. Caramante, P. Piffanelli, and R. Rao, 2014 Genetic diversity in italian tomato landraces: Implications for the development of a core collection. Scientia Horticulturae **168**: 138–144.

Courtois, B., J. Frouin, R. Greco, G. Bruschi, G. Droc, *et al.*, 2012 Genetic diversity and population structure in a european collection of rice. Crop science **52**: 1663–1675.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, *et al.*, 2011 The variant call format and vcftools. Bioinformatics **27**: 2156–2158.

De La Torre, A. R., B. Wilhite, and D. B. Neale, 2019 Environmental genome-wide association reveals climate adaptation is shaped by subtle to moderate allele frequency shifts in loblolly pine. Genome Biology and Evolution **11**: 2976–2989.

de Pedro, M., M. Riba, S. C. González-Martínez, P. Seoane, R. Bautista, *et al.*, 2021 Demography, genetic diversity and expansion load in the colonizing species leontodon longirostris (asteraceae) throughout its native range. Molecular ecology **30**: 1190–1205.

Excoffier, L., M. Foll, and R. J. Petit, 2009 Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics **40**: 481–501.

Finan, J. J., 1948 Maize in the great herbals. Annals of the Missouri Botanical Garden **35**: 149–191.

García-Martínez, S., G. Corrado, J. J. Ruiz, and R. Rao, 2013 Diversity and structure of a sample of traditional italian and spanish tomato accessions. Genetic resources and crop evolution **60**: 789–798.

Gilbert, K. J., S. Peischl, and L. Excoffier, 2018 Mutation load dynamics during environmentally-driven range shifts. PLoS Genetics **14**: e1007450.

Gilbert, K. J., N. P. Sharp, A. L. Angert, G. L. Conte, J. A. Draghi, *et al.*, 2017 Local adaptation interacts with expansion load during range expansion: maladaptation reduces expansion load. The American Naturalist **189**: 368–380.

Glémin, S., 2003 How are deleterious mutations purged? drift versus nonrandom mating. Evolution **57**: 2678–2687.

Glémin, S., J. Ronfort, and T. Bataillon, 2003 Patterns of inbreeding depression and architecture of the load in subdivided populations. Genetics **165**: 2193–2212.

González-Martínez, S. C., K. Ridout, and J. R. Pannell, 2017

Range expansion compromises adaptive evolution in an outcrossing plant. Current Biology **27**: 2544–2551.

Gu, Z., R. Eils, and M. Schlesner, 2016 Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics .

Gutaker, R. M., C. L. Weiß, D. Ellis, N. L. Anglin, S. Knapp, *et al.*, 2019 The origins and adaptation of european potatoes reconstructed from historical genomes. Nature ecology & evolution **3**: 1093–1101.

Haberer, G., N. Kamal, E. Bauer, H. Gundlach, I. Fischer, *et al.*, 2020 European maize genomes highlight intraspecies variation in repeat and gene content. Nature Genetics **52**: 950–957.

Hämälä, T., 2018 Ecological genomics in arabidopsis lyrata: local adaptation, phenotypic differentiation and reproductive isolation .

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, *et al.*, 2020 Array programming with NumPy. Nature **585**: 357–362.

Hewitt, G., 2000 The genetic legacy of the quaternary ice ages. Nature **405**: 907–913.

Huang, X., S. Huang, B. Han, and J. Li, 2022 The integrated genomics of crop domestication and breeding. Cell .

Hudson, R. R., M. Slatkin, and W. P. Maddison, 1992 Estimation of levels of gene flow from dna sequence data. Genetics **132**: 583–589.

Hufford, M. B., E. Martínez-Meyer, B. S. Gaut, L. E. Eguiarte, and M. I. Tenaillon, 2012 Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. PLoS One **7**: e47659.

Hunter, J. D., 2007 Matplotlib: A 2d graphics environment. Computing in Science & Engineering **9**: 90–95.

Inc., P. T., 2015 Collaborative data science.

Janick, J. and G. Caneva, 2005 The first images of maize in europe. Maydica **50**: 71.

Kistler, L., S. Y. Maezumi, J. Gregorio de Souza, N. A. Przelomska, F. Malaquias Costa, *et al.*, 2018 Multiproxy evidence highlights a complex evolutionary legacy of maize in south america. Science **362**: 1309–1313.

Le Corre, V. and A. Kremer, 2012 The genetic differentiation at quantitative trait loci under local adaptation. Molecular ecology **21**: 1548–1566.

Leng, E. R., A. Tavčar, and V. Trifunovič, 1962 Maize of southeastern europe and its potential value in breeding programs elsewhere. Euphytica **11**: 263–272.

Li, H., 2011 A statistical framework for snp calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27**: 2987–2993.

Marcus, J., W. Ha, R. F. Barber, and J. Novembre, 2021 Fast and flexible estimation of effective migration surfaces. Elife **10**: e61927.

Mayer, M., A. C. Hölker, E. González-Segovia, E. Bauer, T. Presterl, *et al.*, 2020 Discovery of beneficial haplotypes for complex traits in maize landraces. Nature communications **11**: 1–10.

Mayer, M., S. Unterseer, E. Bauer, N. de Leon, B. Ordas, *et al.*, 2017 Is there an optimum level of diversity in utilization of genetic resources? Theoretical and applied genetics **130**: 2283–2295.

Mezmouk, S. and J. Ross-Ibarra, 2014 The pattern and distribution of deleterious mutations in maize. G3: Genes, Genomes, Genetics **4**: 163–171.

Miles, A., pyup.io bot, M. R., P. Ralph, J. Kelleher, *et al.*, 2023 cggh/scikit-allel: v1.3.6.

Ortiz, R., F. Reslow, R. Vetukuri, M. R. García-Gil, P. Pérez-Rodríguez, *et al.*, 2023 Inbreeding effects on the performance and genomic prediction for polysomic tetraploid potato offspring grown at high nordic latitudes. Genes **14**: 1302.

Peischl, S., I. Dupanloup, M. Kirkpatrick, and L. Excoffier, 2013 On the accumulation of deleterious mutations during range expansions. Molecular ecology **22**: 5972–5982.

Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics **155**: 945–959.

Purugganan, M. D., 2019 Evolutionary insights into the nature of plant domestication. Current Biology **29**: R705–R714.

Purugganan, M. D. and D. Q. Fuller, 2009 The nature of selection during plant domestication. Nature **457**: 843–848.

R Core Team, 2021 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rebourg, C., M. Chastanet, B. Gouesnard, C. Welcker, P. Dubreuil, *et al.*, 2003 Maize introduction into europe: the history reviewed in the light of molecular data. Theoretical and applied genetics **106**: 895–903.

Revilla Temiño, P., M. d. P. Soengas Fernández, M. E. Cartea González, R. A. Malvar Pintos, and A. Ordás Pérez, 2003 Isozyme variability among european maize populations and the introduction of maize in europe .

Samayoa, L. F., B. A. Olukolu, C. J. Yang, Q. Chen, M. G. Stetter, *et al.*, 2021 Domestication reshaped the genetic basis of inbreeding depression in a maize landrace compared to its wild relative, teosinte. Plos Genetics **17**: e1009797.

Savolainen, O., M. Lascoux, and J. Merilä, 2013 Ecological genomics of local adaptation. Nature Reviews Genetics **14**: 807–820.

Slatkin, M. and L. Excoffier, 2012 Serial founder effects during range expansion: a spatial analog of genetic drift. Genetics **191**: 171–181.

Smaragdov, M. and A. Kudinov, 2020 Assessing the power of principal components and wright's fixation index analyzes applied to reveal the genome-wide genetic differences between herds of holstein cows. BMC genetics **21**: 1–15.

Smith, A. L., T. R. Hodkinson, J. Villellas, J. A. Catford, A. M. Csergő, *et al.*, 2020 Global gene flow releases invasive plants from environmental constraints on genetic diversity. Proceedings of the National Academy of Sciences **117**: 4218–4227.

Szpiech, Z. A., 2024 selscan 2.0: scanning for sweeps in unphased data. Bioinformatics **40**: btae006.

Szpiech, Z. A. and R. D. Hernandez, 2014 selscan: an efficient multithreaded program to perform ehh-based scans for positive selection. Molecular biology and evolution **31**: 2824–2827.

Takou, M., T. Hämälä, E. M. Koch, K. A. Steige, H. Dittberner, *et al.*, 2021 Maintenance of adaptive dynamics and no detectable load in a range-edge outcrossing plant population. Molecular biology and evolution **38**: 1820–1836.

Tenaillon, M. I. and A. Charcosset, 2011 A european perspective on maize history. Comptes rendus biologies **334**: 221–228.

Theodorou, K. and D. Couvet, 2006 Genetic load in subdivided populations: interactions between the migration rate, the size and the number of subpopulations. Heredity **96**: 69–78.

Tian, T., Y. Liu, H. Yan, Q. You, X. Yi, *et al.*, 2017 agrigo v2. 0: a go analysis toolkit for the agricultural community, 2017 update. Nucleic acids research **45**: W122–W129.

Tittes, S., A. Lorant, S. McGinty, J. F. Doebley, J. B. Holland, *et al.*, 2021 Not so local: the population genetics of convergent adaptation in maize and teosinte. BioRxiv pp. 2021–09.

Unterseer, S., E. Bauer, G. Haberer, M. Seidel, C. Knaak, *et al.*, 2014 A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k snp genotyping array. BMC genomics **15**: 1–15.

Unterseer, S., S. D. Pophaly, R. Peis, P. Westermeier, M. Mayer, *et al.*, 2016 A comprehensive study of the genomic differentiation between temperate dent and flint maize. Genome biology **17**: 1–14.

Waldvogel, A.-M., B. Feldmeyer, G. Rolshausen, M. Exposito-Alonso, C. Rellstab, *et al.*, 2020 Evolutionary genomics can improve prediction of species' responses to climate change. Evolution Letters **4**: 4–18.

Wang, J., W. G. Hill, D. Charlesworth, and B. Charlesworth, 1999 Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. Genetics Research **74**: 165–178.

Wang, L., T. M. Beissinger, A. Lorant, C. Ross-Ibarra, J. Ross-Ibarra, *et al.*, 2017 The interplay of demography and selection during maize domestication and expansion. Genome biology **18**: 1–13.

Yang, J., S. Mezmouk, A. Baumgarten, E. S. Buckler, K. E. Guill, *et al.*, 2017a Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. PLoS genetics **13**: e1007019.

Yang, N., Y. Wang, X. Liu, M. Jin, M. Vallebueno-Estrada, *et al.*, 2023 Two teosintes made modern maize. Science **382**: eadg8940.

Yang, N., X.-W. Xu, R.-R. Wang, W.-L. Peng, L. Cai, *et al.*, 2017b Contributions of zea mays subspecies mexicana haplotypes to modern maize. Nature communications **8**: 1874.

Zimmermann, H., P. Brandt, J. Fischer, E. Welk, and H. von Wehrden, 2014 The human release hypothesis for biological invasions: human activity as a determinant of the abundance of invasive plant species. F1000Research **3**.

Supplement

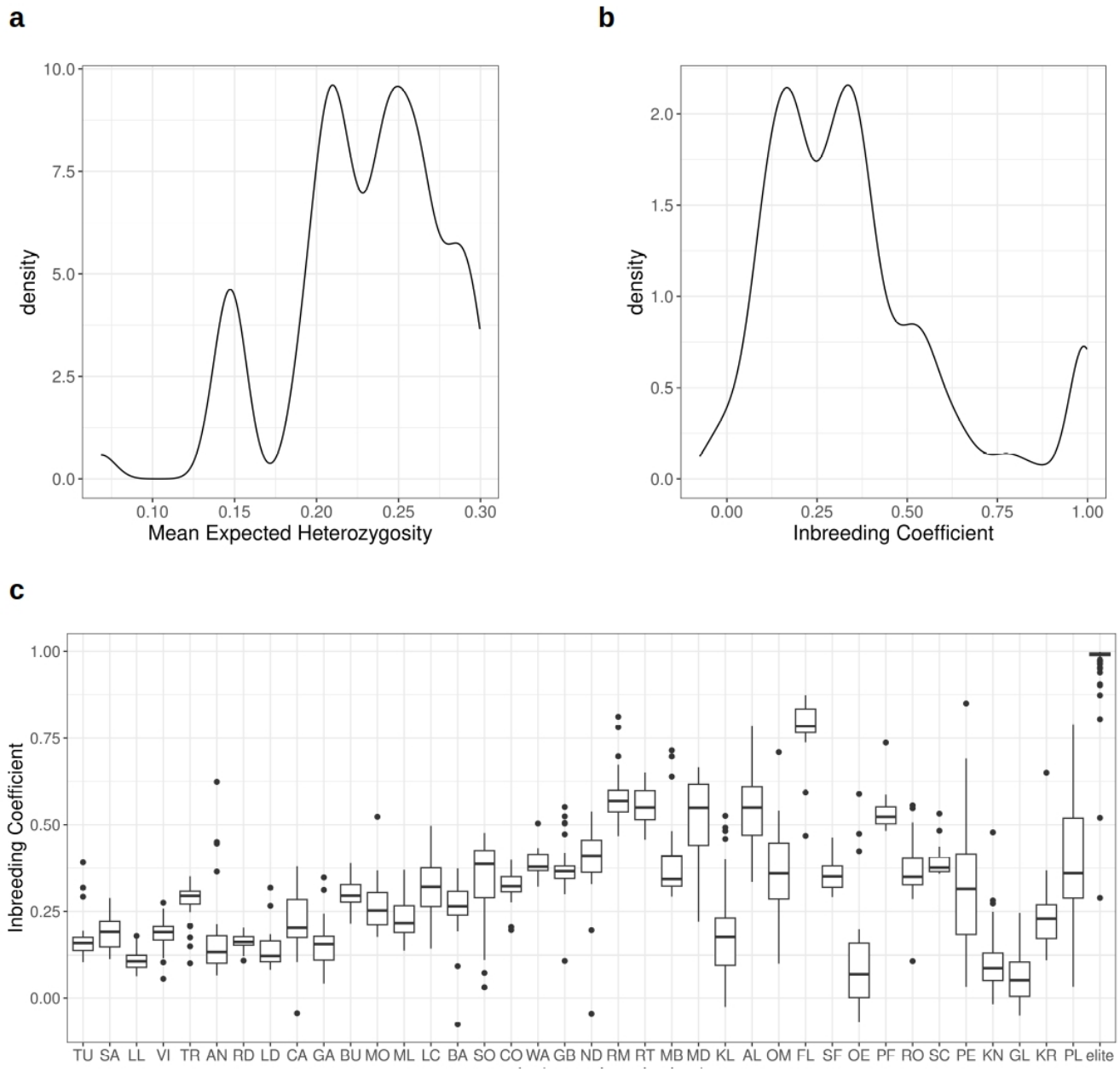

Figure S1 Genetic diversity within European maize. a. The density of the mean H_{Exp} for all populations is shown. b. The density of the F_{IS} values for each individual of all landraces and elite lines. c. The F_{IS} values per population, ordered from West to East with the elites depicted at the far right of the plot.

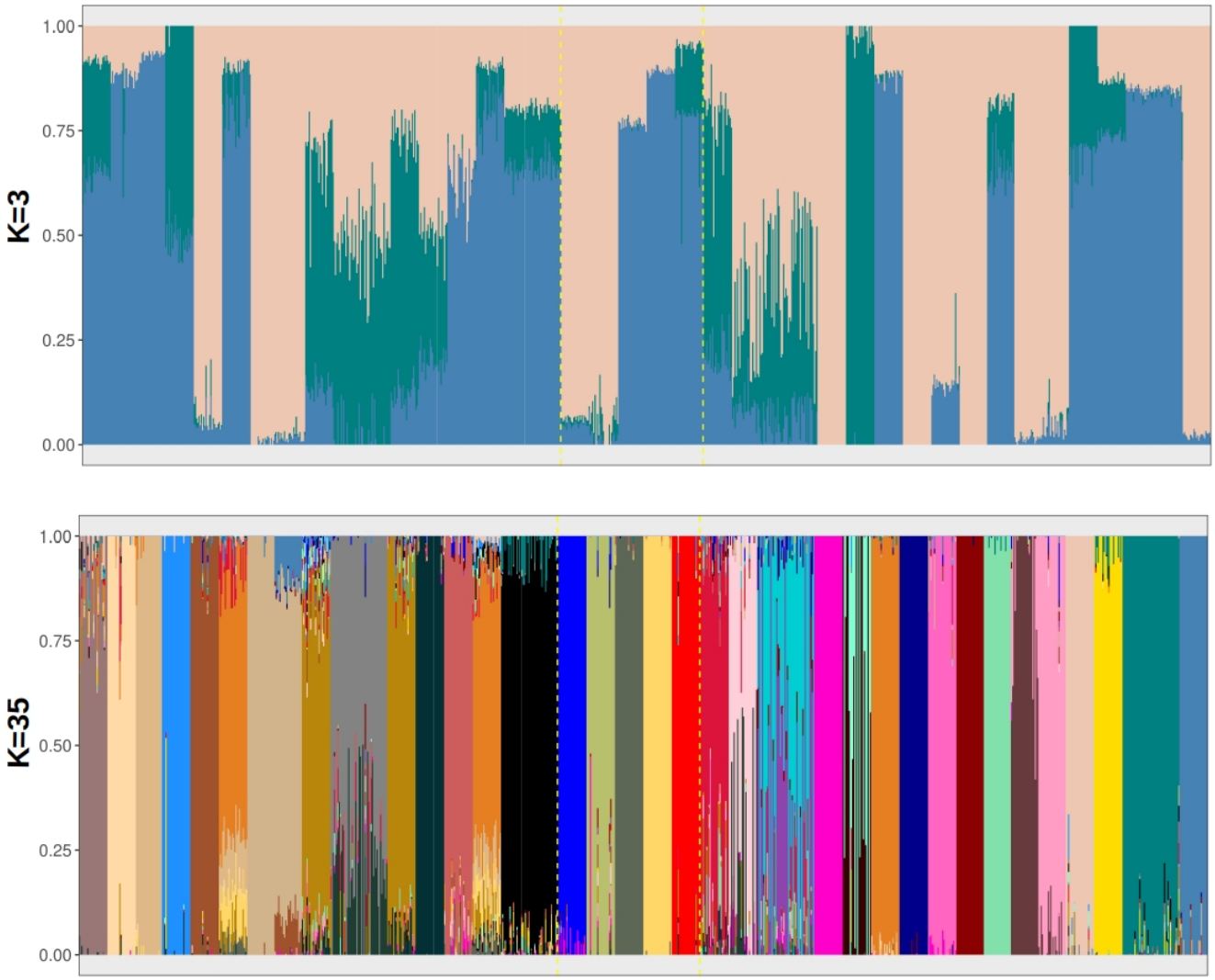

Figure S2 Structure analysis of the European maize landraces. On top, the results for cluster K=3 and on the bottom K=35 are shown. The populations are ordered as in the F_{ST}
heatmap (Figure 2). The dashed yellow lines in both plots indicate where each F_{ST} group starts and ends.

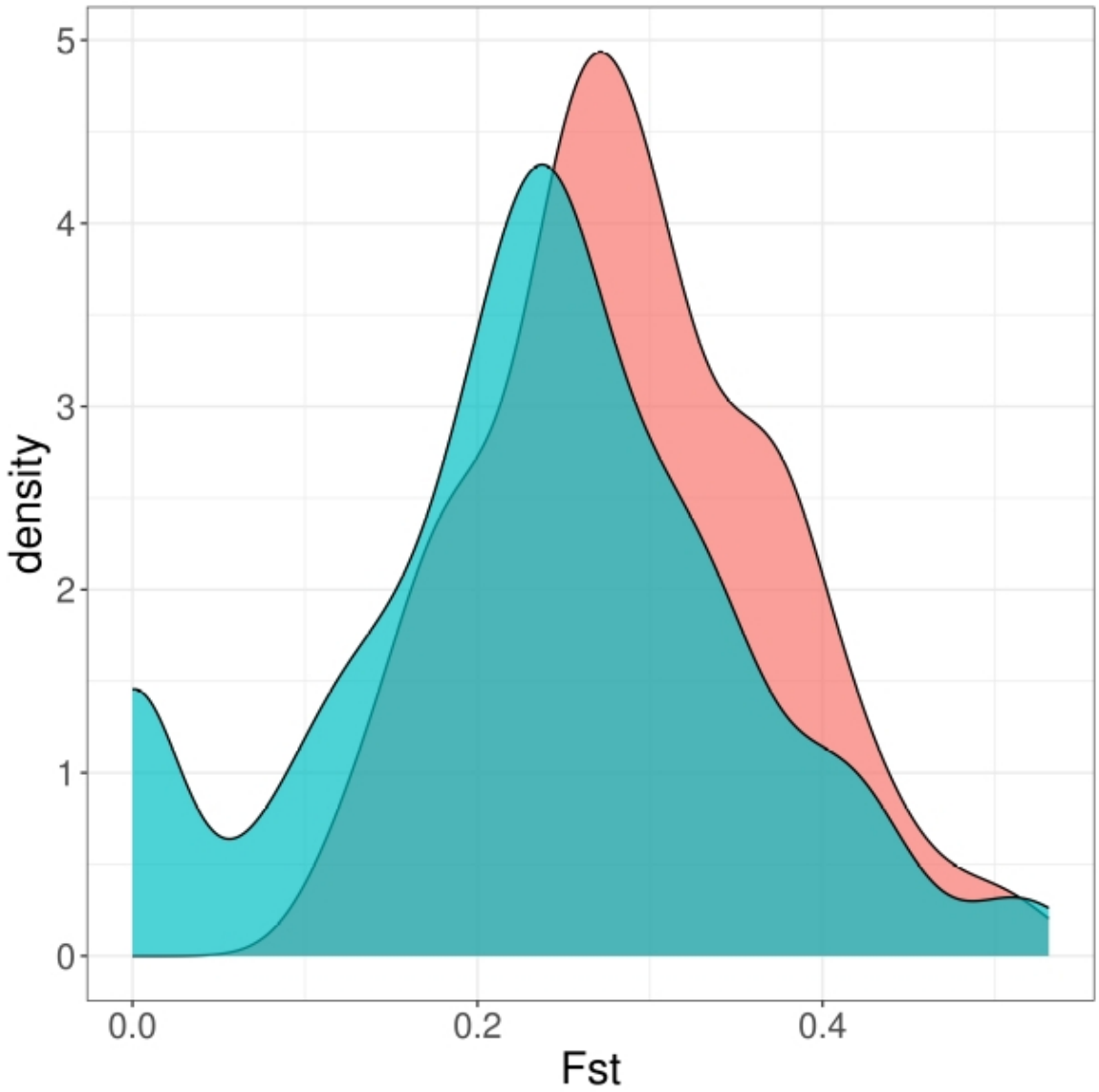

Figure S3 **The distributions of pairwise F_{ST} values for all the pairs within the same cluster (blue), or between the two clusters (red).**

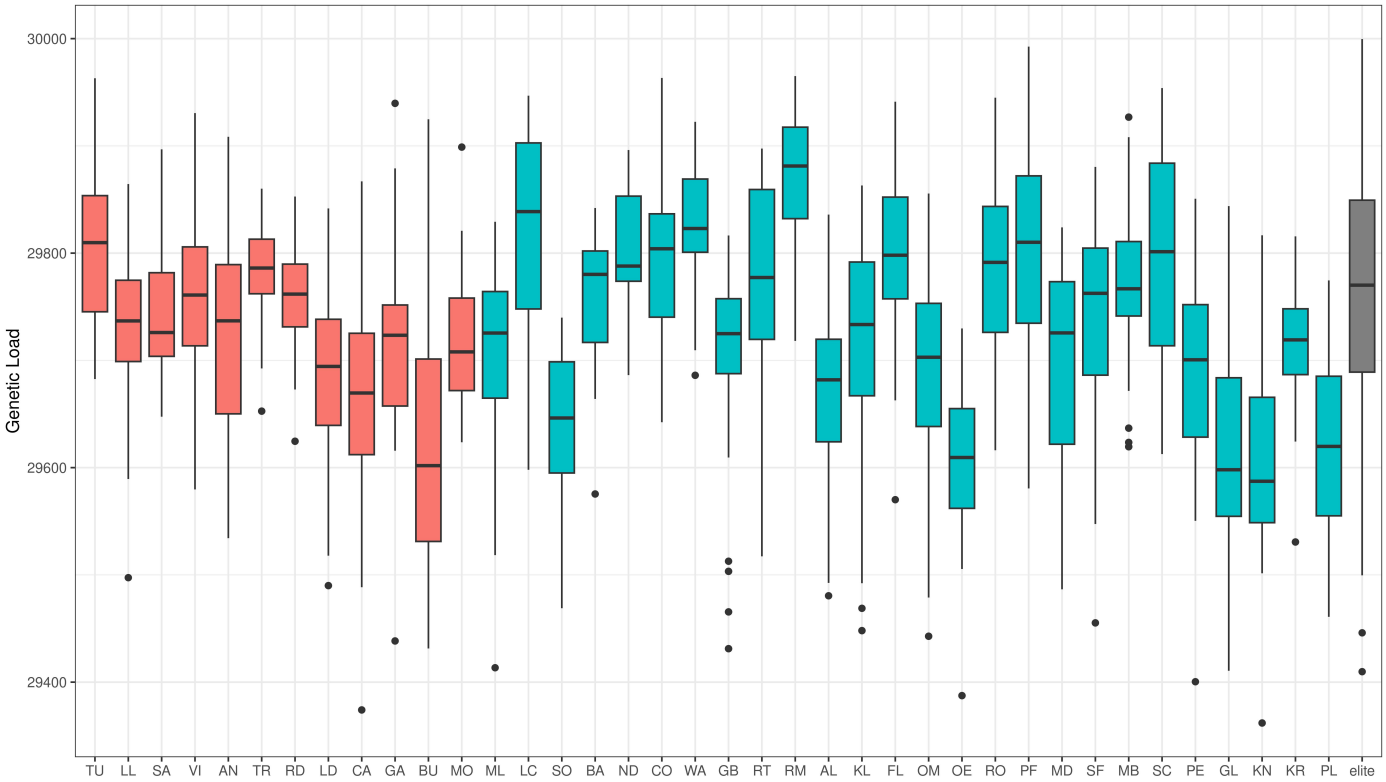

Figure S4 **The genetic load per European landrace population ordered by their distance from the hypothetical introduction point in Spain. In red are marked all the populations belonging in the West/Spanish cluster, in blue the populations in the Central European cluster and in grey the elite lines.**