PNAS

www.pnas.org

Supplementary Information for

Adaptive introgression from maize has facilitated the establishment of teosinte as a noxious weed in Europe

Valérie Le Corre[a][1], Mathieu Siol[a], Yves Vigouroux[b], Maud I. Tenaillon[c], Christophe Délye[a]

[a]Agroécologie, AgroSup Dijon, INRAE, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France
[b]DIADE, Université de Montpellier, IRD, F-34394, Montpellier, France
[c]Génétique Quantitative et Evolution – Le Moulon, Université Paris-Saclay, INRAE, CNRS, AgroParisTech, F-91190, France

[1] Corresponding author. Email**:** valerie.le-corre@inrae.fr

This PDF file includes:

> Figures S1 to S9
> Tables S1 to S4
> Legends for Datasets S1 to S2
> SI References

Other supplementary materials for this manuscript include the following:
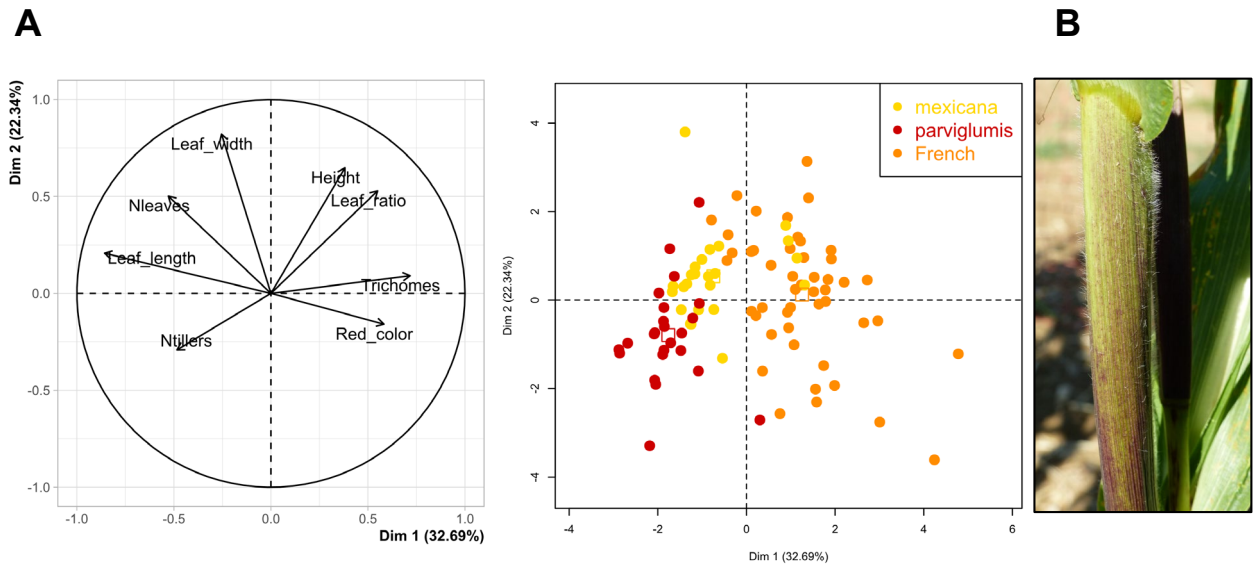
> Datasets S1 to S2

A

B

Figure S1. Phenotypic variation in teosintes assessed from a common garden experiment in Dijon, France. (A) A principal component analysis was conducted on 8 morphological and architectural traits measured on 48 plants of French teosintes, 24 plants of Mexican teosintes of the subspecies mexicana and 24 plants of Mexican teosintes of the subspecies parviglumis. Height: height of plants 45 days after sowing; Leaf_width: largest width of leaves (mean over 3 leaves); Leaf_length: length of leaves (mean over 3 leaves); Leaf_ratio: ratio of mean leaf width to mean leaf length; Nleaves: number of leaves 35 days after sowing; Ntillers: number of primary tillers 40 days after sowing; Trichomes: presence of trichomes on the leaves' sheaths (qualitative variable with 3 levels); Red_color: presence of a red coloration on leaves' sheaths (qualitative variable with 3 levels). (B) Leaf sheath pubescence and red coloration on a French teosinte plant.

Figure S2. Geographical origins of the Mexican teosinte populations included in the dataset (retrieved from Pyhäjärvi et al. 2013 and Aguirre-Liguori et al. 2018). Colors indicate the main genetic cluster assigned to each population according to our FastStructure analysis at *K*=11 (see also Fig. 1 and Dataset S1): red=PARV1, dark red = PARV2, light pink = PARV3, brown = PARV4, yellow = MEX1, gold = MEX2. For parviglumis, the main genetic cluster PARV1 included populations from the Mexican States of Jalisco, Michoacan and Guerrero, two other clusters (PARV2, PARV3) grouped populations from south of Jalisco in the West, and the 4th cluster grouped two populations from Oaxaca in the East (PARV4). For mexicana, the two clusters MEX1 and MEX2 match respectively with the two geographical races «Chalco» and «Central Plateau».
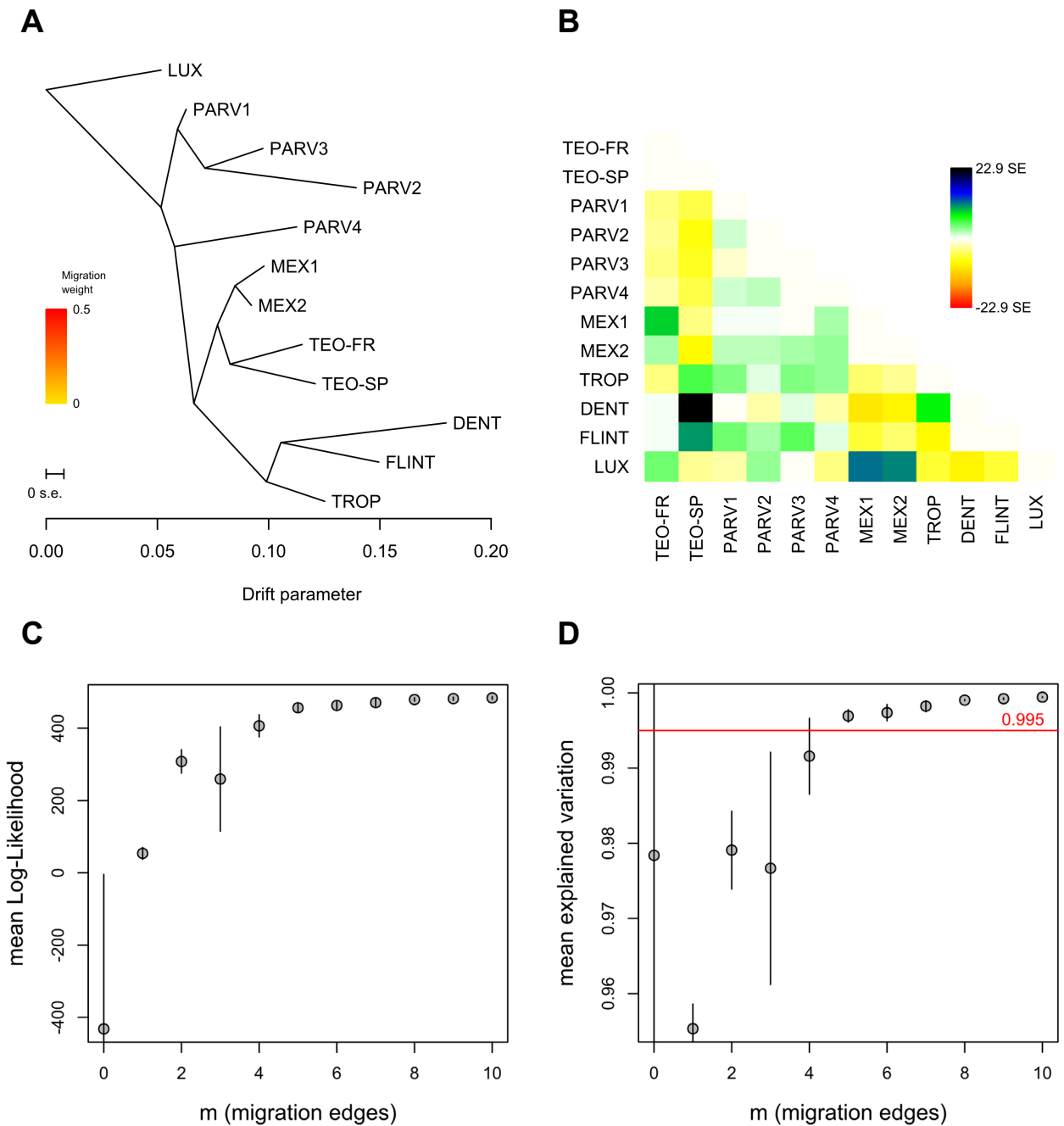
Figure S3. Treemix analyses. (A) Maximum-likelihood tree with no migration showing the relationships among French teosintes (TEO-FR), Spanish teosintes (TEO-SP) and the 9 reference groups identified from the fastStructure analysis in parviglumis (4 groups : PARV1, PARV2, PARV3 and PARV4), mexicana (2 groups : MEX1 and MEX2) and maize (3 groups: tropical landraces TROP, Dent inbred lines DENT and Flint inbred lines FLINT). *Zea luxurians* (LUX) was used as an outgroup. (B) Visualization of residual covariances from the maximum likelihood tree in (A). Residuals above zero represent genetic groups that are more closely related to each other than modelled in the tree and thus are candidates for admixture events. The number of migration events (m) among populations was inferred based on (C), the mean likelihood of the inferred trees as a function of the number of migration events allowed and (D), the mean proportion of explained covariance as a function of the number of migration events allowed. For each value of m, ten maximum likelihood trees were inferred. Based on (C) and (D), we chose m=5 as the most meaningful number of migration events.
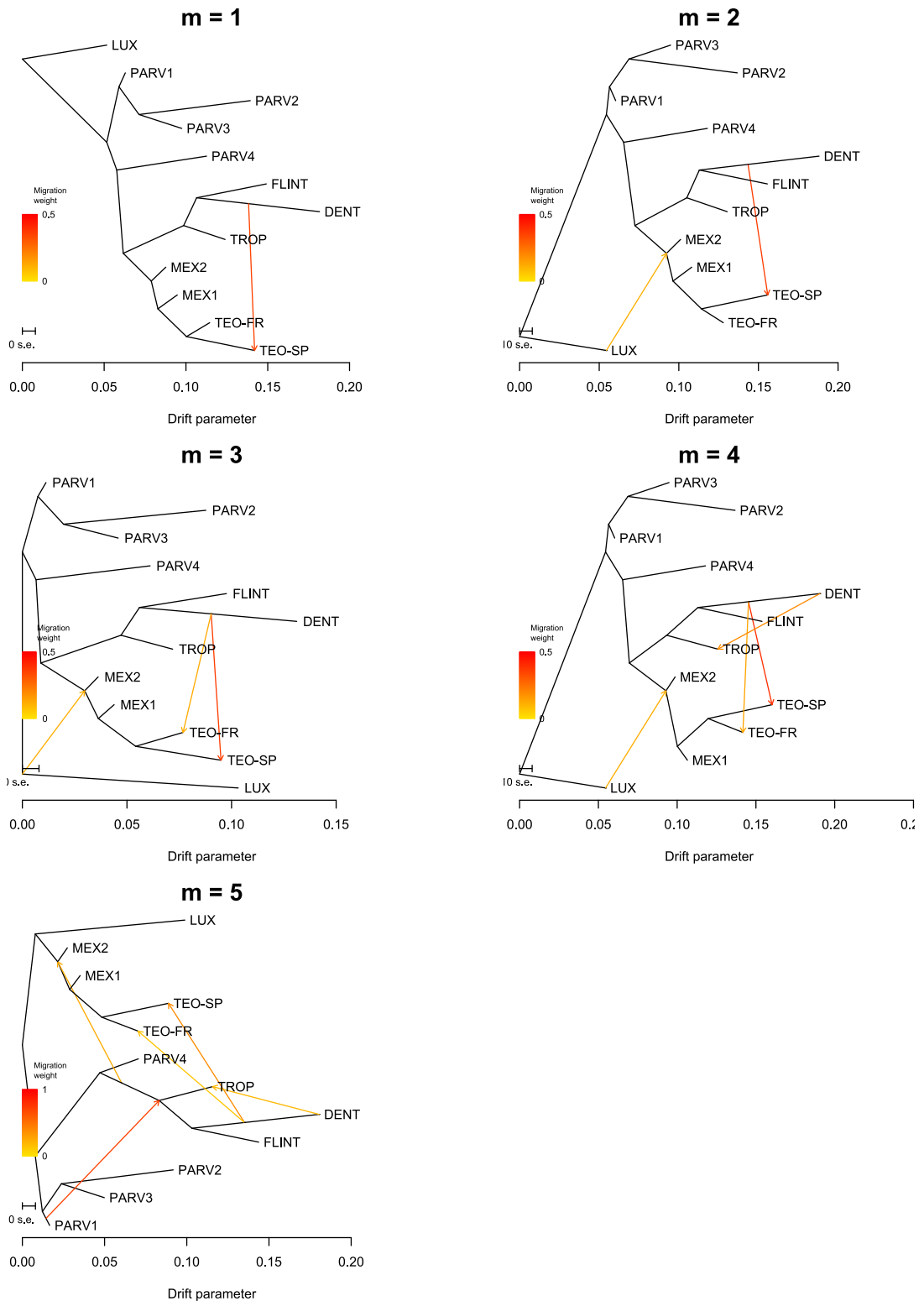
Figure S4. Treemix analyses. Maximum-likelihood trees incorporating 1 to 5 migration events and showing the relationships among French teosintes (TEO-FR), Spanish teosintes (TEO-SP) and the 9 reference groups identified from the fastStructure analysis in parviglumis (4 groups : PARV1, PARV2, PARV3 and PARV4), mexicana (2 groups : MEX1 and MEX2) and maize (3 groups: tropical landraces TROP, Dent inbred lines DENT and Flint inbred lines FLINT). *Zea luxurians* (LUX) was used as an outgroup.
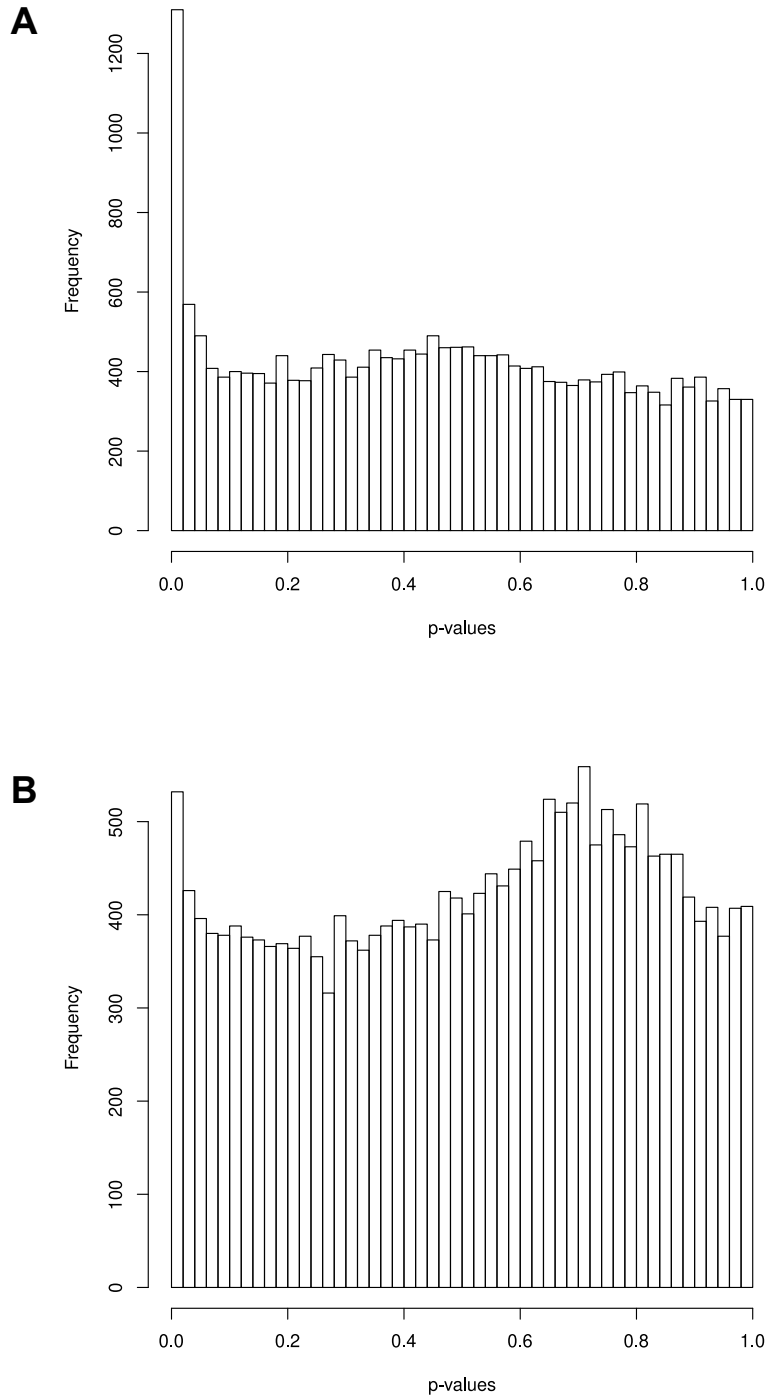
Figure S5. Detection of outlier SNPs associated with divergence between mexicana and European teosintes. Distribution of P-values of the squared loadings of SNPs on (A) the first principal component and (B) the second principal component in pcadapt. An excess of small P-values indicates the presence of outliers.
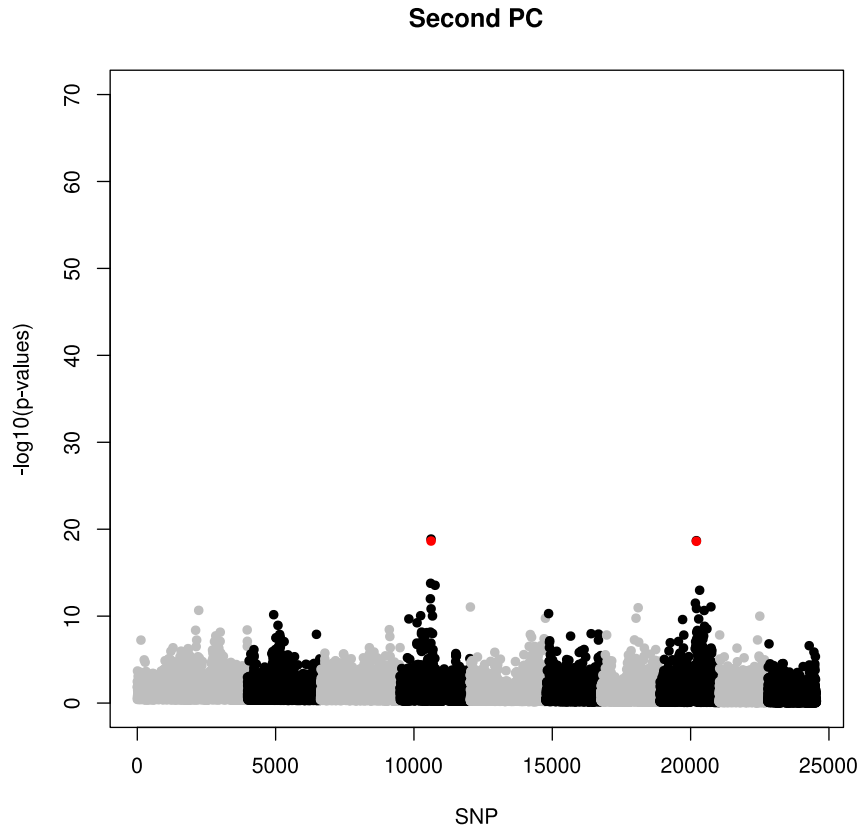
Figure S6. Detection of outlier SNPs along the 10 chromosomes of the genome based on a Principal Component Analysis of mexicana accessions, French teosintes and Spanish teosintes using pcadapt. Manhattan plot of the squared loadings of SNPs with the second principal component of the PCA. The 2 top SNPS having q-values less than 0.1% are displayed in red.
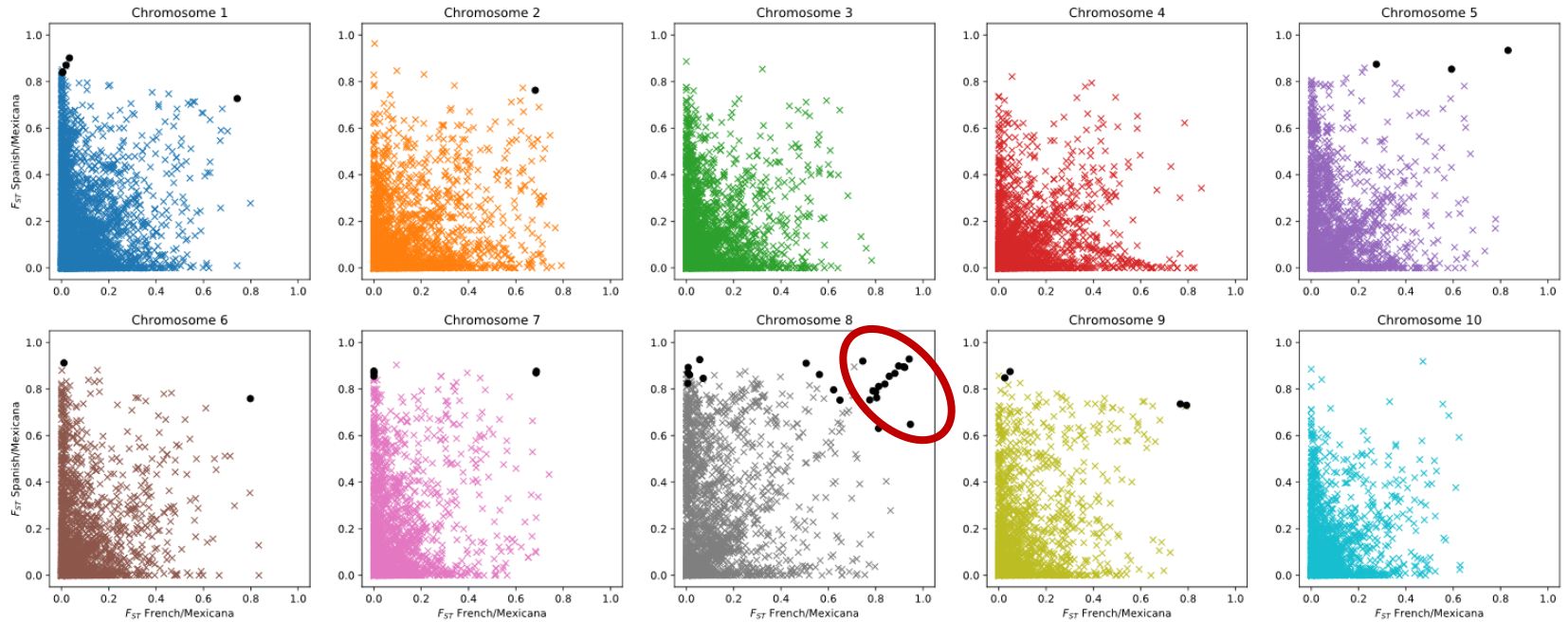
Figure S7. Biplots showing for each chromosome the genetic differentiation (F_{ST}) values between French teosintes and Mexican mexicana teosintes (on the x-axis) *versus* the genetic differentiation (F_{ST}) values between Spanish teosintes and Mexican mexicana teosintes (on the y-axis). Each point represents the F_{ST} value at a single SNP. Outlier SNPs detected by pcadapt are marked with black dots. The red ellipse surrounds a subset of 12 outlier SNPs with elevated F_{ST} values that are all located on a same short region of chromosome 8 (from position 123,949;082 to position 129,332,265 on maize B73 reference v4.0). This region encompasses the candidate gene ZCN8.

Figure S8. Graphical genotype view of genotypes at SNPs along the chromosome 8 (top) and in a region spanning *ZCN8* (bottom), for French teosintes (TEO-FR), Dent maize (DENT), Spanish teosintes (TEO-SP) and Mexican mexicana teosintes (MEX). At each SNP, genotypes homozygous for the allele most frequent within Dent maize are shown in brown, genotypes homozygous for the alternate allele in blue and heterozygotes in green; white stands for missing data. Each line represents an individual, while each column is a SNP. Genetic groups are delimited by white lines. The region considered for the bottom plot is about 6 Mb in length (positions 123,211,593 to 129,332,265 on reference genome B73 v4.0).
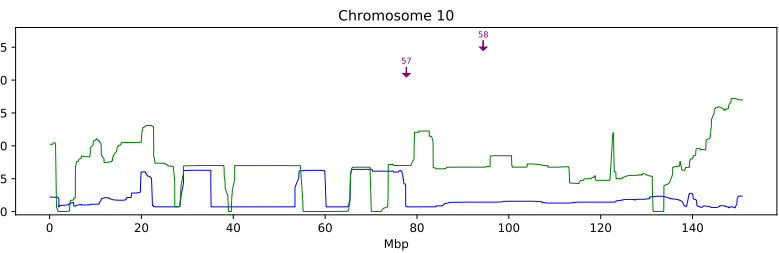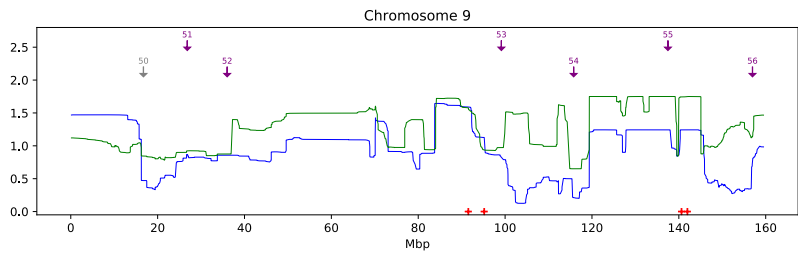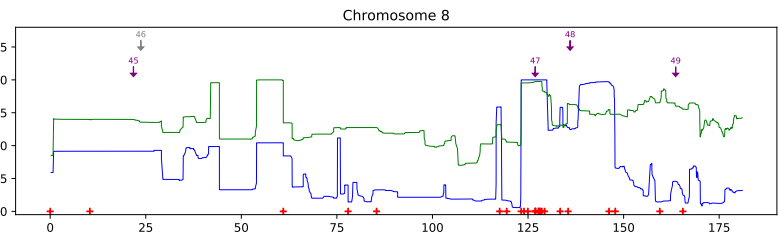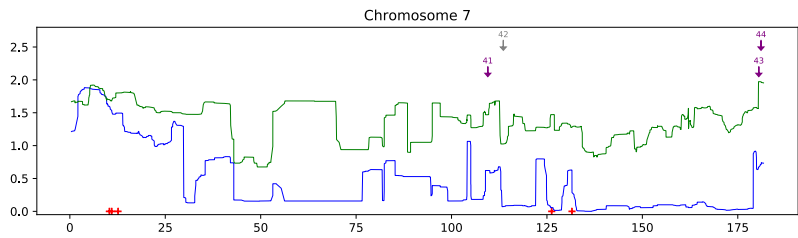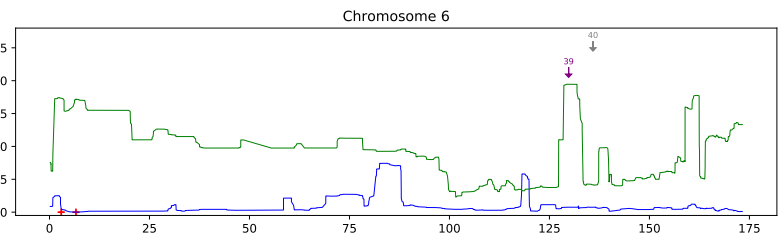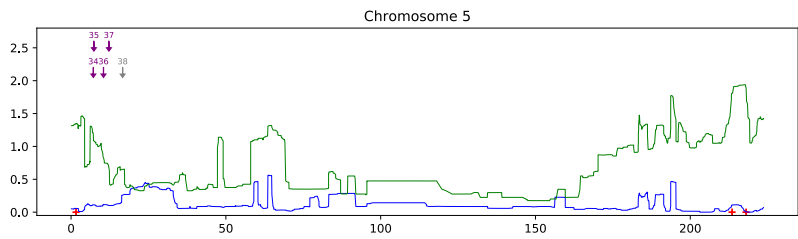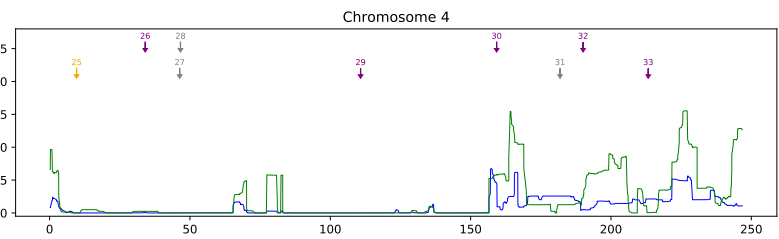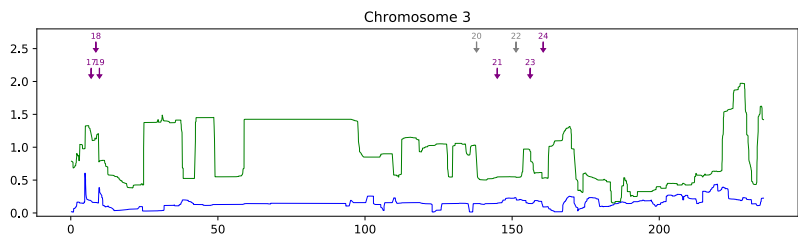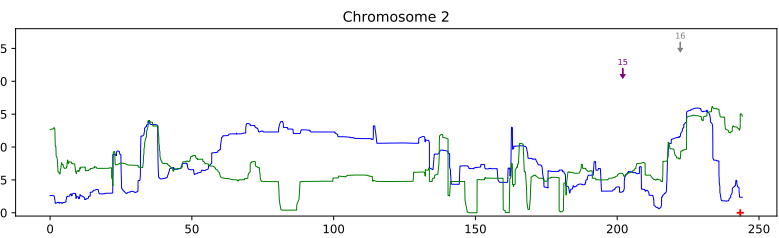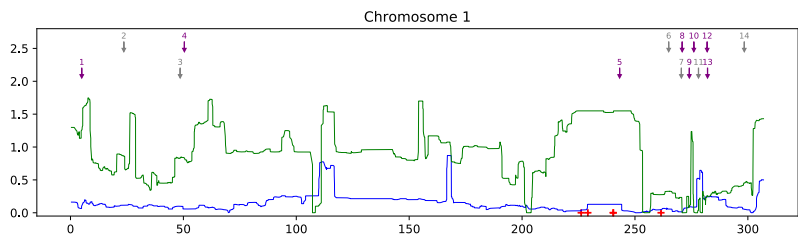
Figure S9. Patterns of local ancestry in European teosintes along the 10 chromosomes of the genome, as inferred by ELAI. The plot shows mean ancestry allele dosages (y-axis) for Dent cultivated maize along the chromosomes (x-axis). The blue line is for French teosintes, the green one for Spanish teosintes. Red dots are the positions of outlier SNPs associated with the first component in the pcadapt analysis and associated with divergence between Mexican mexicana teosintes and European teosintes. The positions of known genes involved in gene networks that control flowering time (Alter et al., 2016; Huang et al., 2018; Liang et al., 2019; Tenaillon et al., 2019); reproductive barriers (gene *GA1*, Zhang et al., 2019) and domestication traits (Chen et al., 2019; Dong et al., 2019; Lin et al. 2012; Studer et al., 2017) are indicated with purple, orange, and grey arrows, respectively. The corresponding list of genes, numbered from 1 to 58 according to their physical position on the genome, is as follows. 1: *ZAG6*, 2: *GT1*, 3: *P1*, 4: *PHYB1*, 5: *ID1*, 6: *YAB12*, 7: *TB1*, 8: *D8*, 9: *PHYA1*, 10: *KN1*, 11: *GIF1*, 12: *PHYC1*, 13: *MADS4,* 14: *TS6*, 15: *PRRH1*, 16: *SMK1*, 17: *CG1*, 18: *GI2*, 19: *D1*, 20: *ZAG2*, 21: *TS4*, 22: *TRU1*, 23: *LDP1*, 24: *MADS69*, 25: *GA1* ,26: *CCA1*, 27: *TGA1*, 28: *NOT1*, 29: *PEBP2*, 30: *SRR1*, 31: *TU1*, 32: *ELF9*, 33: *FKF2*, 34: *ZMM31*, 35: *PHYC2*, 36: *PHYA2*, 37: *D9*, 38: *YAB6*, 39: *GA2ox1*, 40: *ZAG1*, 41: *ELF4*, 42: *RA1*, 43: *PRR37-like*, 44: *DLF1*, 45: *GI1*, 46: *ZMM2*, 47: *PEBP8,* 48: *RAP2.7*, 49: *ELM1*, 50: *ZMM3*, 51: *D3*, 52: *CONZ1*, 53: *GL15*, 54: *CCT2*, 55: *PHYB2*, 56: *MADS1*, 57: *LHY1*, 58: *CCT1*. See Dataset S2 for more details on the proposed list of candidate genes.

Figure S10. Pictures of teosinte spikes collected in maize fields in France. (A), Typical teosinte spike. (B) and (C), rarely observed, hybrid-like spikes.

Table S1. Nei's expected heterozygosity (*He*) for each of the French teosintes, the Spanish teosintes, the parviglumis teosintes from Mexico, the mexicana teosintes from Mexico and the maize inbred lines, and pairwise genetic differentiation (F_{ST}) values between them. Mexican teosinte accessions and maize accessions were grouped according to the main genetic cluster assigned by the fastStructure analysis at K=11.

	FRA	SPA	PARV1	PARV2	PARV3	PARV4	MEX1	MEX2	TROP	DENT	FLINT
He	0.221	0.251	0.305	0.131	0.220	0.202	0.273	0.258	0.350	0.360	0.350
FRA		0.237	0.188	0.447	0.347	0.352	0.153	0.158	0.25	0.294	0.316
SPA			0.205	0.456	0.353	0.354	0.213	0.206	0.222	0.226	0.275
PARV1				0.222	0.133	0.172	0.117	0.095	0.151	0.214	0.215
PARV2					0.332	0.473	0.354	0.327	0.323	0.365	0.394
PARV3						0.335	0.267	0.237	0.246	0.297	0.314
PARV4							0.261	0.237	0.243	0.289	0.315
MEX1								0.056	0.208	0.275	0.284
MEX2									0.193	0.257	0.264
TROP										0.109	0.147
DENT											0.148

Table S2. Tests for admixture with maize using the four-populations test $f4$(H, mexicana ; maize, parviglumis). The statistics was never different from zero when the tested population H was MEX1, a result consistent with the reference populations MEX1 and MEX2 being admixed with neither maize nor parviglumis. Significantly positive values were observed when H was a European teosinte population (TEO-FR, TEO-SP), which supported the occurrence of admixture with maize.

H	Mexicana	Maize	$f4$	SE	Z-score	P-value
MEX1	MEX2	TROP	0.00066	0.000505	1.32	0.19
MEX1	MEX2	DENT	0.00040	0.000584	0.69	0.49
MEX1	MEX2	FLINT	0.00031	0.000497	0.64	0.52
TEO-FR	MEX1	TROP	0.00377	0.000767	4.91	< 10-6
TEO-FR	MEX2	TROP	0.00444	0.000722	6.15	< 10-6
TEO-FR	MEX1	DENT	0.01084	0.001177	9.2	< 10-6
TEO-FR	MEX2	DENT	0.01124	0.001108	10.15	< 10-6
TEO-FR	MEX1	FLINT	0.00760	0.000990	7.68	< 10-6
TEO-FR	MEX2	FLINT	0.00792	0.000964	8.21	< 10-6
TEO-SP	MEX1	TROP	0.01335	0.001100	12.16	< 10-6
TEO-SP	MEX2	TROP	0.01401	0.001022	13.72	< 10-6
TEO-SP	MEX1	DENT	0.03099	0.001853	16.72	< 10-6
TEO-SP	MEX2	DENT	0.03139	0.001759	17.84	< 10-6
TEO-SP	MEX1	FLINT	0.01828	0.001304	14.02	< 10-6
TEO-SP	MEX2	FLINT	0.01860	0.001220	15.25	< 10-6

Table S3. List of population samples used for phenotypic assays (common garden experiment and herbicide sensitivity bioassays).

ID	subspecies	Country	Locality	Latitude	Longitude	Year	Common garden (number of plants)	Herbicide Sensitivity (number of plants)
TEO17-01	-	France	Guidier	46.0121	-0.1104	2017	4	25
TEO17-02	-	France	Les Gours	45.9601	-0.0844	2017	6	25
TEO17-03	-	France	Chives	45.9654	-0.1326	2017	6	25
TEO17-04	-	France	Lupsault	45.9404	-0.0614	2017	11	25
TEO17-05	-	France	Lupsault	45.9435	-0.0656	2017	3	25
TEO17-06	-	France	Lupsault	45.9449	-0.0630	2017	6	25
TEO17-07	-	France	Les Gours	45.9702	-0.0632	2017	6	25
TEO17-08	-	France	Saint-Fraigne	45.9407	-0.0268	2017	6	25
CM10	mexicana	Mexico	VillaSeca	19.4075	-99.6271	2010	4	-
SMH573	mexicana	Mexico	Puerto de Cabras	19.9828	-100.9599	2009	5	-
SMH575	mexicana	Mexico	San Rafael	20.0532	-101.0881	2009	3	-
SMH576	mexicana	Mexico	Tejocote de Clalera	20.1607	-101.3731	2009	3	-
SMH579	mexicana	Mexico	Armadillo	20.1338	-101.4342	2010	4	-
SMH580	mexicana	Mexico	Capacho	19.9646	-101.2354	2010	5	-
CM01	parviglumis	Mexico	Teloloapan	18.3498	-99.8411	2010	4	-
CM04	parviglumis	Mexico	Tecoanapa	16.9811	-99.2855	2010	4	-
CM08	parviglumis	Mexico	Paso_Morelos	18.2377	-99.218	2010	4	-
CM12	parviglumis	Mexico	Guachinango	20.6273	-104.4079	2010	4	-
MIT14	parviglumis	Mexico	San_Lorenzo	19.9458	-103.9958	2010	4	-
SMH566	parviglumis	Mexico	El_Estanco	19.9027	-104.1572	2009	4	-

4

Table S4. Primers used for *ACC1* sequencing and genotyping

Primer code	Primer sequence (3'-5')	Primer pair target (nucleotide positions in Genbank accession XM_020548014)
AC1ZM-2	CTTGGAGGAATTCCTGTGGGCGTC	12,593-13,926
AC1ZM-2R	GTGTTGGATGTGTTGGTATCATCAAT	
AC1ZM-3	GAAGTCATGGCAGTCCAATGGC	11,491-12,692
AC1ZM-3R	CAGATCGCTCATGGGAATCAAGC	
AC1ZM-4	TTCCTCTGGTTGGATGACAAGAGT	10,239-11,579
AC1ZM-4R	CAGCAAACACTAGCTCAGTTGCC	
AC1ZM-5	GAGGAACAAAGCTAAGCTTGTAACG	9,058-10,301
AC1ZM-5R	AGGCTCCACATGCCGGAGAATC	
AC1ZM-6	TTTGGTATGCTGCCTGGACAAC	7,919-9,143
AC1ZM-6R	AGCGAACTAACAGATCCCTGTAAC	
AC1ZM-7	ACTGTAAGGACTGGACATGGTAGC	6,767-7,973
AC1ZM-7R	AGCCTCGTTGCTAGAACAGACATA	
AC1ZM-14	GAGGAAAACAGCAGCTCTTGCTA	5,488-6,826
AC1ZM-14R	CGCTTCAACTGTTGAATCATTCATTCT	
AC1ZM-9	GTTCAGTTGCTTTGTGATCAATATGGT	4,839-5,637
AC1ZM-9R	CTAGAAAACTTACCTTCACTTTCCCA	
AC1ZM-10	GTGAAATATTGCAGATGGCACAAAAAC	3,593-4,922
AC1ZM-10R	GACCTTCTGGTGTCGCCGTTGC	
AC1ZM-11	CGACCTCCCAAGTGAGGCACC	2,568-3,672
AC1ZM-11R	TGCGGTCAATACATCTGGCAGT	
AC1ZM-12	CTGGGCCACGGAACGACAATGTCA	1,876-3,149
AC1ZM-12R	GATTGATAATCCCATTCATTTGATAA	

Dataset S1 (excel file). Full list of the1005 accessions of European teosintes, Mexican teosintes and maize inbred lines genotyped at 24,544 SNPs, and their assignment to 11 genetic clusters as determined by the software fastStructure.

Dataset S2 (excel file). A proposed list of 58 candidate genes involved in gene networks that control flowering time, reproductive barriers and domestication traits in maize, with their physical positions on B73 reference genome v4.0.

SI References

1. Aguirre-Liguori JA et al. (2017) Connecting genomic patterns of local adaptation and niche suitability in teosintes. Molecular Ecology 26, 4226-4240. DOI: https://doi.org/10.1111/mec.14203.
2. Alter et al. (2016) Flowering time-regulated genes in maize include the transcription factor ZmMADS1. Plant Physiol 172, 389-404. DOI: https://doi.org /10.1104/pp.16.00285.
3. Chen et al. (2019) TeoNAM: A Nested Association Mapping Population for Domestication and Agronomic Trait Analysis in Maize. Genetics 213, 1065-1078. DOI: https://doi.org/10.1534/genetics.119.302594.
4. Dong et al. (2019) The regulatory landscape of a core maize domestication module controlling bud dormancy and growth repression. Nature Communications 10, 3810. DOI: https://doi.org/10.1038/s41467-019-11774-w.
5. Huang et al. (2018) ZmCCT9 enhances maize adaptation to higher latitudes. Proc Natl Acad Sci E334-E341. DOI: https://doi.org /doi/10.1073/pnas.1718058115
6. Liang et al. (2019) ZmMADS69 functions as a flowering activator through the ZmRap2.7-ZCN8 regulatory module and contributes to maize flowering time adaptation. New Phytologist 221, 2335–2347. DOI: https://doi.org/10.1111/nph.15512.
7. Lin et al. (2012) Parallel domestication of the Shattering1 genes in cereals. Nature Genetics 44, 720-725. DOI: https://doi.org/10.1038/ng.2281.
8. Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in teosinte. Genome Biology and Evolution 5, 1594–1609. DOI: https://doi.org/10.1093/gbe/evt109.
9. Studer et al. (2017) Selection during maize domestication targeted a gene network controlling plant and inflorescence architecture. Genetics 707, 755-765. DOI: https://doi.org/10.1534/genetics.117.300071.
10. Tenaillon et al. (2019) Transcriptomic response to divergent selection for flowering time in maize reveals convergence and key players of the underlying gene regulatory network. BioRxiv, 461947, ver. 5 peer-reviewed and recommended by PCI Evolutionary Biology. https://www.biorxiv.org/content/10.1101/461947v5.
11. Zhang et al. (2019) A PECTIN METHYLESTERASE gene at the maize Ga1 locus confers male function in unilateral cross-incompatibility. Nature Communications 9, 3678. DOI: 10.1038/s41467-018-06139-8.