Theor Appl Genet (2013) 126:2671–2682
DOI 10.1007/s00122-013-2164-z

ORIGINAL PAPER

Out of America: tracing the genetic footprints of the global diffusion of maize

C. Mir · T. Zerjal · V. Combes · F. Dumas · D. Madur · C. Bedoya · S. Dreisigacker · J. Franco · P. Grudloyma · P. X. Hao · S. Hearne · C. Jampatong · D. Laloë · Z. Muthamia · T. Nguyen · B. M. Prasanna · S. Taba · C. X. Xie · M. Yunus · S. Zhang · M. L. Warburton · A. Charcosset

Received: 13 September 2012 / Accepted: 12 July 2013 / Published online: 7 August 2013

Abstract Maize was first domesticated in a restricted valley in south-central Mexico. It was diffused throughout the Americas over thousands of years, and following the discovery of the New World by Columbus, was introduced into Europe. Trade and colonization introduced it further into all parts of the world to which it could adapt. Repeated introductions, local selection and adaptation, a highly diverse gene pool and outcrossing nature, and global trade in maize led to difficulty understanding exactly where the diversity of many of the local maize landraces originated. This is particularly true in Africa and Asia, where historical accounts are scarce or contradictory. Knowledge of post-domestication movements of maize around the world would assist in germplasm conservation and plant breeding efforts. To this end, we used SSR markers to genotype multiple individuals from hundreds of representative landraces from around the world. Applying a multidisciplinary approach combining genetic, linguistic, and historical data, we reconstructed possible patterns of maize diffusion throughout the world from American "contribution" centers, which we propose reflect the origins of maize worldwide. These results shed

C. Mir and T. Zerjal contributed equally to this paper.

Communicated by T. Luebberstedt.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2164-z) contains supplementary material, which is available to authorized users.

C. Mir · V. Combes · F. Dumas · D. Madur · A. Charcosset
Unité Mixte de Recherche de Génétique Végétale, Institut National de la Recherche Agronomique, Université Paris Sud, Centre National de la Recherche Scientifique (INRA), AgroParisTech, Ferme du Moulon, 91190 Gif-sur-Yvette, France

T. Zerjal · D. Laloë
UMR 1313 Génétique Animale et Biologie Intégrative INRA-AgroParisTech, Domaine de Vilvert, 78352 Jouy en Josas, France

C. Bedoya · S. Dreisigacker · S. Hearne · S. Taba
Centro International de Mejoramiento de Maíz y Trigo (CIMMYT), Applied Biotechnology Center, El Batan, Apdo Postal 6-641, 06600 Mexico, D.F., Mexico

J. Franco
Departamento de Biometria, Facultad de Agronomia, Estadistica y Computacion, Estación Experimental EEMAC, Ruta 3, Km. 363, 60000 Paysandú, Uruguay

P. Grudloyma
Department of Agriculture, Nakhon Sawan Field Crops Research Center, PhahonYothin Road, Tak Fa, Nakhon Sawan 60190, Thailand

P. X. Hao · T. Nguyen
National Maize Research Institute, Phuong Town, Dan Phuong, Hanoi, Vietnam

C. Jampatong
National Corn and Sorghum Research Center, Kasetsart University, 298 Moo 1, Klang Dong, Pak Chong, Nakhon Ratchasima 30320, Thailand

Z. Muthamia
Kenya Agriculture Research Institute, P. O. Box 57811, Nairobi, Kenya

B. M. Prasanna
Indian Agricultural Research Institute, New Delhi 110012, India

B. M. Prasanna
CIMMYT, P. O. Box 1041, Village Market, Nairobi 00621, Kenya

C. X. Xie · S. Zhang
Chinese Academy of Agricultural Sciences, Beijing 100094, China

new light on introductions of maize into Africa and Asia. By providing a first globally comprehensive genetic characterization of landraces using markers appropriate to this evolutionary time frame, we explore the post-domestication evolutionary history of maize and highlight original diversity sources that may be tapped for plant improvement in different regions of the world.

Introduction

Maize (*Zea mays* ssp. *mays* L.) domestication began around 9,000 BP in a restricted valley in south-central Mexico (Matsuoka et al. 2002), but has evolved into the crop with the third highest cultivation area (Leff et al. 2004) and second highest production (http://faostat.fao.org) worldwide. As a critical staple and the most geographically ubiquitous cereal (Leff et al. 2004), maize has been the focus of numerous genetic and genomic studies on domestication (Matsuoka et al. 2002; van Heerwaarden et al. 2011), diversity (Camus-Kulandaivelu et al. 2006; Dubreuil et al. 2006; Vigouroux et al. 2008) and plant improvement (Duvick 2005; Menkir et al. 2006; Witcombe et al. 2003). Maize was first spread throughout the Americas over thousands of years, allowing progressive adaptation to new environments and a gradual evolution of hundreds of landraces, or farmer's varieties. These tend to be low yielding, but contain more phenotypic and genetic variation than improved modern varieties (Liu et al. 2003; Vigouroux et al. 2008; Warburton et al. 2008), and represent an extended "center" of maize diversity (Matsuoka et al. 2002; van Heerwaarden et al. 2011). In contrast, historical diffusion out of the Americas by early explorers was rapid and subsequent adaptation and selection by local farmers, repeatedly around the world, led to the creation of hundreds of new landraces in the past 500 years (Dubreuil et al. 2006).

Knowledge of post-domestication movements of maize around the world would assist in germplasm conservation and plant breeding efforts, to identify genetic groups which may not yet have been considered with recent breeding approaches and which may deserve specific evaluation. This is particularly true in countries with smaller maize improvement programs and where maize is a more recent introduction. In some cases, knowledge of diversity organization is instrumental in organizing complementary heterotic groups and guiding crosses in regions where hybrid breeding is recent, such as in many Asian countries (Pray 2006). However, the tangled web of global exchanges of maize germplasm and the lack of precise historical documentation, particularly for introductions into Africa and Asia, have made it difficult to understand relationships among landraces. Meanwhile, the absence of genetic studies at a truly global level, with representative populations from all inhabited continents, makes it more difficult to identify variation in these genetic resources that may be crucial to conserving and exploiting them for the efficient production and management of new varieties (Vigouroux et al. 2008). Identifying the evolutionary sources of genetic diversity from within the Americas and then evaluating the American genetic ancestry of non-American landraces would allow patterns of maize migrations out of the Americas and into the rest of the world to be inferred. Matching the original diversity, sources of maize landraces from many countries will in turn highlight the most closely related (and probably most similarly adapted) sources of new diversity that can be tapped to expand the genetic pool for breeders in any given environment.

Maize is highly heterogenous, and most of the diversity is partitioned within each population, rather than between populations (Warburton et al. 2008). Many characterization studies of maize have used only one, or at most two or three, individuals to characterize each population. Unless a very high number of markers are used, this strategy has the potential to miss much of the within population variation present in the populations under study. To overcome this limitation, a bulking strategy has been used, where 15 individuals per population were analyzed simultaneously (Dubreuil et al. 2006). Hypervariable SSR markers, which have the best signature for this time scale (Ellegren 2000), can be used to characterize the diversity within maize landrace populations. Combined with a bulking strategy, SSRs allow the use of fewer markers for a complete genetic characterization (Hamblin et al. 2007). SSRs also do not suffer from the ascertainment bias that Single Nucleotide Polymorphism (SNP) markers do.

The purpose of this study was to contribute to reconstruction of the genetic footprints of maize origins worldwide. To do so, we genotyped multiple individuals from 799 different landrace accessions sampled from across the globe with SSR markers to compare diversity and evaluate the contribution of different American genetic groups to those cultivated in other regions of the world. Based on this genetic information and also historical and linguistic records, we propose a number of hypotheses on main migration routes that moved maize into new continents

M. Yunus
Indonesian Department of Agriculture, Bogor, Indonesia

M. L. Warburton (✉)
United States Department of Agriculture, Corn Host Plant Research Resistance Unit, Mississippi State University, Box 9555, Oktibbeha 39762, USA
e-mail: Marilyn.warburton@ars.usda.gov

from original sources of diversity in the maize center of origin, providing a comparison basis for future historical investigations and genetic studies of maize landraces based on other marker types.

Materials and methods

Plant materials

A total of 11,985 maize plants, from 784 different landrace populations represented by 799 accessions, were sampled for genetic characterization. This sampling covered most of the range of maize over the Americas (258 accessions), Africa (237), the Middle East (13), Europe (148) and Asia (143) (Fig. 1) (Monfreda et al. 2008), and was thus expected to represent the global landrace gene pool. This sample included 269 American and European maize accessions previously analyzed with SSR markers (Camus-Kulandaivelu et al. 2006). In addition, one teosinte (*Zea perennis* accession Piedra Ancha from Mexico) was included as an outgroup to be consistent with prior studies of this type. For each maize accession, 15 plants were analyzed. All accessions were part of institutional collections. Complete passport data are detailed in Table S1.

Molecular analyses

DNA was extracted from lyophilized leaf tissue collected from individual 3-week-old plants grown in the greenhouse, using a CTAB method, and quantified (see details from http://www.cimmyt.org/english/docs/manual/protocols/abc_amgl.pdf). Each accession was analyzed as a bulk of DNA from 15 individual plants, mixed in equal amounts as described in Reif et al. (2005) and Dubreuil et al. (2006). To confirm the characterization of the landraces presented here, we compared our findings to the outcomes of a previous study of American landraces based on a large number of SNP markers, but only one individual per accession (van Heerwaarden et al. 2011).

All accessions were genotyped with 17 unlinked SSR markers (Table S2) that had good reproducibility in allele size determination. All genotyping reactions were performed following standard PCR amplification protocols (Dubreuil et al. 2006), using appropriate controls to enable the incorporation of the preexisting American and European data from Camus-Kulandaivelu et al. (2006) into this study. Fluorescently labeled PCR products were separated on a Li-Cor® sequencer as described in Camus-Kulandaivelu et al. (2006). Images were processed using the One-Dscan v. 2.05 software (Scanalytics, Fairfax, VA) to estimate fragment sizes and band intensities from peak height. For each DNA bulk, allele frequencies were estimated from the intensity of the bands corresponding to alleles after filtering the background noise resulting from persistent stuttering phenomena of SSR amplification using the deconvolution method (Dubreuil et al. 2006). Homogeneity in allele calling between gels was guaranteed by allele standards and the re-analysis of accessions from different gels together. In very rare cases, alleles displaying close sizes and not systematically discernible were pooled (i.e., classed as equivalent).

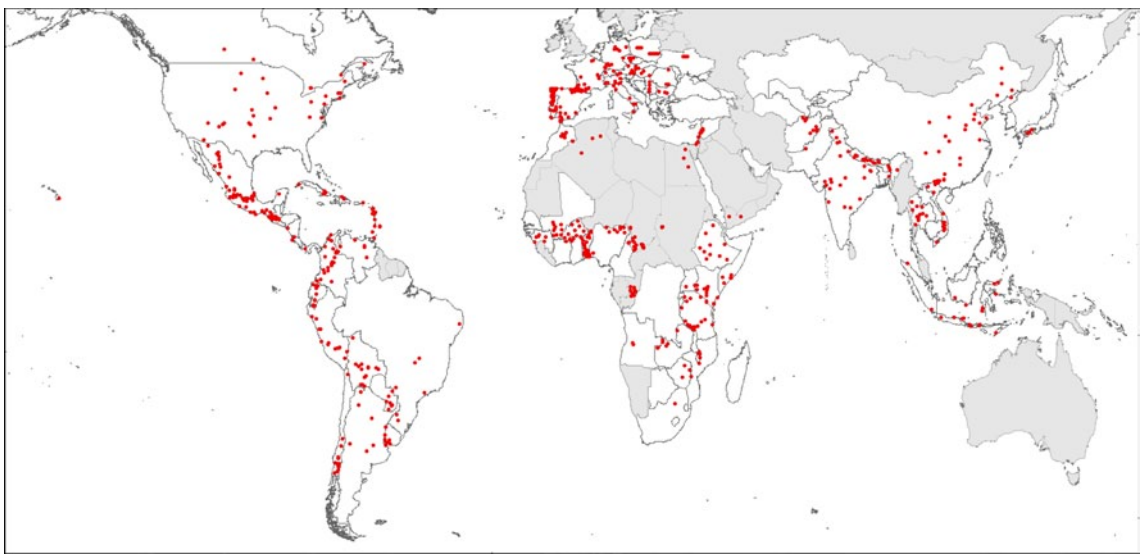

Fig. 1 Geographic origin of the 799 maize landrace accessions included in the study; each landrace represented with a *red dot*. Countries with only negligible maize cultivation (according to Leff et al. 2004) are *shaded*

Population structure analyses

Maize population structure was inferred based on two methods, one using Bayesian clustering algorithms and the second proposing a multivariate analysis that does not require data to meet Hardy-Weinberg expectations or linkage equilibrium to exist between loci (Jombart et al. 2008; Jombart et al. 2009; Jombart et al. 2010). We chose to use both methods to assess the various assumptions and criticisms of each method when interpreting the results.

The Bayesian clustering analysis was carried out using the *STRUCTURE* 2.2 program (Pritchard et al. 2000) with a simulated matrix of five individual genotypes per maize accession to satisfy allele frequencies and Hardy-Weinberg equilibrium assumptions with the R statistical software package (R Development Core Team 2012). Potential bias of accession clustering when using simulated individuals was discounted by comparing our results with previous results based on individual genotypes (Vigouroux et al. 2008) for American landraces. Analyses were performed assuming independent allele frequencies among clusters, an admixture model, and 1,000,000 iterations following a burn-in period of 500,000 iterations. Five replicates were performed for each cluster K, from 2 to K_{max}. The smallest K value that captured most of the structure in the data before the log likelihood curve leveled off was retained (Fig. S1A). The contributions of each cluster to each accession genome (mean admixture proportions $Q_{1\ to\ K}$) were estimated by averaging admixture proportions over the corresponding five simulated individuals. Each cluster contained all "representative accessions", those displaying Q values higher than a threshold of 80 %; other accessions were referred to as "admixed accessions". The optimal K value was also used to identify "biologically sensible clusters", i.e., clusters identified by at least few accessions related by pedigree, origin, or breeding program.

STRUCTURE was first run on only the American accessions ($K = 1$ to $K_{max} = 11$) to identify "American genetic clusters", which are those that could have contributed to the genetic background of non-American landraces. To document the diffusion process of maize out of the Americas, a *STRUCTURE* "contribution" analysis was then performed on the entire dataset. In this case, a priori information based on the American clustering result was added (Pritchard et al. 2000) to allow the ancestry estimation for non-American landraces (Murgia et al. 2006). We allowed for some misclassification of American individuals by setting the *Migrprior* option to 0.01. Mean admixture proportions were estimated for each accession; for each non-American accession, these were considered as reflecting accession origin(s). Population structure patterns were geographically mapped by coloring accessions according to their ancestry in each specific cluster (i.e., according to their mean admixture proportions for that cluster), using the *rgdal* and *sp* packages of R.

The second method used to identify genetic patterns relied on Principal Component Analysis. As an alternative to Bayesian clustering algorithms (van Heerwaarden et al. 2011; Pray 2006), it is free of assumptions about underlying population genetic models and can be applied to any type of genetic data, without having to simulate individual genotypes. The analysis was performed on the global dataset using the *dapc* function implemented in the R package "Adegenet" (Jombart 2008). This method, called DAPC, relies on a PCA data transformation step, followed by a discriminant analysis on the retained principal components to partition genetic variation into a between-group and a within-group component, minimizing variation of the within-group component. To account for more cryptic spatial patterns of genetic structure across the landscape, a spatial Principal Component analysis (sPCA) was also run for each continent independently, using the *spca* function of Adegenet, which is particularly adapted to complex situations such as hierarchical clustering or genetic clines (Jombart et al. 2008; Gautier et al. 2010). While the PCA optimization criterion only accounts for genetic variance, the sPCA includes also spatial autocorrelation measured by Moran's I index. This is obtained via a neighbor network that connects geographically closed populations to a model of spatial structure among populations. For a thorough description of this method see Jombart et al. 2008.

The global dataset was analyzed with DAPC without any prior grouping information. From the preliminary PCA data transformation, 120 principal components were retained, which accounted for 94 % of the total variability, and this was followed by a Discriminant Analysis. To identify the best supported number of clusters, we used the successive K-means clustering procedure as implemented in the *find. cluster* function of Adegenet, with increasing number of clusters (K from 1 to 50). With each increasing value of K, different clustering solutions were compared using the Bayesian Information Criterion (BIC), where the optimal clustering solution should correspond to the lowest BIC value (Fig. S1B). Because of the large number of populations analyzed and the complexity of interpretation of the DAPC plot, we projected the PC components on a geographic map using the function *colorplot* of Adegenet. This function summarizes up to three PC components at the time by recoding each population score as intensities of a given color channel of the RGB system: red for the first PC, green for the second PC and blue for the third. Each color-recoded population score was then plotted onto a map using population geographical coordinates. sPCA results for the first three sPC axes were also projected on geographic maps (Fig. S2A–D) using a bubble chart representation.

Phylogenetic analyses

Phylogenetic analyses were performed over (1) all maize accessions and (2) the representative maize accessions of the *STRUCTURE* clusters as defined in the American and contribution analyses, respectively. Analyses were based on pairwise distances estimated between accessions from the allele frequencies using the natural log transformation of the proportion of shared alleles, which is free from the assumptions of the mutation model (Matsuoka et al. 2002). We only considered the representative accessions in the second analysis to limit the blurring effect of strong admixture on the phylogenetic signals and to reduce the computational time. Trees were constructed using *Phylip* 3.69 (Felsenstein 2005) with neighbor joining and the Fitch algorithm according to Matsuoka et al. (2002) and Vigouroux et al. (2008), using a *Z. perennis* accession as outgroup. Statistical significance of tree topology was estimated by bootstrap re-sampling among loci 1,000 times (Felsenstein 1985). Limited statistical significance was found for trees based on representative accessions from clusters, and thus the same accession matrix was also used to construct phylogenetic networks using the Neighbor-Net algorithm in *SplitsTree* 4.11.3 (Huson and Bryant 2006) to enable a more realistic inference of the complex maize evolutionary history (Bryant and Moulton 2004). On Neighbor-Net graphs, each split between groups of accessions is visualized by parallel lines, whose length reflects their weight. The major phylogenetic signals are thus displayed together with the conflicts resulting from reticulation, and visualized as boxes. In each box, line lengths thus indicate support for two competing patterns of phylogenetic relationships.

Genetic diversity analyses

Allele richness (A) was reported for each locus and over all loci; values were averaged over all accessions (A_a) and calculated for each continent. Nei's unbiased genetic diversity was estimated following the method of Pons and Chaouche (1995) for meta-populations subdivided into a large number of populations (considering each accession as a population). Total diversity was computed at each locus (h_t) over all maize accessions (Table S2). Mean population diversity (H_k), total genetic diversity (H_t) and between-population differentiation (G_{st}) were estimated over all loci and all accessions. Parameters were calculated for each continent, and G_{st} values were estimated for all pairs of continents. The significance of the differentiation between continents was tested by comparing the G_{st} values with 1,000 simulated G_{st} values obtained assuming no genetic differentiation by random permutations of accessions between continents or clusters, respectively (Sokal and Rohlf 1995). Partitioning of maize SSR diversity was estimated for each continent and significance of continent comparisons were performed using a Wilcoxon signed-rank test using the package *Statistica 6.1* (StatSoft, France).

Results

Cluster analyses and genetic variation in American Landraces

The Bayesian clustering software *STRUCTURE 2.2* was initially used to infer population structure in American landraces only, to identify major clusters in the maize center of diversity (Fig. S3). In Vigouroux et al. (2008), an analysis of 350 American maize races indicated that $K = 4$ was the optimal clustering model. In our analysis of 258 accessions at $K = 4$, 61 % of the American accessions showed admixed origins (i.e., the four clusters contributed <80 % to the genetic diversity of these accessions), which was higher than that reported by Vigouroux et al. (2008). The admixed accessions identified at $K = 4$ were mainly from three geographical zones: the southern end of the Great Plains of the United States, the northern part of South America (Columbia and Venezuela) and the middle part of South America (Bolivia, Paraguay and southeastern Brazil). These accessions were gradually assigned to a specific cluster when the K values were increased. At $K = 6$ (Fig. S3E), two distinct South American clusters formed from what was a single cluster at $K = 5$ (Fig. S3D), and many of the admixed accessions joined one of these two. Seven clusters captured most of the structure in the data before the log likelihood curve reached a plateau (Fig. S1A, Fig. S3F). For this $K = 7$ value, in North America, the Northern US flints were distinguished from the Corn Belt dents which we referred to as the "Middle North-America" group. At $K = 6$ and $K = 7$, the separate "Mexican Highlands" group encompasses the Southern Dents.

In this seven-cluster model, significant genetic differentiation was found among clusters (G_{st} 0.10–0.33, all $P < 0.001$, Table 1). We kept the names of the four clusters also identified by Vigouroux et al. (2008) as they had reported: "Northern US flints", "Mexican highlands", "Tropical lowlands" and "Andes". We called the remaining three clusters "Middle North-American", "Northern South-American", and "Middle South-American", based on their geographic origins. Residual admixture was still observed for 63 % of American accessions, which may reveal a true composite ancestry of accessions from different clusters reflecting gene exchanges. On the other hand, it may reflect the difficulty of using *STRUCTURE* to identify additional cluster(s) in the case of continuous allele frequency gradation between populations (Vigouroux et al. 2008), and/or too limited sample size

Table 1 Pairwise genetic differentiation between the seven maize clusters defined in the American *STRUCTURE* analysis (G_{st}, expressed in percentage) estimated over all loci

Clusters	Tropical lowlands	Northern South-America	Middle South-America	Northern US flints	Mexican highlands	Middle North-America
Northern South-America	14.95*	–	–	–	–	–
Middle South-America	17.60*	12.35*	–	–	–	–
Northern US flints	26.58*	26.61*	27.43*	–	–	–
Mexican highlands	14.28*	16.57*	15.34*	22.35*	–	–
Middle North-America	12.39*	18.80*	17.28*	16.19*	10.28*	–
Andes	18.54*	14.16*	11.33*	32.72*	18.71*	22.56*

* Refer to significant values ($P < 0.001$)

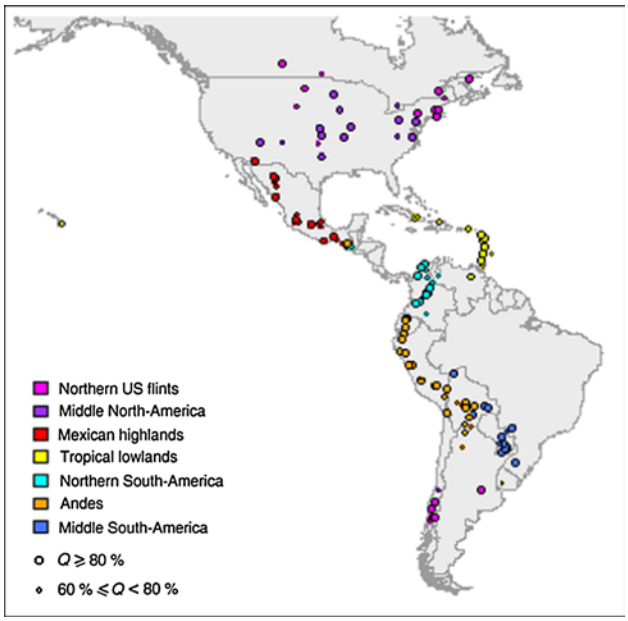

Fig. 2 Geographical location of the seven *STRUCTURE* clusters identified in the Americas ($K = 7$). Each of the 258 accessions from the Americas is colored according to their major ancestry (mean admixture proportion Q) in one of the seven clusters only when Q exceeds 60 %. Size of the circles is proportional to the variation in ancestry according to *STRUCTURE*

(Rosenberg et al. 2002). The seven clusters identified clear geographic patterns (Fig. 2), including strong latitudinal differences and contrasting mean elevations (Northern US flints 270 m, Middle North-America 406 m, Mexican highlands 1517 m, Tropical lowlands 76 m, Northern South-America 928 m, Middle South-America 392 m, and Andes 2279 m). The same pattern was also identified by sPCA analysis of American accessions (Fig. S2A) and in the phylogenetic analysis of these accessions (Fig. S4), as well as previous suggestions that different environmental conditions have shaped the genetic composition of maize landraces (Vigouroux et al. 2008; Ducrocq et al. 2008).

Diversity and phylogenetic analysis of global maize landraces

Overall genetic diversity in the study was very high (Table 2), and while there was significant genetic differentiation between continents, many of the alleles were shared between geographical regions (Table 2). The mean population diversity (H_k) calculated for each continent was higher in North America and diminished progressively eastwards with the smallest value in Asia (Table 2). The opposite trend occurred for genetic differentiation among accessions (G_{st}), where values were higher in Asia and diminished westward toward the Americas. The neighbor-joining tree based on the pairwise distances between the whole set of maize accessions often displayed tree branching that reflected the common origin of accessions (Fig. 3; Fig. S4). Nevertheless, inference of global diffusion of maize from this tree remained overly complex, as many clades joined accessions from diverse geographical origins and the global tree branching remained of difficult interpretation. Many of the accessions displaying multiple American contributions or with diverse entries would be due to hybrid ancestry. The Pyrenean accessions displaying predominantly Northern South-American ancestry and the Galician accessions displaying Northern US flint ancestry are a case in point, as these accessions are hybrids of the two parental American ancestral clusters. We were nevertheless able to draw better conclusions from the phylogenetic networks constructed from genetic distances within each diffusion route, which largely supported the *STRUCTURE* results and both were used in creating potential genetic routes of diffusion of maize landraces, as discussed below.

A global *STRUCTURE* analysis of all accessions, using the seven American clusters as prior information, was performed to infer their genetic contribution to the non-American landraces. This over-simplification does not infer total global population structure, but was done to identify main trends of maize diffusion. Because maize has experienced a series of repeated introductions over the last 400 years at most, analysis of allele frequency gradients, such as

Table 2 Partition of maize diversity measured by SSRs

Area	Accession number	A^a	A_A^b (%)	$H_t \pm SE^‡$	$H_k \pm SE^c$	G_{st}^d
World	799	174	–	0.596 ± 0.003	0.409 ± 0.001	0.313
America	258	153	–	0.596 ± 0.003	0.443 ± 0.003	0.257
Europe-Asia-Africa/ Middle East	541	149	83.66	0.589 ± 0.002	0.393 ± 0.002	0.333
Europe	148	110	69.93	0.585 ± 0.003	0.411 ± 0.004	0.300
Asia	143	118	66.01	0.550 ± 0.004	0.376 ± 0.004	0.316
Africa/Middle East	250	117	73.20	0.568 ± 0.003	0.392 ± 0.003	0.310

[a] Total number of alleles for 17 SSRs

[b] Percentage of American alleles represented

[c] Mean genetic diversity per accession estimated over all loci, ± standard error (SE)

[d] Genetic differentiation among accessions within each area

[‡] Total genetic diversity estimated over all loci, over all accessions for each area, ± standard error (SE). No significant variation was found for the contrast of America versus Europe, Asia and Africa/Middle East, nor for the contrast of global accessions versus Europe, Asia, Africa/Middle East. (pairwise Wilcoxon signed-rank test, all $P > 0.05$)

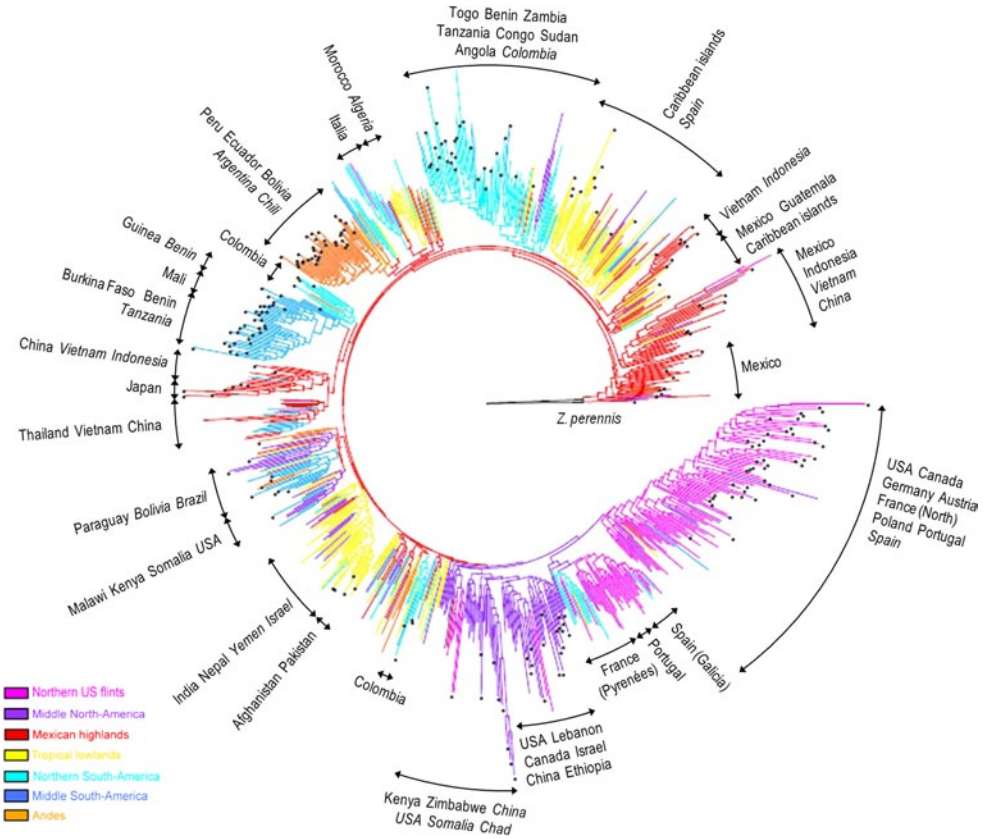

Fig. 3 Global phylogeny of 799 maize landraces. Phylogenetic relationships between maize accessions were calculated based on the log-transformed proportion of shared alleles and visualized using Neighbor-Joining algorithm. A teosinte accession of *Z. perennis* was used as outgroup. Accessions were colored according to their major ancestry as defined in the *STRUCTURE* contribution analysis ($K = 7$), and labeled with an *asterisk* indicating ancestry within that *STRUCTURE* cluster of at least 80 %. For the major phylogenetic clusters, the main origins of the accessions are mentioned, with *italics* referring to less frequent origins

presented by Van Etten and Hijmans (2010) was not appropriate. In this "contribution analysis", most accessions (70 %) received contributions from multiple ancestral clusters (admixed-ancestry) (Table S3), but the predominant contribution of each ancestral cluster was geographically highly delimited (Fig. S5).

DAPC analysis of the global dataset was run to verify the *STRUCTURE* "contribution" results. Based on the Bayesian Information Criterion, we retained 15 clusters for the analysis (Fig. S1B), representing the minimum number of clusters after which the BIC either increased or decreased by only a negligible amount. The DAPC analysis showed that most of the genetic diversity was summarized by the first three principal components (Fig. 4) and the genetic partition obtained closely resembled the one identified by *STRUCTURE*. The first principal component

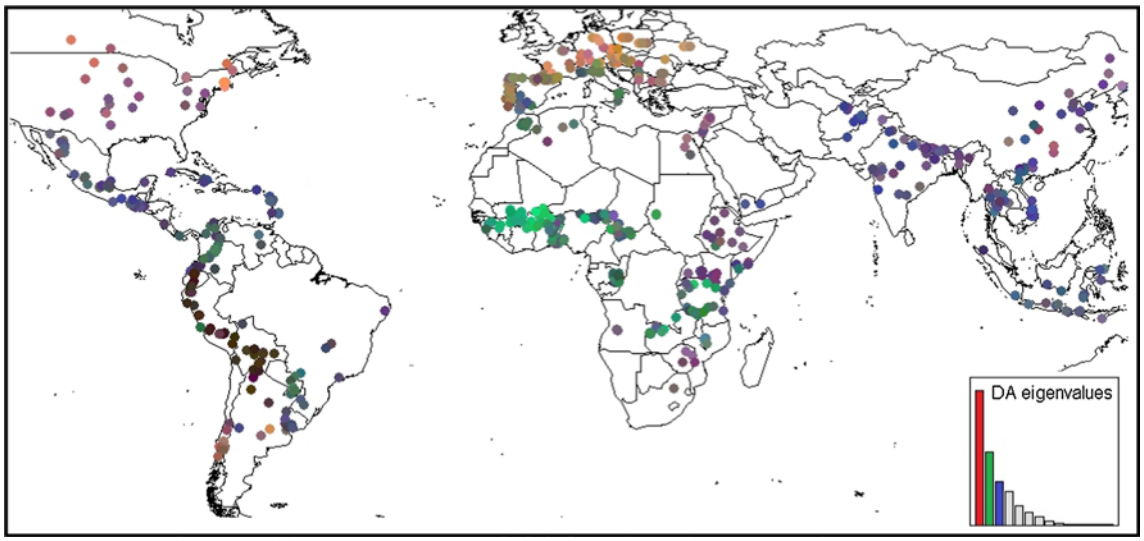

Fig. 4 World-wide projections of the colorplot synthesizing accession's coordinates on the first three DAPC components. Each *dot* corresponds to a population analyzed in the study. Each principal component is represented as intensity of a given color channel; the first PC is shown in *red*, the second PC in *green* and the third PC in *blue*. The *inset* indicates the eigenvalues of the DAPC

(red channel, Fig. 4) clearly differentiated flint germplasm from North America and Northern Europe from the rest of the world. The second principal component (green channel, Fig. 4) identified a cluster of accessions from Western Africa, Colombia and Middle South America, it also included some Italian and Pyrenean accessions. Finally, the third component (blue channel, Fig. 4) grouped Tropical and Mexican accessions with those from Southern Asia, and included accessions from Southern Spain and Eastern Africa. Accessions from the Andes appear as a separate group. The large number of accessions with mixed proportion of colors reflects the large admixed genetic component of most accessions, in accordance with the *STRUCTURE* analysis.

The sPCA run for each continent (combining North and South America in one analysis) to better understand the geographic pattern of local genetic structuring are presented in Fig. S2A–D. In the Americas, sPCA supported a strong geographic component in the organization of maize diversity, with clines of genetic differentiation diffusing northward and southward from Central America (Fig. S2A, axis 1 and 2). A genetic clustering of Mexican accessions and the Andean accessions are revealed by axis 3 (Fig. S2A). In Europe, a different origin of Northern and Southern European accessions was suggested by the clearly delineated spatial pattern of this maize diversity (Fig. S2B, axis 1). Contrasting genetic groups were also identified in the Iberian Peninsula (Fig. S2B, axis 1–3), suggesting multiple introductions and perhaps admixture of accessions from this area. In Africa, populations that are geographically near but genetically distinct (Fig. S2C, axis 1–3) made the identification of a clear geographical pattern for the continent very difficult. In Asia, the sPCA revealed the existence of a clear separation of Western Asian accessions with the existence of a genetic gradient eastward (Fig. S2D, axis 1); the separation of Eastern/Northeastern Asian accessions with a genetic gradient westward (Fig. S2D, axis 2); and the separation of accessions from Southeast Asia with a contribution toward the northeast and, to a lesser extent, toward the west (Fig. S2D, axis 3).

Discussion

Both the *STRUCTURE* and the DAPC clustering patterns presented here are congruent with the pattern of genetic structure of American Landraces inferred from SNP markers and principal components analysis reported recently (van Heerwaarden et al. 2011). All clusters were geographically well defined (Fig. S3F) and corresponded to regions or races previously identified as important intermediaries in the American maize evolution (Leff et al. 2004; van Heerwaarden et al. 2011; Vigouroux et al. 2008). Maize genetic origins in the Americas have been extensively studied, but the history of maize outside of the Americas is very recent, starting with the discovery of the American Continent in 1492. Things moved very rapidly from that point; maize quickly became an important staple crop worldwide, and dispersal followed human colonization and trade. Consequently, maize diffused rapidly and repeated maize founder events occurred outside the Americas, leading to a reduction of genetic diversity within populations and to an increased differentiation among populations. This has been documented in European introductions of maize (Rebourg

et al. 2003), and is extended in this study to Africa and Asia.

Genetics meets history: reconstruction of maize diffusion out of the Americas

In light of the clustering results and combining historical and linguistic data, we have attempted a hypothetical reconstruction of maize diffusion out of the Americas (Fig. 5), but we are aware that this is necessarily an oversimplification of the overall events that shaped maize world genetic diversity. Previously documented diffusion of Northern US flints through Europe from northern France eastwards starting in the sixteenth century is confirmed in our analysis, as well as their strong contribution, through hybridization, to Pyrenean-Galician landraces (Dubreuil et al. 2006; Camus-Kulandaivelu et al. 2006) (Figs. 4, 5; Fig. S2B, Fig. S5A). The predominance of Northern US flints in the admixed-ancestry of Portuguese landraces in our data suggests a hybrid origin and perhaps a second independent introduction of Northern US flint populations into Portugal, possibly via Portuguese expeditions in North America in the early sixteenth century (Harrisse 2006). It is probable that multiple Northern US flint introductions into Europe occurred, as supported by the scattering of American accessions throughout the phylogenetic network (Fig. S6A), and by the relatively high genetic diversity present in Europe compared to the Africa and Asia continents.

In the Middle East and Eastern Africa, maize introductions trace back to the Middle North-American maize source (Figs. 4, 5; Fig. S5B). This apparently disagree with reports of early diffusion of Caribbean maize through Southern Europe into Egypt (presumably in 1517) and onwards throughout eastern Africa (e.g., in Ethiopia in 1623) (Portères 1955). Unless the current study did not sufficiently sample Caribbean germplasm, and the related materials simply were not included in the sample, the route identified here is probably more recent than those documented by Portères (1955), leading to a replacement of early Caribbean introductions by the more productive twentieth century US varieties. Phylogenetic clustering (Fig. S6B) confirms the distinct introduction of Middle North-American material into Eastern Africa rather than diffusion from the Middle East. The same ancestral cluster was evidently introduced into Northeastern China as well (Figs. 4, 5; Fig. S5B) where historical data on maize origins are scarce.

Ancestry from the Mexican highlands cluster was found throughout Eastern Asia, especially along the coasts, suggesting maritime introduction(s) (Figs. 4, 5; Fig. S5C). In agreement with historical hypotheses (Chacornac-Rault 2004), phylogenetic clustering (Fig. S6C) suggests an initial introduction into Indonesia, diffusing northwards and toward Japan. This agrees with reports of Portuguese introductions as early as 1496 (Chacornac-Rault 2004) or colonization of the Philippines by Spain in the sixteenth century

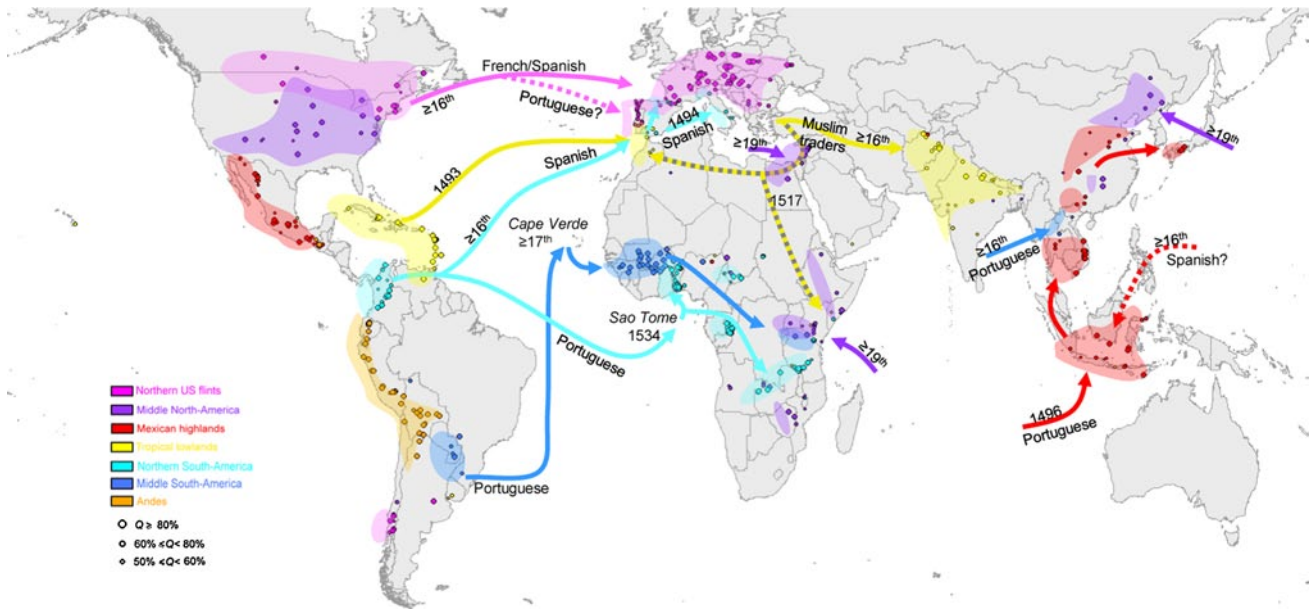

Fig. 5 Hypothetical reconstruction of major routes of maize diffusion out of the Americas combining *STRUCTURE* clustering results and historical data. Each of the 541 non-American accessions and 96 representative American accessions is colored according to their major ancestry Q in one of the seven American ancestral clusters (see legend) applying a cutoff of $Q \geq 50$ %. The size of the circles is proportional to the variation in ancestry. *Hatched arrows* stand for historically unresolved diffusion routes (see text)

(Zaide and Zaide 2004). The Mexican contribution identified here seems more congruent with a Spanish introduction of maize into Eastern Asia, possibly replacing earlier Portuguese introductions. An analysis of landraces from the Philippines (lacking in this study) is needed to further investigate this potential point of introduction.

The Tropical lowlands cluster contributed to southern Spanish maize (Figs. 4, 5; Fig. S5D), in agreement with reports that Columbus returned to Spain in 1493 with maize from the Caribbean (Anghiera 1907; Dubreuil et al. 2006). Tropical lowland ancestry also occurs in Moroccan landraces, perhaps via Spain. However, this is not supported by linguistic studies on Moroccan maize names (Chastanet 1998), which instead suggest an Egyptian-mediated origin of maize. As mentioned previously, an early introduction of Tropical lowland landraces into Egypt may have occurred, only to be later replaced by Middle North-American materials, which is supported by our data. Tropical lowland ancestry was also detected in western Asia from Nepal to Afghanistan. Although maize is recorded in this region from the late sixteenth century, its origin is obscure (Desjardins et al. 2004). Phylogenetic analysis and sPCA analysis (Fig. S2D, Fig. S6D,) support a common origin for the western Asian landraces, with a certain level of sub-clustering within each country. Such a scenario is congruent with diffusion via overland routes from Europe and the Middle East (e.g., the Silk Road) and linguistic data support a role of Arabic traders in the introduction of maize into Western Asia (Desjardins et al. 2004). Tropical lowland ancestry decreases southeastwards through Asia, where Mexican ancestry becomes predominant, making Asia the contact zone between these two diffusion routes, as supported by the clustering pattern identified by sPCA.

The Northern South-American cluster represents an unexpected second contribution to southern European landraces. This is revealed by ancestry found in some Pyrenean, Italian, southern Spanish and Galician landraces, exceeding the contribution of the Tropical lowlands cluster (Figs. 4, 5; Fig. S5E), and is historically supported. In 1514, maize was sent to the Papal court in Italy by the Portuguese, who were established in Colombia (Janick and Caneva 2005). In Spain and France, this contribution may stem from the Spanish colonization of Colombia in the early sixteenth century (Heers 1991). Some Northern South-American and Middle South-American contributions to western sub-Saharan African landraces are identified by both clustering analyses (Figs. 4, 5; Fig. S5E–F). This probably reflects overland diffusion of two independent Portuguese maize introductions as a cheap staple food for slaves in Sao Tome and in the Cape Verde islands, two key transit points in the Portuguese slave trade between Africa and the Americas (Portères 1955; Juhé-Beaulaton 1998; Madeira Santos and Ferraz Torrão 1998). In Cape Verde, historical and linguistic data support the cultivation of maize introduced from Brazil (the Middle South-American cluster) after the sixteenth century (Madeira Santos and Ferraz Torrão 1998). In Sao Tome, the same study reports cultivation of Caribbean maize by 1534, but our genetic data indicate maize from Northern South-America instead. This apparent incongruence may be due to a common confusion in maize names (Juhé-Beaulaton 1998; Madeira Santos and Ferraz Torrão 1998), and since Colombia played a major role in the Portuguese slave trade, the probability is high that maize was sent to Africa from there. Maize diffusion to the African mainland probably occurred first from Sao Tome, where it was cultivated earlier (Madeira Santos and Ferraz Torrão 1998). In the phylogenetic analysis, the absence of strong geographical clustering of Sao Tome (Fig. S6E) and Cape Verde (Fig. S6F) derived accessions suggests long-distance diffusions, congruent with local human trade and movements (Campbell and Tishkoff 2008). We also detected a small Middle South-American contribution in a few eastern Asian landraces (Figs. 4, 5; Fig. S5F), congruent with sixteenth century Portuguese expeditions in Eastern Asia (Lach and Van Kley 1994).

Finally, the Andean ancestral cluster showed no clear evidence of direct diffusion out of the Americas (Figs. 4, 5; Fig. S5G). This may be due to its relative geographical isolation from main trading routes of the sixteenth century, as well as its adaptation to extreme environmental conditions, making its landraces less productive outside their ecological niches (Gouesnard et al. 2002); this continues to be a problem today, and Andean maize is rarely used as a source of maize breeding material.

In addition to these global genetic trends, a few landraces were found to display different ancestry than their geographical neighbors (e.g., one Mexican-like population in Nigeria, one Andean-like population in Central America). This may be interpreted as modern point introductions, although stochasticity of *STRUCTURE* analysis (Vigouroux et al. 2008) and sampling errors cannot be ruled out. In particular, Northern US flint contribution in the southernmost part of South America is likely due to recent introductions as reported by Timothy et al. (1961) and was also found in the study of Vigouroux et al. (2008).

Conclusions

Cluster analyses of American accessions support clear geographical patterns and was also confirmed by Fitch and Neighbor-Net phylogenetic representations (Fig. S4). These patterns are probably the result of demographic and adaptive events accompanying maize expansion north and south from its domestication center in Mexico (Vigouroux et al. 2008; Ducrocq et al. 2008). From our dataset, we

identified seven genetic clusters, which we considered the possible "contribution sources" from which non-American landraces originated, thus offering a first representation of maize diffusion out of the Americas.

The problems presented by hybrid ancestry in hierarchical bifurcating phylogenetic representations were resolved in this study using two complementary clustering analyses (*STRUCTURE* and DAPC), which are better adapted to elucidate complex evolutionary scenarios. Beyond confirming the key role of Northern US flints in maize adaptation to temperate climate outside America, both clustering methods identified introductions of Northern South-American origins into Europe, in addition to those from tropical lowland and Northern Flint origins previously hypothesized by Dubreuil et al. (2006). This study highlighted for the first time Tropical lowland and Mexican origins of Asian landraces, and their contact zone. In Africa, both methods highlighted contributions from the Northern South-American, the Tropical lowlands, and the Middle South-American clusters, partially replaced by more recent introductions of Middle North-American origins.

Overall, our findings were highly congruent with recent reports on American maize history, but also further clarified the patterns of maize diffusion after its domestication. Our study reveals a complex history of maize diffusion out of the Americas. In Europe, Asia and Africa, landraces from distinct American origins coexist. In each of the three continents, this pattern accounts for the substantial genetic differentiation found among accessions (G_{st} ~ 0.30–0.32) and certainly contributed to the transfer of a large proportion of the American genetic diversity worldwide. These transfer patterns can be used to direct the search and use of adaptive characters and guide breeding programs, particularly to expand the genetic base of narrow breeding pools (Goodman 1999) and organize hybrid breeding into complementary heterotic groups (Pray 2006).

Acknowledgments This work was supported by the Generation Challenge Program (grant 3005.14). We thank Drs. N. Lauter, M. Sawkins, M. Chastanet and C. Welcker for comments on manuscript; and the Banco Português de Germoplasma Vegetal for providing Portuguese material. Genetic dataset from this study is available on the Generation Challenge Program Central Registry website, at http://www.generationcp.org/gcp_central_registry.

References

Anghiera P (1907) De Orbe Novo (1st complete ed. 1530). Leroux, Paris

Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255–265

Campbell MC, Tishkoff SA (2008) African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet 9:403–433

Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172:2449–2463

Chacornac-Rault M (2004) Thesis, Museum National d'Histoire Naturelle

Chastanet M (1998) Chapter 9. In: Chastanet M (ed) Plantes et Paysages d'Afrique. Karthala et Cra, Paris, pp 251–282

Desjardins A, McCarthy E, Milho SA (2004) Makka and yu mai: early journeys of Zea mays to Asia National Agricultural Library www.nal.usda.gov/research/maize. Accessed 18 May, 2012

Dubreuil P, Warburton M, Chastanet M, Hoisington D, Charcosset A (2006) More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. Maydica 51:281–291

Ducrocq S, Madur D, Veyrieras JB, Camus-Kulandaivelu L, Kloiber-Maitz M, Presterl T, Ouzunova M, Manicacci D, Charcosset A (2008) Key impact of Vgt1 on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. Genetics 178:2433–2437

Duvick DN (2005) Genetic progress in yield of United States maize (Zea mays L.). Maydica 50:193–202

Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet 16:551–558

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (2005) PHYLIP (Phylogeny inference package). http://evolution.genetics.washington.edu/phylip.html. Accessed 18 May, 2012

Gautier M, Laloë D, Moazami-Goudarzi K (2010) Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PLoS One 5:9

Goodman MM (1999) Broadening the genetic diversity in maize breeding by use of exotic germplasm. In: Coors JC, Pandey S (eds) The Genetics and Exploitation of Heterosis in Crops. WI ASSA-CSSA-SSSA, Madison, pp 139–148

Gouesnard B, Rebourg C, Welcker C, Charcosset A (2002) Analysis of photoperiod sensitivity within a collection of tropical maize populations. Gen Res Crop Evol 49:471–481

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS One 2:367. doi:10.1371/journal.pone.0001367

Harrisse H (2006) Discovery of North America: a critical, documentary and historic investigation. Martino Publishing, Eastford

Heers J (1991) La découverte de l'Amérique. Editions Complexe, Brussels

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Janick J, Caneva G (2005) The first images of maize in Europe. Maydica 50:71–80

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101:92–103

Jombart T, Pontier D, Dufour DA (2009) Genetic markers in the playground of multivariate analysis. Heredity 102:330–341

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94

Juhé-Beaulaton D (1998) Chapter 2 In: Chastanet M (ed) Plantes et Paysages d'Afrique. Karthala et Cra, Paris, pp 45–67

Lach DF, Van Kley EJ (1994) Asia in the making of Europe. University of Chicago Press, Chicago

Leff B, Ramankutty N, Foley JA (2004) Geographical distribution of the major crops across the world. Global Biogeochemical Cycles 18:GB1009. doi:10.1029/2003GB002108

Liu K, Goodman M, Muse S, Smith JS, Buckler E, Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128

Madeira Santos ME, Ferraz Torrão MM (1998) Chapter 3 In: Chastanet M (ed) Plantes et Paysages d'Afrique. Karthala et Cra, Paris, pp 69–83

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

Menkir A, Olowolafe MO, Ingelbrecht I, Fawole I, Badu-Apraku B, Vroh BI (2006) Assessment of testcross performance and genetic diversity of yellow endosperm maize lines derived from adapted x exotic backcrosses. Theor Appl Genet 11:90–99

Monfreda C, Ramankutty N, Foley JA (2008) Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochimical Cycles 22:GB1022. doi:10.1029/2007GB002947

Murgia C, Pritchard JK, Kim S, Fassati A, Weiss R (2006) Clonal origin and evolution of a transmissible cancer. Cell 126:477–487

Pons O, Chaouche K (1995) Estimation, variance and optimal sampling of gene diversity. II. Diploid locus. Theor Appl Genet 91:122–130

Portères R (1955) L'introduction du maïs en Afrique. J Agric Trop Bot Appl 2:221–231

Pray C (2006) The Asian Maize Biotechnology Network (AMBIONET): a model for strengthening national agricultural research systems. CIMMYT, Mexico

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

R Development Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903

Reif JC, Hamrit S, Heckenberger M, Schipprack Wm, Maurer HP, Bohn M, Melchinger AE (2005) Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. Theor Appl Genet 111:906–913

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Sokal RR, Rohlf FJ (1995) Biometry. Freeman and Company, New York

Timothy DH, Peña BV, Ramirez RE, Brown WL, Anderson E (1961) The races of maize in Chile. Natl Acad Sci - Natl Res Counc publ 847, Washington, DC

van Etten J, Hijmans RJ (2010) A geospatial modelling approach integrating archaeobotany and genetics to trace the origin and dispersal of domesticated plants. PLoS One 5(8):e12060. doi:10.1371/journal.pone.0012060

van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, de Jesus Sanchez Gonzalez J, Ross-Ibarra J (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci USA 108:1088–1092

Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Sanchez GJ, Doebley J (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot 95:1240–1253

Warburton ML, Reif JC, Frisch M, Bohn M, Bedoya C, Xia XC, Crossa J, Franco J, Hoisington D, Pixley K, Taba S, Melchinger AE (2008) Genetic diversity in CIMMYT nontemperate maize germplasm: landraces, open-pollinated varieties, and inbred lines. Crop Sci 48:617–624

Witcombe JR, Joshi A, Goyal SN (2003) Participatory plant breeding in maize: a case study from Gujarat, India. Euphytica 130:413–422

Zaide GF, Zaide SM (2004) Philippine History and Government. All-Nations Publishing Company, Quezon City