

Characterization of introgression from the teosinte *Zea mays* ssp. *mexicana* to Mexican highland maize

Eric Gonzalez-Segovia¹, Sergio Pérez-Limon¹, G. Carolina Cántora-Martínez¹, Alejandro Guerrero-Zavala¹, Garrett M Janzen², Matthew B Hufford², Jeffrey Ross-Ibarra³, Ruairidh J H Sawers^{Corresp. 1}

¹ Unidad de Genómica Avanzada (LANGEBIO), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, Mexico

² Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, United States

³ Department of Plant Sciences, Center for Population Biology, and Genome Center, University of California, Davis, Davis, United States

Corresponding Author: Ruairidh J H Sawers

Email address: rusawers@cinvestav.mx

Background. The spread of maize cultivation to the highlands of central Mexico was accompanied by substantial introgression from the endemic wild teosinte *Zea mays* ssp. *mexicana*, prompting the hypothesis that the transfer of beneficial variation facilitated local adaptation.

Methods. We used whole-genome sequence data to map regions of *Zea mays* ssp. *mexicana* introgression in three Mexican highland maize individuals. We generated a genetic linkage map and performed Quantitative Trait Locus mapping in an F₂ population derived from a cross between lowland and highland maize individuals.

Results. Introgression regions ranged in size from several hundred base pairs to Megabase-scale events. Gene density within introgression regions was comparable to the genome as a whole, and over one thousand annotated genes were located within introgression events. Quantitative Trait Locus mapping identified a small number of loci linked to traits characteristic of Mexican highland maize.

Discussion. Although there was no strong evidence to associate quantitative trait loci with regions of introgression, we nonetheless identified many Mexican highland alleles of introgressed origin that carry potentially functional sequence variants. The impact of introgression on stress tolerance and yield in the highland environment remains to be fully characterized.

Title: Characterization of introgression from the teosinte *Zea mays* ssp. *mexicana* to Mexican highland maize

Running Title: Teosinte introgression to maize

Area 4: Molecular Ecology and Evolution

Eric Gonzalez-Segovia^{1*}, Sergio Pérez-Limon¹, G. Carolina Cíntora-Martínez^{1†}, Alejandro Guerrero-Zavala¹, Garrett M. Janzen², Matthew B. Hufford², Jeffrey Ross-Ibarra³, Ruairidh J. H. Sawers^{1‡}

¹Unidad de Genómica Avanzada (LANGEBIO), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, Mexico

²Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, United States

³Department of Plant Sciences, Center for Population Biology, and Genome Center, University of California, Davis, Davis, United States

*present address: School of Life Sciences, Technical University of Munich, Freising, Germany.

†present address: Max Planck Institute for Plant Breeding Research, Cologne, Germany.

‡present address: Department of Plant Science, Pennsylvania State University, State College, United States

Corresponding Author:

Ruairidh Sawers

Unidad de Genómica Avanzada (LANGEBIO), Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, Mexico

e-mail address: ruairidh.sawers@gmail.com

28 **ABSTRACT**

29 **Background.** The spread of maize cultivation to the highlands of central Mexico was
 30 accompanied by substantial introgression from the endemic wild teosinte *Zea mays* ssp.
 31 *mexicana*, prompting the hypothesis that the transfer of beneficial variation facilitated local
 32 adaptation.

33 **Methods.** We used whole-genome sequence data to map regions of *Zea mays* ssp. *mexicana*
 34 introgression in three Mexican highland maize individuals. We generated a genetic linkage map
 35 and performed Quantitative Trait Locus mapping in an F₂ population derived from a cross
 36 between lowland and highland maize individuals.

37 **Results.** Introgression regions ranged in size from several hundred base pairs to Megabase-scale
 38 events. Gene density within introgression regions was comparable to the genome as a whole, and
 39 over one thousand annotated genes were located within introgression events. Quantitative Trait
 40 Locus mapping identified a small number of loci linked to traits characteristic of Mexican
 41 highland maize.

42 **Discussion.** Although there was no strong evidence to associate quantitative trait loci with
 43 regions of introgression, we nonetheless identified many Mexican highland alleles of
 44 introgressed origin that carry potentially functional sequence variants. The impact of
 45 introgression on stress tolerance and yield in the highland environment remains to be fully
 46 characterized.

INTRODUCTION

Maize (*Zea mays* ssp. *mays*) was domesticated in southwestern Mexico approximately 9,000 years ago from an annual wild grass, the teosinte *Zea mays* ssp. *parviglumis* (Matsuoka et al. 2002; Piperno et al. 2009). Following domestication, maize dispersed across Mexico and diversified to give rise to locally-adapted landrace varieties (Wellhausen et al. 1952; Sanchez G. and Goodman 1992; Ruiz Corral et al. 2008; Perales and Golicher 2014). Cultivated maize spread rapidly beyond the ancestral niche occupied by *parviglumis* (Piperno 2006; Merrill et al. 2009; Grobman et al. 2012), raising questions as to the origin and nature of the genetic variation underlying adaptive phenotypic change. Within Mexico, the colonization of the highland Central Plateau represents a clear example of niche expansion with respect to the *parviglumis* ancestor (Piperno 2006). The Mexican highlands present a unique combination of environmental challenges to maize cultivation: low temperatures necessitate planting early in the year in order that plants might reach maturity, in turn risking exposing seedlings to frost and water deficit before the onset of annual rains; throughout the growing season, low-temperature, high-levels of UV radiation and hail storms pose further difficulties (Eagles and Lothrop 1994; Lafitte and Edmeades 1997; Jiang et al. 1999; Mercer et al. 2008); the volcanic origin of the central highland region, and the associated acidic nature of the soils, restricts the bioavailability of phosphorus (Bayuelo-Jiménez and Ochoa-Cadavid 2014).

The first maize to be cultivated in the Mexican highlands encountered not only new environmental challenges, but also the highland-adapted teosinte *Zea mays* ssp. *mexicana* (hereafter, *mexicana*. Hufford et al. 2012a). Mexican highland maize and *mexicana* share morphological traits (most obviously, pronounced stem pigmentation and pubescence) that are rarely seen in *parviglumis* and lowland maize (Wellhausen et al. 1953; Wilkes 1976). This

common morphology presented the first indication of introgression between *mexicana* and maize, interpreted variously as the adoption of adaptive traits by maize or as mimicry by teosinte to evade removal from cultivated fields (Wilkes 1972; Lauter et al. 2004). Subsequent molecular studies have demonstrated shared ancestry between highland maize and *mexicana* (Doebley 1990; Matsuoka et al. 2002; van Heerwaarden et al. 2011; Hufford et al. 2013). Introgression from *mexicana* to maize is more common than in the opposite direction, with *mexicana* contributing around 20% of the genome of Mexican highland maize varieties (van Heerwaarden et al. 2011; Hufford et al. 2013; Romero Navarro et al. 2017; Wang et al. 2017). The use of increasing numbers of molecular markers and whole genome sequence datasets has allowed the mapping of large-scale regions of introgression within Mexican highland maize genomes (Hufford et al. 2013; Wang et al. 2017). A number of introgressed haplotypes are found at high frequency in Mexican highland maize, while being rare or absent in lowland populations, consistent with a possible role in local adaptation (Hufford et al. 2013; Takuno et al. 2015; Romero Navarro et al. 2017; Wang et al. 2017).

Introgression has been proposed as a mechanism whereby invasive populations can rapidly acquire the genetic variation necessary to adapt to local environmental challenges (Hedrick 2013; Martin and Jiggins 2017). For example, introgression of genes from *Heliconius melpomene* butterflies to *H. timareta* has been linked to the acquisition of mimetic red wing patterns (Pardo-Diaz et al. 2012). Similarly, it has been suggested that the adaptation of *Arabidopsis arenosa* to serpentine soils was facilitated by introgression from the related species *A. lyrata* (Arnold et al. 2016). In cultivated barley, introgression from wild relatives appears to have aided expansion and adaptation (Poets et al. 2015). The expansion of weedy *Helianthus annuus* into certain regions of Texas was driven by admixture with the wild relative *H. debilis*,

promoting increased herbivore resistance (Whitney et al. 2006, 2015). Introgression from crops to their wild-relatives has also been documented, directly enhancing vigor (Gutierrez et al. 2011), or promoting persistence of weedy forms in cultivated fields through mimicry (Chen et al. 2004; Xia et al. 2011).

The fate of any given locus following hybridization will depend not only on the adaptive value of its associated variants but also the local genomic landscape in terms of recombination rates and the nature of linked genes (Martin and Jiggins 2017). For example, studies in mice and humans have found weaker signals of introgression in areas of the genome where gene density is high and/or the recombination rate is low (Sankararaman et al. 2014; Janoušek et al. 2015; Sankararaman et al. 2016; Martin and Jiggins 2017). In the context of crop plants, the potential benefits of introgression from wild relatives with respect to local adaptation or stress tolerance may be offset by decreased agronomic value as a result of linked pre-domestication gene content. It has been hypothesized that introgression will be facilitated by high local levels of recombination, allowing beneficial or neutral alleles to readily recombine away from undesirable variants (Barton and Bengtsson 1986; Martin and Jiggins 2017). Nonetheless, previous reports have described a number of Megabase-scale introgression events that have been transferred from *mexicana* to maize, associated with regions of low recombination in pericentromeric regions or chromosomal inversions (Hufford et al. 2013; Wang et al. 2017). If such large scale events are to be maintained by positive selection, the net effect of many hundreds of wild alleles must be beneficial. Although a low rate of recombination makes it hard to break linkage-drag, the flipside may be to promote introgression by holding together groups of beneficial alleles that collectively constitute an adaptive haplotype (Kirkpatrick and Barton 2006; Twyford and Friedman 2015; Marques 2017).

In this study, we performed a detailed characterization of *mexicana* introgression in three Mexican highland maize individuals. Introgression events were characterized with respect to their size, gene content and variation. In addition, we generated a linkage map to estimate local recombination rate. To assess the impact of gene flow on plant morphology and phenology, we mapped QTL for a number of traits in an F₂ mapping population and compared their location to the introgression map.

MATERIALS AND METHODS

Plant material

Sequence data reported here was generated from two outbred individuals of accession Mexi5 of the landrace Palomero Toluqueño (PT; collected near the city of Toluca, Mexico state at 19.29N, -99.57W, 2,597 masl), obtained from the collection of the International Center for Maize and Wheat Improvement (CIMMYT) seed bank, and one outbred individual of accession TC313 of landrace Mushito de Michoacán (MM; collected south of Pátzcuaro, Michoacán at 19.31N, -101.68W, 2271 masl), from the collection of Alfredo Carrera, Universidad Autónoma Chapingo, Michoacán. Samples for sequencing were collected from mature field grown plants RS16-1032.6 (PT1), RS16-1032.9 (PT2) and RS13-1261.1 (MM). Individual PT1 was crossed as male to a partially inbred stock derived from the Mexican landrace Reventador (RV; an S6 derivative of accession Nay15, INIFAP. Bukowski et al. 2018), and a single individual of the resulting F₁ stock self-pollinated to generate an F₂ family. A total of 170 RV x PT1 F₂ individuals were evaluated in a lowland field site (Valle de Banderas, Nayarit. Winter cycle 2017) and genotyped to generate a genetic linkage map and perform QTL mapping (see below).

Whole genome sequencing

Total genomic DNA was extracted by LANGEBIO-CINVESTAV Genomic Services (<http://www.langebiodcinvestav.mx/?pag=458>). Whole genome sequencing was performed by NGX-Bio (San Francisco, California, USA. 94107) using the Illumina HiSeq 3000/4000 platform, using HiSeq 4000 SBS chemistry to generate 150 bp paired-end reads. A total of 85.5, 87.7, and 144.7 Gb of sequence was generated for PT1, PT2, and MM1, respectively. Genome sequence data is available from National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (PRJNA511379).

Public sequence data

Additional whole genome sequence data for the lowland maize landraces Nal Tel (RIMMA0703) and Zapalote Chico (RIMMA0733) was taken from Wang et al., 2017 (NCBI SRA SRP065483), and an extra lowland sequence (BKN022), *mexicana* (TIL08 and TIL25) *Tripsacum* (TDD39103) and parviglumis (TIL01, TIL05, TIL10) from (Bukowski et al. 2018) (/iplant/home/shared/panzea/hapmap3/bam). Genome sequence was obtained in bam format aligned to the B73 reference genome v3 using BWA mem (BWA v.0.7.12. Li 2013).

Pre-processing of whole genome sequence data

The PT1, PT2 and MM1 sequence was processed using Trimmomatic v.0.32 (Bolger et al. 2014), set for pair-end data with the following parameters: LEADING: 3 TRAILING: 3 SLIDINGWINDOW: 4:15 MINLEN: 36. The resulting trimmed sequences, both paired and single end, were mapped against the B73 reference genome v3 using BWA v.0.7.12 (Li. 2013), under the default settings, using the -M option for Picard compatibility. The resulting sam

output was sorted, and converted to bam format using Picard tools v.2.4.1 (<http://broadinstitute.github.io/picard>). Single-end and pair-end ordered bam files for each individual were merged using samtools v.1.3.1 (Li et al. 2009), and duplicated molecules removed with Picard tools using default parameters, with the flag REMOVE_DUPLICATES=true. The files were indexed using Picard Tools and indel realignment carried out with the Genome Analysis Toolkit (GATK) v.3.5.0 (McKenna et al. 2010). Publicly available lowland maize, *mexicana* and *Tripsacum* bam format sequences were processed using the same pipeline from the removal of duplicate molecules onwards.

Calculation of D and f_d

Genotype likelihoods (GL) were calculated using ANGSD v.0.912 (Korneliussen et al. 2014) with the following parameters: -GL 1 -remove_bads 1 -nThreads 8 -doGlf 3 -doMajorMinor 1 -doMaf 1 -SNP_pval 1e-6 -minInd 2 -minMapQ 30 -minQ 20. Inbreeding values were calculated for each individual with ngsF v1.2.0 (Vieira et al. 2013), using the script ngsF.sh with the following parameters: --n_threads 20 --n_ind 9 --min_epsilon 1e-6 --glf <GL calculated with ANGSD> --n_sites <number of called SNPs>, giving the following values: RIMMA0703, 0.22; RIMMA0733, 0.30; TIL25, 0.59; TIL08, 0.66; TDD39103, 0.64; BKN022, 0.69; PT1, 0.00084; PT2, 0.00156; MM, 0.00042; TIL10, 0.49; TIL05, 0.56; TIL01, 0.48. Inbreeding coefficients were used in a ANGSD SNP calling to account for deviations of the HWE using the following parameters: -SNP_pval 1e-6 -GL 1 -doMajorMinor 1 -doMaf 1 -rf -remove_bads 1 -minMapQ 30 -minQ 20 -minInd 4 -doGeno 4 -doPost 1 -postCutoff 0.95 -indF <inbreedingValues> . Allele frequencies were used to calculate f_d and D using the script ABBA_BABA.v1.pl (Owens et al. 2016), based on the tree (((P 1 , P 2), P 3), O), where the P1 position was the three

lowland genomes BKN022, RIMMA0703 and RIMMA0733, P2 was the three highland genomes PT1, PT2 and MM, P3 was the two teosinte *mexicana* genomes TIL08 and TIL25, and O was the *Tripsacum* genome TDD39103. From this analysis, we calculated the average value of D across the genome, along with an associated p value estimated from the distribution of D values calculated in non-overlapping windows of 5 Mb across the genome (ABBA_out_blocker.pl; Jackknife_ABBA_pipe.R; ABBA_pvalue.R. Owens et al. 2016). To calculate D at the level of individual chromosomes, we used non-overlapping windows of 1 Mb. To map introgression within the genome, a custom R script was used to calculate f_d in non-overlapping windows of 50 informative (ABBA/BABA) sites, based on the ABBA_BABA.v1.pl output. We considered the sets of the top 1 % and 10 % scoring windows as positive for introgression (Table S1, S2). Raw f_d output is provided in Table S3. We also generated a null data set, substituting South American for Mexican maize (Table S4). The average of f_d at the whole genome was taken as the proportion of introgression. The location of gene models in introgression regions is listed in Table S5.

Genotypic analysis of a RV x PT F₂ family and construction of genetic linkage map

Total genomic DNA was extracted from 170 F₂ individuals derived from the cross of RV x PT1 using the Qiagen (GmbH) DNeasy Plant Mini kit DNA extraction kit, according to the manufacturer's instructions. Samples were analysed by the International Center for Maize and Wheat Improvement (CIMMYT) using the DaRT platform (<http://seedsofdiscovery.org/es/catalogo/saga-servicio-de-analisis-genetico-para-la-agricultura/>. (Sansaloni et al. 2011). Using tag sequences of ~65 bp, a total of 26, 727 SNPs were identified, with < 50 % missing data. Tags were anchored to physical positions in the B73 v3 reference

208 genome using BLAST (Altschul 1990), under the following parameters: min % for each base =
 209 3, max % for each base = 60, e-value = $5e^{-10}$, max hits per sequence = 10, percent overlap = 90,
 210 percent identity = 90. Tags aligning to multiple positions and those that contained multiple
 211 SNPs, were discarded, as were tags derived from heterozygous sites for which only one allele
 212 could be aligned under the defined parameters, and sets of two or more tags that aligned to a
 213 common position. The exact position for each SNP was calculated on the basis of the position of
 214 the SNP within the tag and the position of the alignment of the tag against B73. The resulting set
 215 of 10,323 SNPs were transformed to hapmap format, and filtered to identify segregating sites
 216 (allele frequency > 0.2 and < 0.8), thinned to minimum spacing of 1 kb, and transformed to ABH
 217 format (A: PT; B: RV) using TASSEL v5.0 (Bradbury et al. 2007). The data were inspected
 218 visually using ABHgenotypeR v1.0.1 (Furuta et al. 2017), and passed to the ABHgenotypeR
 219 pipeline to impute missing data, corrected for under-called heterozygous sites, and corrected for
 220 single interspersed alleles using a maximum haplotype length of 6. The proportion of missing
 221 sites dropped from 0.57 to 0.02 following imputation. The final proportion of sites was 24.4%
 222 A, 23.6% B, 52.0% H. Markers were assigned to linkage groups and ordered based on the B73
 223 v3 physical map prior to estimation of the genetic map using R v3.4.0 (R Core Team 2014) with
 224 R/QTL v1.41.6 (Broman et al. 2003), following the recommendations available at
 225 <http://www.rqtl.org/tutorials/geneticmaps.pdf>. The marker set was reduced once more on the
 226 basis of redundancy in the genetic map using the functions `qtl::findDupMarkers` and
 227 `qtl::drop.markers`, resulting in a final set of 1,166 SNPs that was passed to the function
 228 `qtl::est.map` with the kosambi mapping function. Genetic and physical distances were extracted
 229 per chromosome using `qtl::pull.map`. Local estimates of recombination rates (RR) (cM/Mb) were
 230 obtained using R/MareyMap v1.3.4 (Rezvoy et al. 2007), fitting a cubic spline across each

chromosome, using the parameters $\text{spar} = 0.05$ and $\text{df} = 10$ (Table S6, S7). Markers that distorted the monotonic increase of the fitted spline were removed by hand to avoid negative rates. The genetic map of the Maize Nested Association Mapping Population was obtained from MaizeGDB (Andorf et al. 2016) and physical positions were taken as the midpoint of genes associated with the markers, with local estimates of RR calculated as described above.

Functional annotation of sequence variants

The SNPs obtained with ANGSD were converted to hapmap format using custom scripts, passed to TASSEL v5.2.43, and converted to vcf format, the reference allele at any given site being defined based on the B73 reference genome v3, set using bcftools v1.5. To perform functional annotation, the vcf file was passed to SnpEff v4.3 (Cingolani et al. 2012; Table S8). A custom R script was used to select SNPs homozygous for the alternative allele in the three highland maize genomes. Population differentiation data for Mexican highland and lowland maize populations was taken from (Takuno et al. 2015). Maize gene annotation was taken from maize-GAMER (<https://dill-picl.org/projects/gomap/maize-gamer/>).

QTL mapping of variation in morphological traits and flowering time

The 170 F_2 RV x PT1 individuals used for the linkage mapping were grown to maturity in a lowland winter nursery (Valle de Banderas, Nayarit, Mexico. 20.8 N, -105.2 W, 54 masl), and evaluated for the following traits: plant height (PH), ear height (EH), stem pigment intensity (INT), stem pigment extent (EXT), stem macrohair pattern (MPAT); stem macrohair density (MDEN), tassel (male inflorescence) branch number (TBN), tassel length (TL) and days-to-anthesis (DTA). INT was evaluated on a semi-quantitative scale from 0 - 4. EXT was scored as

0%, 25%, 50%, 75% or 100%. MAPT was scored as 0 (no stem macrohairs), 1 (marginal macrohairs only), 2 (patchy macrohair production on the sheath) or 3 (uniform macrohair production on the sheath). MDEN was scored semi-quantitatively from 0 to 4. Other traits were evaluated as described previously (Flint-Garcia et al. 2005). QTL mapping was conducted using a single-scan in R/QTL (Broman et al. 2003), with the support of R/QTLtools (Lovell et al. 2018). Phenotypic data and the genetic map are provided as an R/QTL cross object in Table S9.

RESULTS

Introgression from *mexicana* is distributed throughout the genome of Mexican highland maize

To characterize introgression from *mexicana* to Mexican highland maize, we generated whole genome sequence data from two outbred individuals of the landrace Palomero Toluqueño (PT1 and PT2) and a single outbred individual of the landrace Mushito de Michoacán (MM1), yielding a coverage of ~40 fold for PT1 and PT2, and ~70 fold for MM1 (Table S10). In total, we identified 71,623,944 single nucleotide polymorphisms (SNPs) across the three individuals. To estimate the extent of *mexicana* introgression, we calculated Patterson’s *D* statistic (Durand et al. 2011) and genome-wide f_d (Martin et al. 2017). Briefly, working with genomic sequence from highland and lowland maize, *mexicana* and the related grass *Tripsacum* (see Materials and Methods), we identified those sites that were polymorphic between *Tripsacum* and *mexicana*, and compared the frequency with which highland maize carried the *mexicana* allele and lowland maize the *Tripsacum* allele (the “ABBA” pattern) to the frequency of the complementary case (the “BABA” pattern) (Green et al. 2010). A total of 905,537 SNPs were characterized as following either the ABBA or BABA pattern, and, therefore, to be informative for the analysis.

Our analysis revealed strong evidence of shared ancestry between *mexicana* and our highland maize samples (D, Z score > 9.56; Table 1), with *mexicana* introgression estimated by f_d to account for ~7% of the highland maize genomes.

To localize introgression within the genome, we calculated f_d in non-overlapping windows of 50 informative sites. We considered relatively high-scoring windows to be positive for introgression, selecting the sets of the top 1% and top 10% outliers for further analysis. Adjacent positive windows were concatenated, defining 80 (8.86 Mb or 0.43% of the genome) and 679 (101.98 Mb or 4.95% of the genome) introgression events for 1% and 10% sets, respectively (Fig. 1; Tables S1, S2). In light of our genome-wide estimation of 7% *mexicana* ancestry, both the 1% and 10% sets appear to be conservative. Introgression events were distributed across all ten chromosomes, although not uniformly, with chromosome (chr) 4 particularly enriched (Fig. 1A; Table 1). In the 1% outlier set, 29 of a total of 80 events were located on chr 4, representing 58% of the total introgressed DNA by size. For the 10% outlier set, 90 of 679 events were located on chr 4, representing 24% of the total by size. Genomewide, the size of introgression events in the 1% outlier set ranged from 0.53 to 630 kb. The upper limit was substantially increased in the 10% set, with events ranging from 0.34 to 4,700 kb, indicating that many of the windows between the first and tenth percentile were clustered in the genome. In both 1% and 10% outlier sets, the majority (95% in both cases) of the events were less than 0.5 Mb in size, with these small events also constituting the majority of the total physical introgression size (Fig. 2B). Nine events were identified that were > 1 Mb in size, all from the 10% outlier set. These included events that co-localized with previously reported Megabase-scale introgression regions (Fig. 1A; Table S11; Hufford et al. 2013; Romero Navarro et al. 2017; Wang et al. 2017): the *Inv4m* inversion polymorphism on chr 4 (located at 169 - 180 Mb;

represented as fifteen closely located events in our analysis), a region on chr 6 (located at 46 - 57 Mb; four events in our analysis) and a region on chr 3 (located at 75 - 90 Mb; three events in our analysis) .

To assess the potential functional significance of *mexicana* introgression, we examined the gene content of the introgression events on the basis of the B73 reference genome (Table S5). While previously reported Megabase-scale regions harbor a large number of annotated genes, we considered the possibility that smaller regions were largely distributed in gene-poor sections of the genome, perhaps as remnants of historical gene flow, experiencing little purifying selection as a result of limited functionality. Overall, the cumulative distribution of gene number as a function of ordered physical size conformed with the expectation of genome-wide gene density, with a total of 1,380 genes (3.5% of the tested genes) found inside the 10% outlier events (Fig. 1C; Table S5). There was no indication that small introgression events were gene-poor, and it was at the higher end of the size spectrum that gene-density fell slightly below the genome wide value, consistent with the location of many of the larger events in pericentromeric regions.

Large introgression events are located in regions of low genetic recombination

The most reproducible signals of introgression from *mexicana* to Mexican highland maize are associated with Megabase-scale events that co-localize with putative chromosomal inversions (e.g. regions on Chr 3 and Chr 4 reported here and previously by Hufford et al. 2013; Romero Navarro et al. 2017; Wang et al. 2017), consistent with the hypothesis that a low local rate of genetic recombination can favor introgression (Kirkpatrick and Barton 2006). Of course, it is also clear that such large scale events are easier to detect. To characterize the recombination landscape of the PT genome, we generated a genetic linkage map from the cross of PT1 and the

lowland Mexican landrace Reventador (RV; partially inbred accession used in the f_d analysis). The total map length was 1,275 cM, with a global recombination rate (RR) of 0.61 cM/Mb (based on the size of the B73 v3 physical map). At the level of individual chromosomes, RR ranged from 0.69 on chr 1 to 0.49 on chr 8 (Table 1). As is typical, local RR values were high in the telomeric regions and low around the centromeres (Fig. 2A, Table S6). In addition, we observed variation across the genome with clear recombination hot and cold spots (Fig. 2A). For each introgression event, RR was estimated based on the midpoint location. The RR differed depending on the size of the introgression events (Kruskal-Wallis test, $p < 0.001$, Fig 2B): while small (< 250 kb) events were distributed across a range of RR, large regions (> 250 kb) were constrained to regions where $RR < 0.5$ (Fig 2B), with the exception of one event on chr3 ($RR = 0.58$) and one on chr 4 ($RR = 1.91$), although, in both cases, RR was reduced compared with their surroundings (Tables S6, S7). Of the 32 events > 0.5 Mb identified from the 10% outlier set, 18 were located in pericentromeric regions (defined as the region for which $RR \leq 0.2$ extending from the estimated position of the centromere, Fig. 2A; Table S12), suggesting that we have more large introgression events at pericentromeric regions than expected ($\chi^2 = 7.73$, $df = 1$, p value = 0.005), consistent with the idea that highland maize carry centromeric or pericentromeric regions from *mexicana* (Hufford et al. 2013). We compared our map with the maize Nested Association Mapping (NAM) population reference (McMullen et al. 2009), which did not include highland maize material in its construction. We found that the non-pericentromeric regions harboring large (> 0.5 Mb) introgression events in our analysis presented lower local rates of recombination than the corresponding positions in the NAM map (paired-sample Wilcoxon test, p value < 0.001 ; Table S12), suggesting that these regions themselves may be suppressing recombination. The strong signal of introgression across *Inv4m* coincided

clearly with a region of low RR , consistent with the segregation of inverted and standard haplotypes in our PT x RV cross (Fig. 3. Hufford et al. 2013; Romero Navarro et al. 2017; Wang et al. 2017).

Introgression events contribute to the differentiation of Mexican highland and lowland maize

To better understand the importance of introgression in the differentiation of highland and lowland Mexican maize, we examined a previously published F_{ST} data set (Takuno et al. 2015), comparing genes inside and outside of our introgression events (Table S5). For 409 genes reported to show significant differentiation (from a total of 21, 029 genes for which an F_{st} estimate was available), 62 (15%) were located in the set of 10% outlier introgression events, an enrichment over the genome-wide expectation ($\chi^2 = 205$, $df = 1$, p value < 0.001 . Table S13). This trend was driven, in part, by the large number of high F_{ST} genes within introgression events on chr 4 (34 of the total of 62 high F_{ST} genes within introgression), although the enrichment remained even after removal of chr 4 from the data set ($X^2 = 57$, $df = 1$, p -value < 0.001 . Table S14). Across the 21, 029 genes for which an estimate was available, the median F_{ST} value was significantly higher for 636 genes located in introgression regions than for the remaining 20, 393 genes (Wilcox test, p value < 0.001). When chromosomes were considered individually, chr 3, 4, 8 and 9 showed a significant ($p < 0.01$) difference in F_{ST} (Fig. 4A).

Introgression events contain variation of potential significance for protein function

To assess functional variation, we categorized the SNPs identified in genes on the basis of their possible impact on encoded proteins using SnpEff (Cingolani et al. 2012). Genes located in

369 introgression events showed an excess of high ($\chi^2 = 35.03$, $df = 1$, p value < 0.001 . Table S15)
 370 and moderate ($\chi^2 = 222.72$, $df = 1$, p -value < 0.001 . Table S16) effects fixed in our sample of
 371 three genomes (six alleles). In total, 40 of the 1, 380 genes in introgression events were
 372 homozygous for the alternate allele at one of more high-effect SNPs across all three highland
 373 maize individuals (45 SNPs in total; Table S8). An additional 502 genes were fixed for the
 374 alternate allele at one or more moderate-effect SNPs (1, 740 SNPs in total; Table S8). We further
 375 categorized fixed high- and moderate- effect SNPs with respect to the other samples in our
 376 analysis (Fig. 4B. Table S8). The majority of SNPs in introgression regions fixed for the
 377 alternate allele in our highland maize samples were also fixed for the alternate allele in *mexicana*
 378 (1, 346 of 1, 785 SNPs). Of these, 61 highland-fixed SNPs (located in 40 genes) unambiguously
 379 followed the “ABBA” pattern used initially for the selection of introgression regions (*i.e.* fixed
 380 for the alternate allele in highland maize and *mexicana*; fixed for the reference allele in
 381 *Tripsacum*, lowland maize, and, although not included in the earlier analysis, also *parviglumis*).
 382 A further 168 highland-fixed SNPs (located in 96 genes) were not called in *Tripsacum* (and
 383 therefore, were not used in our estimation f_d), but were fixed for the alternate allele in highland
 384 and *mexicana* samples and fixed for the reference allele in lowland and *parviglumis* samples,
 385 consistent with introgression. The largest category resulting from this grouping consisted of 193
 386 SNPs (located in 104 genes) that were not called in *Tripsacum*, were fixed alternate in highland
 387 and *mexicana* samples, and segregating in lowland and *parviglumis* samples. Such SNPs
 388 differentiate highland and lowland individuals in both teosinte and maize, and their distribution
 389 within putative introgression events is consistent with an introgressed origin. A total of 70
 390 highland-fixed SNPs (located in 41 genes), including 48 in the *Inv4m* region, were private to our
 391 highland genomes.

To combine variant effect prediction with annotated gene function, we cross referenced the list of genes in introgression events with the classical maize gene list, a curated set of 4,908 well characterized genes (the “combined set” gene list was obtained from www.maizegdb.org/gene_center/gene and filtered for unique gene identifiers). Considering the classical genes located in introgression regions as a whole, a diverse range of functions are represented, many that are potentially significant to morphology or environmental responses (Table 2). Intriguing examples include the *Bx8* gene required for the biosynthesis of benzoxazinoid defense compounds (Frey et al. 1997), the phosphorus homeostasis gene *Pho1;2a* (Salazar-Vidal et al. 2016), the flowering-time locus *Gi2* (Mendoza et al. 2012), and various genes related to phytohormone biosynthesis, ear morphology and grain development (Table 2). These last include the genes *Compact Plant2* (*Ct2*), *Fasciated Ear3* (*Fea3*) and *Tunicate* (*Tu1*) that play a role in the regulation of plant meristems (Han et al. 2012; Je et al. 2016; Wu et al. 2018; Han et al. 2012), the genes *Nana Plant1* (*Na1*) and *Nana2-like1* (*Natl1*) involved in brassinosteroid biosynthesis (Hartwig et al. 2011; Best et al. 2016), *Aminocyclopropane carboxylate oxidase20* (*Acco20*) involved in ethylene biosynthesis (Mira et al. 2016), and the genes *Dwarf8* (*D8*) and *Kaurene oxidase1* (*Ko1*) that play a role in gibberellic acid signaling (Peng et al. 1999).

To look at the possible implication of post-domestication gene flow during the early development of cultivated maize, introgression events were compared with the location of previously reported domestication and improvement genes (Hufford et al. 2012b). Of 420 reported domestication candidates present in the B73 v3 reference genome annotation (*i.e.* genes showing a reduction in diversity and increased differentiation between teosinte and landrace maize), 17 (3.6 %) were located in introgression events (based on 10% outliers). Similarly, of

529 annotated improvement candidates (*i.e.* genes showing a reduction in diversity between landrace maize and modern inbred lines), 22 (3.8 %) were located in introgression regions. For both domestication and improvement candidates the proportion within introgression events mirrored the genome-wide value of 3.5% (domestication: $\chi^2 = 0.2185$, $df = 1$, p value = 0.64; improvement: $\chi^2 = 0.48451$, $df = 1$, p value = 0.48). As such, we see no evidence that these candidates are refractory to introgression.

Introgression events on chromosome 9 co-localize with a previously-reported QTL for sheath pubescence

One of the most striking morphological characteristics of Mexican highland maize is the presence of pronounced stem pubescence (Fig. 5A. Wellhausen 1952). In a previous study, evidence of *mexicana* introgression was identified on chr 9 (106.5 Mb - 125.5 Mb. Hufford et al. 2013), co-localizing with *macrohairless1* (*mh1*), a locus linked with production of macrohairs on the adaxial surface of the leaf blade in inbred maize lines (~115 Mb. Moose 2004). We recovered a single introgression event in this region of chr 9 in our 1% outlier set, and a number of events in our 10% outlier set (Table S1; S2). A previous experiment to map sheath pubescence in a cross between *parviglumis* and *mexicana* identified a major effect QTL on the long-arm of chr 9, consistent with the action of *mexicana*-specific neomorphic allele of *mh1* extending the production of macrohairs from leaf blade to sheath (Lauter et al. 2004). In a further study, using recombinant inbred lines derived from the cross between B73 and PT, there was also evidence to link a QTL in the *mh1* region to stem pubescence (Aguilar-Rangel, PhD Thesis, 2018). Here, we attempted to map stem pubescence in the F₂ progeny of our PT x RV cross. Upon evaluation, however, we found that the majority of the F₂ plants (140 of 157) presented stem macrohairs

(scored on a semi-quantitative scale). While this provided insufficient variation for successful QTL mapping, it may indicate the action of multiple dominant-acting factors. We confirmed that the lowland RV parent did not show stem pubescence. We anticipate that the use of inbred material to reduce the confounding effects of dominance, along with fine-scaled quantitative evaluation, would provide a better characterization of the genetic architecture of stem macrohair production in the PT x RV cross.

QTL peaks associated with morphological and flowering traits do not co-localize with introgression events

In addition to stem pubescence, we evaluated the PT x RV F₂ population for a number of further morphological and flowering time traits to explore any possible association with introgression. QTL were identified associated with stem pigment intensity (INT), stem pigment extent (EXT), tassel (male inflorescence) branch number (TBN), tassel length (TL) and days-to-anthesis (DTA, Table 3). The QTL intervals themselves were large (typically, tens of Mbs), and, necessarily, contained multiple introgression events. In no instance, however, did the marker closest to a QTL peak fall directly within one of our introgression events (Fig. 5B, Table 3). Given the limited resolution of our mapping, we looked for instances where we might identify a candidate gene within a given QTL interval for the purpose of evaluating local introgression; two such candidates are discussed below.

The qTBN-7 interval contains the candidate gene *Ramosa1* (*Ra1*, GRMZM2G003927, Chr 7: 110 Mb). The *Ra1* product has been characterized to restrict production of long-branches in both the male and female inflorescence (Vollbrecht et al. 2005). When we examined the window containing *Ra1* in our analysis, we found no evidence of introgression ($f_d = 0.20$). This

is perhaps not too surprising given that *mexicana*, although described to present lower TBN than *parviglumis* (Doebley 1983), does not present the extreme reduction in tassel branching that is characteristic of Mexican highland maize.

The stem pigment QTL qINT-2 and qPAT-2 overlap on chr 2, defining an interval that contains the candidate gene *BI* (GRMZM2G172795. Chr 2: 19 Mb). The *BI* gene encodes a basic helix-loop-helix transcription factor that regulates the tissue-specific biosynthesis of anthocyanins (Goff et al. 1992; Sharma et al. 2011). In contrast to greatly reduced tassel branching, stem pigmentation is a trait shared by Mexican highland maize and *mexicana*. Furthermore, there is evidence that allelic variation at *BI* is linked to stem pigmentation differences between *parviglumis* and *mexicana* (Selinger and Chandler 1999; Lauter et al. 2004). Nonetheless, inspection of the window containing *BI* in our analysis found no evidence of introgression ($fd = 0.22$), indicating that although variation at *BI* may drive stem pigmentation in both *mexicana* and Mexican highland maize, the underlying alleles may have independent origins.

DISCUSSION

Study of a number of crops has begun to indicate the significance of post-domestication gene flow during the diversification and spread of cultivated varieties (Hufford et al. 2013; Poets et al. 2015; Bredeson et al. 2016; Rendón-Anaya et al. 2017). In line with previous reports (Hufford et al. 2013; Wang et al. 2017), we have detected significant genome-wide introgression from *mexicana* to Mexican highland maize. We estimated genome-wide introgression at ~7%, and we mapped specific introgression events constituting ~5 % of the total physical space. Our values are somewhat lower than other estimates, that have ranged up to 20% (Matsuoka et al. 2002;

Hufford et al. 2013; Wang et al. 2017), reflecting the conservative criteria we employed, and, potentially, the result of differences among the individuals in our sampling. Specifically, the differing sources of PT and MM samples might be reflected in distinct patterns of introgression. Similarly, we do not distinguish between introgression proceeding or contemporary with the early diversification of the Mexican highland landrace group, and later events that might be restricted to specific populations or races. Our analysis will largely have identified events that are shared between PT and MM samples, consistent with both high frequency in contemporary populations, and an origin early in the development of the Mexican highland group.

Previous reports of *mexicana* introgression to Mexican highland maize have focused on Mb-scale regions, events also recovered in our analysis. In addition, we also identified large numbers of small (< 500 kb) events, that, collectively, constituted the bulk of the total introgression. Introgression events presented a gene-density equivalent to the genome as a whole, contrary to the hypothesis that they were harbored in gene-poor regions, possibly remnants of historical gene-flow with little functional significance. Nonetheless, 262 events (~10% of the total introgression by size), with an average size of ~37 kb, did not contain any annotated genes. A further 162 events contained only a single annotated gene. Small, single gene events, presumably the result of recombination following historical hybridization, would not be impacted by the negative effects of linked deleterious alleles, nor would they participate in hitchhiking through linkage to beneficial variants (Barton and Bengtsson 1986; Kirkpatrick and Barton 2006). As such, the persistence of single-gene introgression events would be predicted to reflect directly the fitness effects of associated allelic variants. Across all introgression events, we identified a number of genes that have previously been demonstrated to play major roles in maize development and growth, including well-characterized hormone-signaling genes.

Variation at loci related to phytohormone signalling has the potential to trigger pleiotropic effects, impacting, for example, flowering time, morphology and stress tolerance. Indeed, the capacity to retune simultaneously multiple aspects of plant morphology, phenology and stress biology make hormone pathways compelling candidates as drivers of a collective adaptive syndrome. Gene-level analyses were conducted on the basis of the B73 genome annotation. Given the extensive copy-number variation known to be present in maize, it will be interesting to re-evaluate introgression gene content when further genome assemblies become available for landrace maize and maize wild-relatives.

Among the genes identified to be in introgression events, we recovered previously characterized domestication and improvement candidates, at a rate equivalent to that observed genome-wide. This observation somewhat contradicts an earlier report finding that regions of introgression from *mexicana* to maize harbored fewer domestication candidates, while regions resistant to such introgression were enriched for domestication candidates (Hufford et al. 2013). It may be significant that we employed a greater number of markers than were used in the previous report, and identified a larger number of small events. All the domestication candidates we identified in introgression regions were in events < 775 kb in size (although four candidates were located in small events co-localizing with the *Inv4m* region).

Although many SNPs with potential functional relevance were identified in well-supported gene models, we were unable to link introgression to the phenotypic traits we evaluated. In the case of stem pubescence, we did not observe sufficient variation to permit QTL mapping. Nonetheless, we did identify introgression events in the *mh1l* region of chr 9, that has been previously linked with stem pubescence in both *mexicana* and Mexican highland maize in other studies (Lauter et al. 2004; Aguilar-Rangel, PhD Thesis, 2018). For stem pigmentation and

tassel branch number, we mapped large-effect QTL that co-localized with high-confidence candidate genes. The *Ra1* candidate has also been linked to tassel branch number variation in a maize x *parviglumis* population, the *parviglumis* allele increasing branching three-fold with respect to the maize allele (Xu et al. 2017). PT is characterized by greatly reduced tassel branching (often the tassel is a single, unbranched spike), consistent with a gain of *Ra1* function with respect to typical maize varieties, and reminiscent of the phenotype seen in maize *liguleless* and *unbranched* mutants (Wellhausen 1952; Walsh and Freeling 1999; Chuck et al. 2014). As such, we can hypothesize an allelic series of increasing *Ra1* function from *parviglumis*, through lowland maize, to Mexican highland maize. With respect to stem pubescence, a number of functional variants of *BI* have been described, and differences in stem pigmentation linked to transposon insertion in the region upstream of *BI* (Radicella et al. 1992; Selinger and Chandler 1999; Selinger and Chandler 2001). Significantly, two previously characterized *BI* alleles from *mexicana* were reported to present a different upstream structure to an allele from the Mexican highland landrace Cacahuacintle (Selinger and Chandler 1999). Although the resolution of our mapping limited broader conclusions concerning the impact of introgression, in *Ra1* and *BI* we have identified compelling candidates linked to large-effects in two of the most characteristic morphological traits of Mexican highland maize. The fact that we see no evidence for introgression at either *Ra1* or *BI* may suggest that founder populations in the Mexican highlands contained sufficient standing genetic diversity for these characteristic traits to arise without recourse to gene flow. It is interesting to note that stem pigmentation is also prevalent in the demographically distinct maize races of highland South America, where introgression from *mexicana* is considered to be absent (Wellhausen, 1957; Wang et al., 2017).

Notwithstanding the results of our QTL analysis, the extent of *mexicana* introgression,

and the number of variants identified, argues for a functional impact. The *Inv4m* region has previously been linked to flowering time in a large-scale association analysis (Romero Navarro et al. 2017). The apparent discrepancy with respect to our observations may reflect the fact that the evaluation was carried out in a lowland environment, or result from epistatic interactions that we were not capable of detecting in our experiment. Indeed, it may well be that the broader phenotypic effects associated with introgression are conditional on growth at high elevation, acting in the modification of major QTL, in responses to biotic and abiotic stress, and in subtle, but significant, contributions to yield and harvest quality traits.

CONCLUSION

We detected significant genome-wide introgression from *mexicana* to Mexican highland maize. Employing conservative criteria, we mapped specific introgression events within the genome, constituting ~5 % of the total physical space. Introgression events presented a gene-density equivalent to the genome as a whole, and contained a significant number of genes that have previously been demonstrated to play major roles in maize development and growth. Although potential functional variants were identified, we were unable to link introgression to phenotypic traits. While it is possible that that founder populations in the Mexican highlands contained sufficient standing genetic diversity to support the adoption of the basic morphology and phenology characteristic of modern highland varieties, the extent of *mexicana* introgression, and the number of variants identified, nonetheless argues for a functional impact. We suggest that this impact might yet be identified in modification of major QTL, in responses to biotic and abiotic stress, and in contributions to yield and harvest quality traits under highland conditions.

ACKNOWLEDGEMENTS

We thank Sherry Flint-Garcia and Susan Melia-Hancock for assistance with the collection of phenotypic data. We thank Ana-Laura Alonso Nieves for assistance with photography. Maize gene annotations were kindly compiled by Ricardo Chavez Montes.

REFERENCES

- Aguilar-Rangel R (2018) The role of introgression from teosinte (*Zea mays* ssp. *mexicana*) in the adaptation of maize to the highlands of Mexico. PhD Thesis.
- Altschul S (1990) Basic local alignment search tool. **J Mol Biol** 215: 403–410
- Andorf CM, Cannon EK, Portwood JL 2nd, Gardiner JM, Harper LC, Schaeffer ML, Braun BL, Campbell DA, Vinnakota AG, Sribalasu VV, Huerta M, Cho KT, Wimalanathan K, Richter JD, Mauch ED, Rao BS, Birkett SM, Sen TZ, Lawrence-Dill CJ (2016) MaizeGDB update: new tools, data and interface for the maize model organism database. **Nucleic Acids Res** 44: D1195–201
- Arnold BJ, Lahner B, DaCosta JM, Weisman CM, Hollister JD, Salt DE, Bomblies K, Yant L (2016) Borrowed alleles and convergence in serpentine adaptation. **Proc Natl Acad Sci U S A** 113: 8320–8325
- Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridising populations. **Heredity** 57: 357–376
- Bayuelo-Jiménez JS, Ochoa-Cadavid I (2014) Phosphorus acquisition and internal utilization efficiency among maize landraces from the central Mexican highlands. **Field Crops Res** 156: 123–134
- Best NB, Hartwig T, Budka J, Fujioka S, Johal G, Schulz B, Dilkes BP (2016) *nana plant2* encodes a maize ortholog of the *Arabidopsis* brassinosteroid biosynthesis gene *DWARF1*, identifying developmental interactions between brassinosteroids and gibberellins. **Plant Physiol** 171: 2633–2647
- Bredeson JV, Lyons JB, Prochnik SE, Albert Wu G, Ha CM, Edsinger-Gonzales E, Grimwood J, Schmutz J, Rabbi IY, Egesi C, Nauluvula P, Lebot V, Ndunguru J, Mkamilo G, Bart RS, Setter TL, Gleadow RM, Kulakow P, Ferguson ME, Rounsley S, Rokhsar DS (2016) Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nat Biotechnol** 34: 562–570
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics** 30: 2114–2120

610 Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL:
611 software for association mapping of complex traits in diverse samples. **Bioinformatics** 23:
612 2633–2635

613 Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses.
614 **Bioinformatics** 19: 889–890

615 Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z, Wang B, Xu D, Yang B, Xie C, Fan L, Gao S,
616 Xu X, Zhang G, Li Y, Jiao Y, Doebley JF, Ross-Ibarra J, Lorant A, Buffalo V, Romay MC,
617 Buckler ES, Ware D, Lai J, Sun Q, Xu Y (2018) Construction of the third-generation *Zea mays*
618 haplotype map. **Gigascience** 7: 1–12

619 Chen LJ, Lee DS, Song ZP, Suh HS, Lu B-R (2004) Gene flow from cultivated rice (*Oryza*
620 *sativa*) to its weedy and wild relatives. **Ann Bot** 93: 67–73

621 Chuck GS, Brown PJ, Meeley R, Hake S (2014) Maize SBP-box transcription factors
622 *unbranched2* and *unbranched3* affect yield traits by regulating the rate of lateral primordia
623 initiation. **Proc Natl Acad Sci U S A** 111: 18775–18780

624 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM
625 (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms,
626 SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. **Fly** 6: 80–
627 92

628 Doebley J (1990) Molecular evidence for gene flow among *Zea* species. **Bioscience** 40: 443–448

629 Doebley JF (1983) The maize and teosinte male inflorescence: a numerical taxonomic study.
630 **Ann Mo Bot Gard** 70: 32

631 Eagles HA, Lothrop JE (1994) Highland maize from central Mexico - its origin, characteristics,
632 and use in breeding programs. **Crop Sci** 34: 11

633 Flint-Garcia SA, Thuillet A-C, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich
634 S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform
635 for quantitative trait locus dissection. **Plant J** 44: 1054–1064

636 Frey M, Chomet P, Glawischnig E, Stettner C, Grün S, Winklmair A, Eisenreich W, Bacher A,
637 Meeley RB, Briggs SP, Simcox K, Gierl A (1997) Analysis of a chemical plant defense
638 mechanism in grasses. **Science** 277: 696–699

639 Furuta T, Ashikari M, Jena KK, Doi K, Reuscher S (2017) Adapting genotyping-by-sequencing
640 for rice F2 populations. **G3** 7: 881–893

641 Goff SA, Cone KC, Chandler VL (1992) Functional analysis of the transcriptional activator
642 encoded by the maize *B* gene: evidence for a direct functional interaction between two classes of
643 regulatory proteins. **Genes Dev** 6: 864–875

644 Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W,

645 Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C,
646 Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B,
647 Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J,
648 Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Ž, Gušić I, Doronichev VB, Golovanova LV,
649 Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE,
650 Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S
651 (2010) A draft sequence of the Neandertal genome. **Science** 328:710–722

652 Grobman A, Bonavia D, Dillehay TD, Piperno DR, Iriarte J, Holst I (2012) Preceramic maize
653 from Paredones and Huaca Prieta, Peru. **Proc Natl Acad Sci U S A** 109: 1755–1759

654 Gutierrez A, Cantamutto M, Poverene M (2011) Persistence of sunflower crop traits and fitness
655 in *Helianthus petiolaris* populations. **Plant Biol** 13: 821–830

656 Han J-J, -J. Han J, Jackson D, Martienssen R (2012) Pod corn is caused by rearrangement at the
657 *Tunicate1* locus. **Plant Cell** 24: 2733–2744

658 Hartwig T, Chuck GS, Fujioka S, Klempien A, Weizbauer R, Potluri DPV, Choe S, Johal GS,
659 Schulz B (2011) Brassinosteroid control of sex determination in maize. **Proc Natl Acad Sci U S**
660 **A** 108: 19814–19819

661 Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new
662 mutation and standing variation as sources of adaptive variation. **Mol Ecol** 22: 4606–4618

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

666 Hufford MB, Lubinsky P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J (2013) The
667 genomic signature of crop-wild introgression in maize. **PLoS Genet** 9: e1003477

668 Hufford MB, Martínez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon MI (2012a) Inferences from
669 the historical distribution of wild and domesticated maize provide ecological and evolutionary
670 insight. **PLoS One** 7: e47659

671 Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ,
672 Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM,
673 Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler
674 ES, Yang S, Ross-Ibarra J (2012b) Comparative population genomics of maize domestication
675 and improvement. **Nat Genet** 44: 808–811

676 Janoušek V, Munclinger P, Wang L, Teeter KC, Tucker PK (2015) Functional organization of
677 the genome may shape the species boundary in the house mouse. **Mol Biol Evol** 32: 1208–1220

678 Je BI, Gruel J, Lee YK, Bommert P, Arevalo ED, Eveland AL, Wu Q, Goldshmidt A, Meeley R,
679 Bartlett M, Komatsu M, Sakai H, Jönsson H, Jackson D (2016) Signaling from maize organ
680 primordia via FASCIATED EAR3 regulates stem cell proliferation and yield traits. **Nat Genet**
681 48: 785–791

- 682 Jiang C, Edmeades GO, Armstead I, Lafitte HR, Hayward MD, Hoisington D (1999) Genetic
683 analysis of adaptation differences between highland and lowland tropical maize using molecular
684 markers. **Theor Appl Genet** 99: 1106–1119
- 685 Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation.
686 **Genetics** 173: 419–434
- 687 Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation
688 Sequencing Data. **BMC Bioinformatics** 15: 356
- 689 Lafitte HR, Edmeades GO (1997) Temperature effects on radiation use and biomass partitioning
690 in diverse tropical maize cultivars. **Field Crops Res** 49: 231–247
- 691 Lauter N, Gustus C, Westerbergh A, Doebley J (2004) The inheritance and evolution of leaf
692 pigmentation and pubescence in teosinte. **Genetics** 167: 1949–1959
- 693 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
694 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format
695 and SAMtools. **Bioinformatics** 25: 2078–2079
- 696 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013)
697 Preprint at <http://arxiv.org/abs/1303.3997>
- 698 Lovell JT, Jenkins J, Lowry DB, Mamidi S, Sreedasyam A, Weng X, Barry K, Bonnette J,
699 Campitelli B, Daum C, Gordon SP, Gould BA, Khasanova A, Lipzen A, MacQueen A, Palacio-
700 Mejía JD, Plott C, Shakirov EV, Shu S, Yoshinaga Y, Zane M, Kudrna D, Talag JD, Rokhsar D,
701 Grimwood J, Schmutz J, Juenger TE (2018) The genomic landscape of molecular responses to
702 natural drought stress in *Panicum hallii*. **Nat Commun** 9: 5213
- 703 Marques DA (2017) Adaptation despite gene flow? Low recombination helps. **Mol Ecol** 26:
704 4361–4363
- 705 Martin SH, Jiggins CD (2017) Interpreting the genomic landscape of introgression. **Curr Opin**
706 **Genet Dev** 47: 69–74
- 707 Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez G J, Buckler E, Doebley J (2002) A single
708 domestication for maize shown by multilocus microsatellite genotyping. **Proc Natl Acad Sci U**
709 **S A** 99: 6080–6084
- 710 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K,
711 Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a
712 MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Res** 20:
713 1297–1303
- 714 McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry
715 J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N,
716 Mitchell SE, Peterson B, Pressoir G, Romero S, Oropeza Rosas M, Salvo S, Yates H, Hanson M,
717 Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic

- 718 properties of the maize nested association mapping population. **Science** 325: 737–740
- 719 Mercer K, Martínez-Vásquez Á, Perales HR (2008) Asymmetrical local adaptation of maize
720 landraces along an altitudinal gradient. **Evol Appl** 1: 489–500
- 721 Merrill WL, Hard RJ, Mabry JB, Fritz GJ, Adams KR, Roney JR, MacWilliams AC (2009) The
722 diffusion of maize to the southwestern United States and its impact. **Proc Natl Acad Sci U S A**
723 106: 21019–21026
- 724 Mira MM, Hill RD, Stasolla C (2016) Phytoglobins improve hypoxic root growth by alleviating
725 apical meristem cell death. **Plant Physiol** 172: 2044–2056
- 726 Moose SP (2004) The maize *macrohairless1* locus specifically promotes leaf blade macrohair
727 initiation and responds to factors regulating leaf identity. **Genetics** 166: 1451–1461
- 728 Owens GL, Baute GJ, Rieseberg LH (2016) Revisiting a classic case of introgression:
729 hybridization and gene flow in Californian sunflowers. **Mol Ecol** 25: 2630–2643
- 730 Pardo-Díaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, McMillan WO,
731 Jiggins CD (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies.
732 **PLoS Genet** 8: e1002752
- 733 Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ,
734 Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) “Green
735 revolution” genes encode mutant gibberellin response modulators. **Nature** 400: 256–261
- 736 Perales H, Golicher D (2014) Mapping the diversity of maize races in Mexico. **PLoS One** 9:
737 e114657
- 738 Piperno DR (2006) Quaternary environmental history and agricultural impact on vegetation in
739 central america. **Ann Mo Bot Gard** 93: 274–296
- 740 Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence
741 for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. **Proc Natl**
742 **Acad Sci U S A** 106: 5019–5024
- 743 Poets AM, Fang Z, Clegg MT, Morrell PL (2015) Barley landraces are characterized by
744 geographically heterogeneous genomic origins. **Genome Biol** 16: 173
- 745 Radicella JP, Brown D, Tolar LA, Chandler VL (1992) Allelic diversity of the maize *B*
746 regulatory gene: different leader and promoter sequences of two *B* alleles determine distinct
747 tissue specificities of anthocyanin production. **Genes Dev** 6: 2152–2164
- 748 R Core Team (2014) R: A language and environment for statistical computing. R Foundation for
749 Statistical Computing
- 750 Rendón-Anaya M, Montero-Vargas JM, Saburido-Álvarez S, Vlasova A, Capella-Gutierrez S,
751 Ordaz-Ortiz JJ, Aguilar OM, Vianello-Brondani RP, Santalla M, Delaye L, Gabaldón T, Gepts P,

- Winkler R, Guigó R, Delgado-Salinas A, Herrera-Estrella A. (2017) Genomic history of the origin and domestication of common bean unveils its closest sister species. **Genome Biol** 18: 60
- Rezvoy C, Charif D, Guéguen L, Marais GAB (2007) MareyMap: an R-based tool with graphical interface for estimating recombination rates. **Bioinformatics** 23: 2188–2189
- Romero Navarro JA, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, Preciado E, Terron A, Delgado HV, Vidal V, Ortega A, Banda AE, Montiel NO, Ortiz-Monasterio I, Vicente FS, Espinoza AG, Atlin G, Wenzl P, Hearne S, Buckler ES (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. **Nat Genet** 49: 476–480
- Ruiz Corral JA, Durán Puga N, Sánchez González J de J, Ron Parra J, González Eguiarte DR, Holland JB, Medina García G (2008) Climatic adaptation and ecological descriptors of 42 Mexican maize races. **Crop Sci** 48: 1502
- Salazar-Vidal MN, Acosta-Segovia E, Sánchez-León N, Ahern KR, Brutnell TP, Sawers RJH (2016) Characterization and Transposon Mutagenesis of the Maize (*Zea mays*) *Pho1* Gene Family. **PLoS One** 11: e0161882
- Sanchez G. JJ, Goodman MM (1992) Relationships among the Mexican races of maize. **Econ Bot** 46: 72–85
- Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D (2014) The genomic landscape of Neanderthal ancestry in present-day humans. **Nature** 507: 354–357
- Sankararaman S, Mallick S, Patterson N, Reich D (2016) The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. **Curr Biol** 26: 1241–1247
- Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. **BMC Proc** 5: P54
- Selinger DA, Chandler VL (1999) Major recent and independent changes in levels and patterns of expression have occurred at the *b* gene, a regulatory locus in maize. **Proc Natl Acad Sci U S A** 96: 15007–15012
- Selinger DA, Chandler VL (2001) *B-Bolivia*, an allele of the maize *b1* gene with variable expression, contains a high copy retrotransposon - related sequence immediately upstream. **Plant Physiology** 1363–1379
- Sharma M, Cortes-Cruz M, Ahern KR, McMullen M, Brutnell TP, Chopra S (2011) Identification of the *prl* gene product completes the anthocyanin biosynthesis pathway of maize. **Genetics** 188: 69–79
- Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, Buckler ES, Hufford MB, Ross-Ibarra J (2015) Independent molecular basis of convergent highland adaptation in maize. **Genetics** 200:

788 1297-1312

789 Twyford AD, Friedman J (2015) Adaptive divergence in the monkey flower *Mimulus guttatus* is
790 maintained by a chromosomal inversion. **Evolution** 69: 1476–1486

791 Vieira FG, Fumagalli M, Albrechtsen A, Nielsen R (2013) Estimating inbreeding coefficients
792 from NGS data: impact on genotype calling and allele frequency estimation. **Genome Res** 23:
793 1852–1861

794 Vollbrecht E, Springer PS, Goh L, Buckler ES 4th, Martienssen R (2005) Architecture of floral
795 branch systems in maize and related grasses. **Nature** 436: 1119–1126

796 Walsh J, Freeling M (1999) The *liguleless2* gene of maize functions during the transition from
797 the vegetative to the reproductive shoot apex. **Plant J** 19: 489–495

798 Wang L, Beissinger TM, Lorient A, Ross-Ibarra C, Ross-Ibarra J, Hufford M (2017) The
799 interplay of demography and selection during maize domestication and expansion. **Genome**
800 **Biol.** 18, 215

801 Wellhausen EJ (1957) Races of Maize in Central America. National Academies

802 Wellhausen EJ, Roberts LM, Hernandez E, Mangelsdorf PC (1952) Races of maize in Mexico:
803 their origin, characteristics and distribution. Bussey Institution of Harvard Univ

804 Whitney KD, Randell RA, Rieseberg LH (2006) Adaptive introgression of herbivore resistance
805 traits in the weedy sunflower *Helianthus annuus*. **Am Nat** 167: 794–807

806 Whitney, KD, Broman KW, Kane NC, Hovick SM, Randell RA, Rieseberg LH (2015) Quantitative trait
807 locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a
808 wild sunflower. **Mol Ecol** 24: 2194–2211

809 Wilkes HG (1972) Maize and its wild relatives. **Science** 177: 1071–1077

810 Wu Q, Regan M, Furukawa H, Jackson D (2018) Role of heterotrimeric Gα proteins in maize
811 development and enhancement of agronomic traits. **PLoS Genet** 14: e1007374

812 Xia H-B, Wang W, Xia H, Zhao W, Lu B-R (2011) Conspecific crop-weed introgression
813 influences evolution of weedy rice (*Oryza sativa* f. *spontanea*) across a geographical range. doi:
814 10.1371/journal.pone.0016189

815 Xu G, Wang X, Huang C, Xu D, Li D, Tian J, Chen Q, Wang C, Liang Y, Wu Y, Yang X, Tian
816 F (2017) Complex genetic architecture underlies maize tassel domestication. **New Phytol** 214:
817 852–864

Figure 1

Gene flow from teosinte mexicana to Mexican highland maize.

(A) Mapping of the regions of introgression >250 kb in size to the ten chromosomes of maize (1-10). Colored points indicate the midpoint of regions identified using a 10% (blue) or 1% (red) outlier cut-off. Magenta bars show nine regions previously identified by *Hufford et al., 2013*. The cyan bar shows a Mb-scale region on chromosome 3 identified by *Wang et al., 2017*. The gray trace on the outermost track indicates local recombination frequency (\log_2 transformed) estimated from the PT x REV F_2 population. (B) Size distribution (frequency density) of introgression events using a 10% (blue) or 1% (red) outlier cut-off. Events binned by size < 0.5 Mb, 0.5 Mb to <1 Mb (1.0), 1 Mb or greater. Numbers above bars give the number of events in that class. (C) Cumulative gene count as a function of cumulative physical size of 10% outlier introgression regions, ordered by size. Blue points indicate the cumulative count. The black line shows the expected trend based on total genome size and gene number, under the assumption of a uniform spatial distribution. Alternating gray and white bars indicate size thresholds of the individual events of 0.5 Mb, 1 Mb, 1.5 Mb and > 1.5Mb.

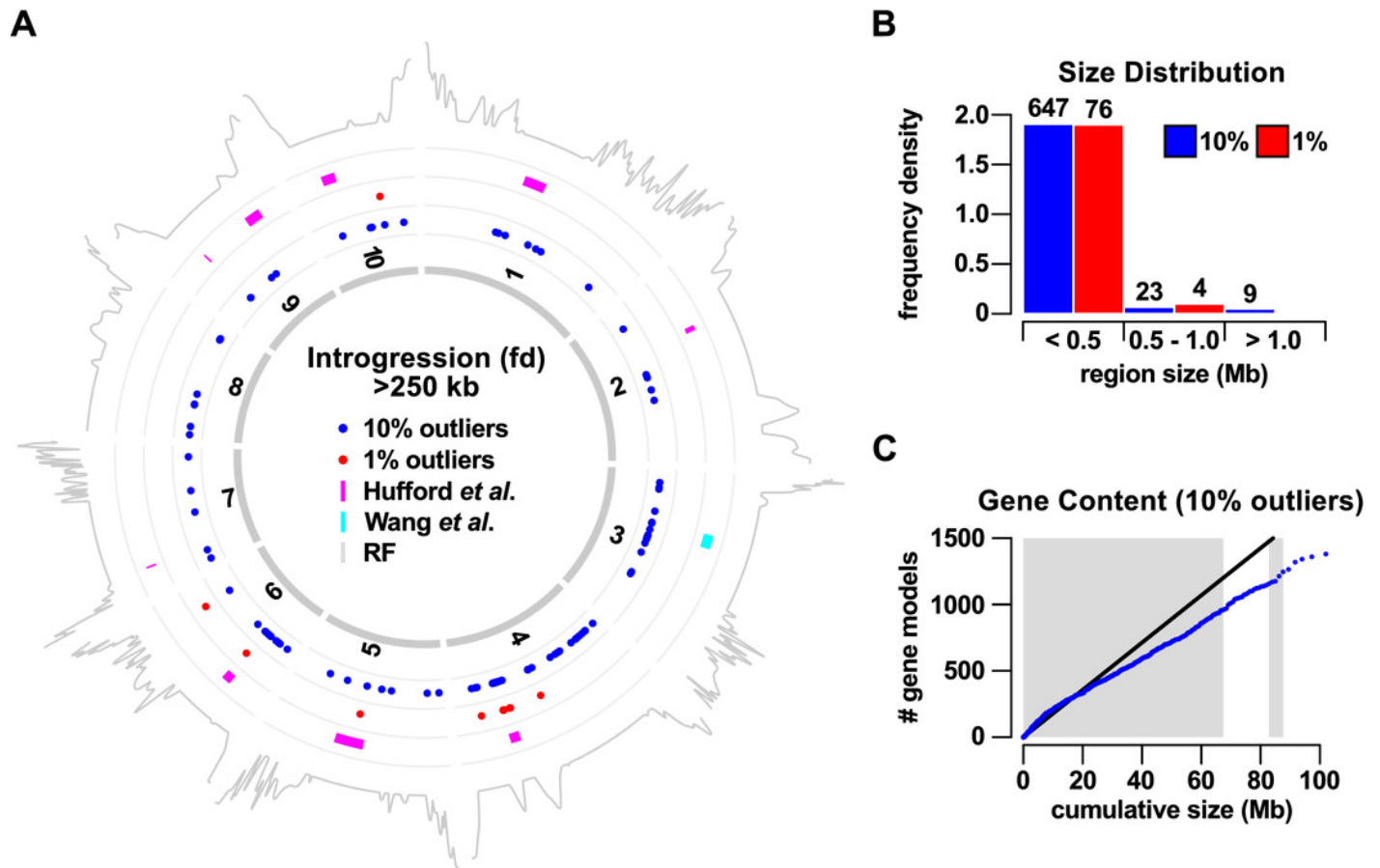
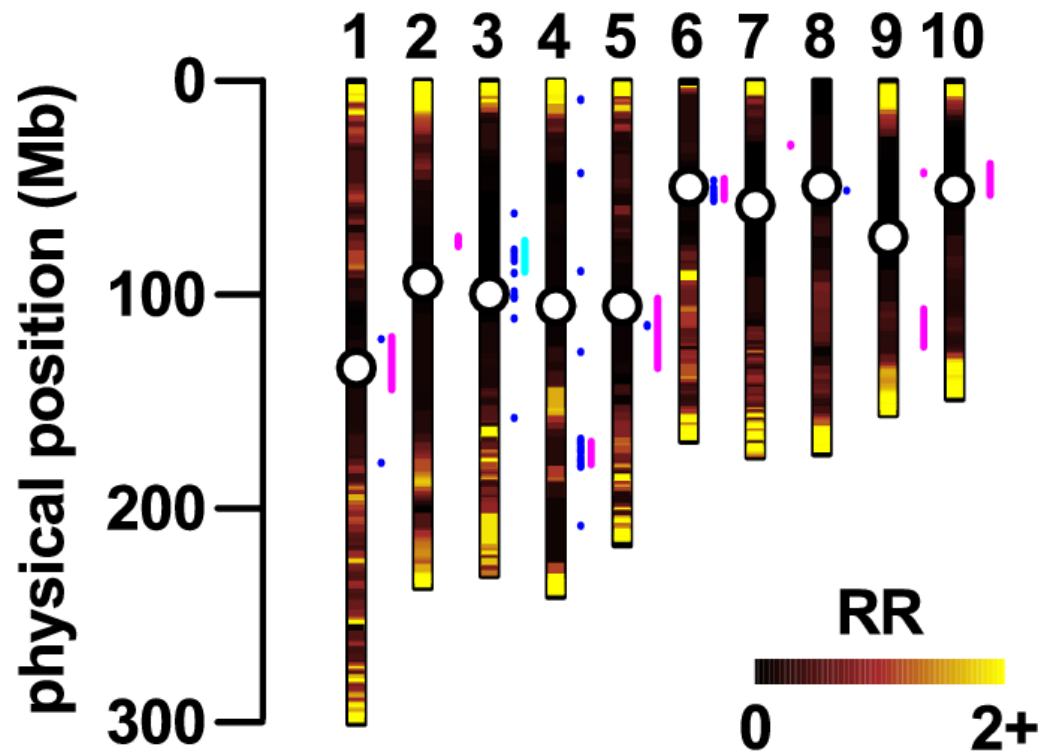


Figure 2

Large regions of introgression are associated with low genetic recombination.

(A) Local recombination rate (RR) across the physical space of the ten maize chromosomes (1 - 10), estimated from the PT x REV F₂ population. RR shown from 0 to 2 on a linear black-red-yellow scale. The RR distribution was truncated, with values > 2 set to 2 (predominantly high values in the telomeric regions). Thirty two introgression events > 500 kb in size that were identified by concatenation of the top 10% outlying fd windows are shown as blue line segments to the right of the individual chromosomes. Magenta and cyan line segments show regions previously identified by *Hufford et al., 2013* and *Wang et al., 2017*, respectively. Centromeres shown as open circles. (B) RR as a function of the physical size of introgression events. Events were grouped into size classes, given as the upper threshold on the plot. Boxes show 1st quartile, median and 3rd quartile. Whiskers extend to the most extreme points within 1.5x box length; outlying values beyond this range are not shown. Numbers above the boxes show the number of events in each size class.

A



B

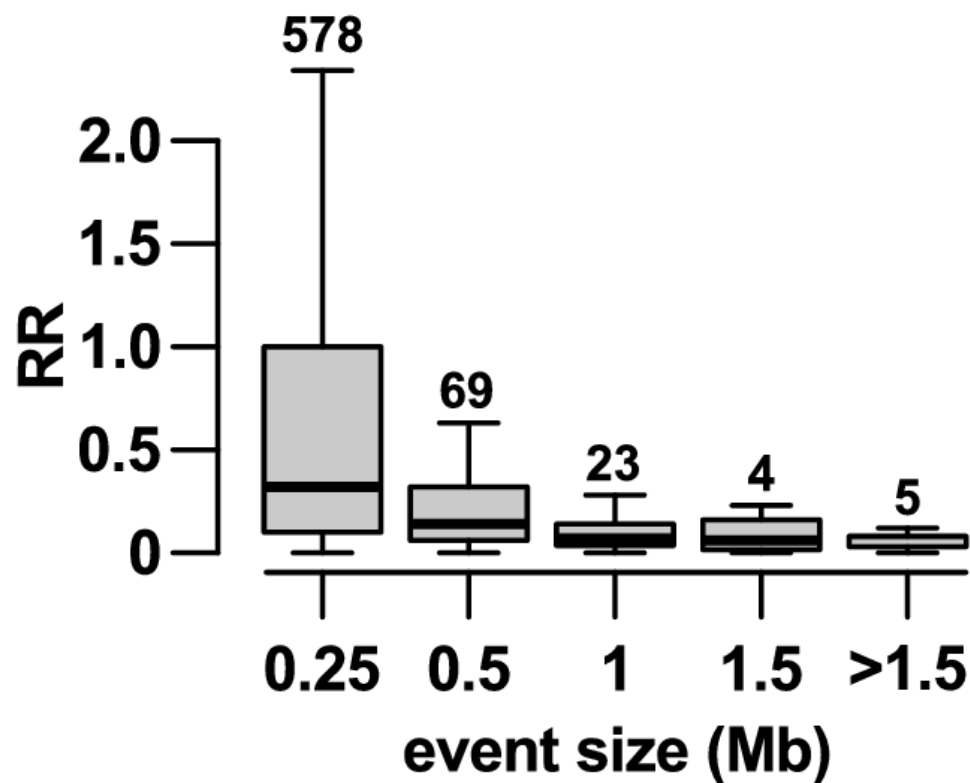


Figure 3

Introgression events on chromosome 4 co-localize with a previously reported inversion.

Marey plot (black dashed line) of genetic position in map units (mu; left-hand axis, outer) against physical position in Mb (x-axis) across chromosome 4. Local recombination rate (RR; solid black line) was calculated as the derivative of the physical-genetic plot (left-hand axis, inner). Introgression from mexicana to Mexican highland maize was estimated as local fd (green line; lowess smoothing of sliding window analysis). The result of a similar analysis substituting South American for Mexican highland maize is also shown (brown line).

Introgression events identified by selection and concatenation of the top 1% or top 10% outlying windows in the Mexican highland analysis are in red and blue, respectively. The position of the inv4m inversion polymorphism, as previously reported by *Hufford et al., 2013*, is shown as a gray rectangle. The position of the centromere is indicated as an open circle on the chromosome schematic.

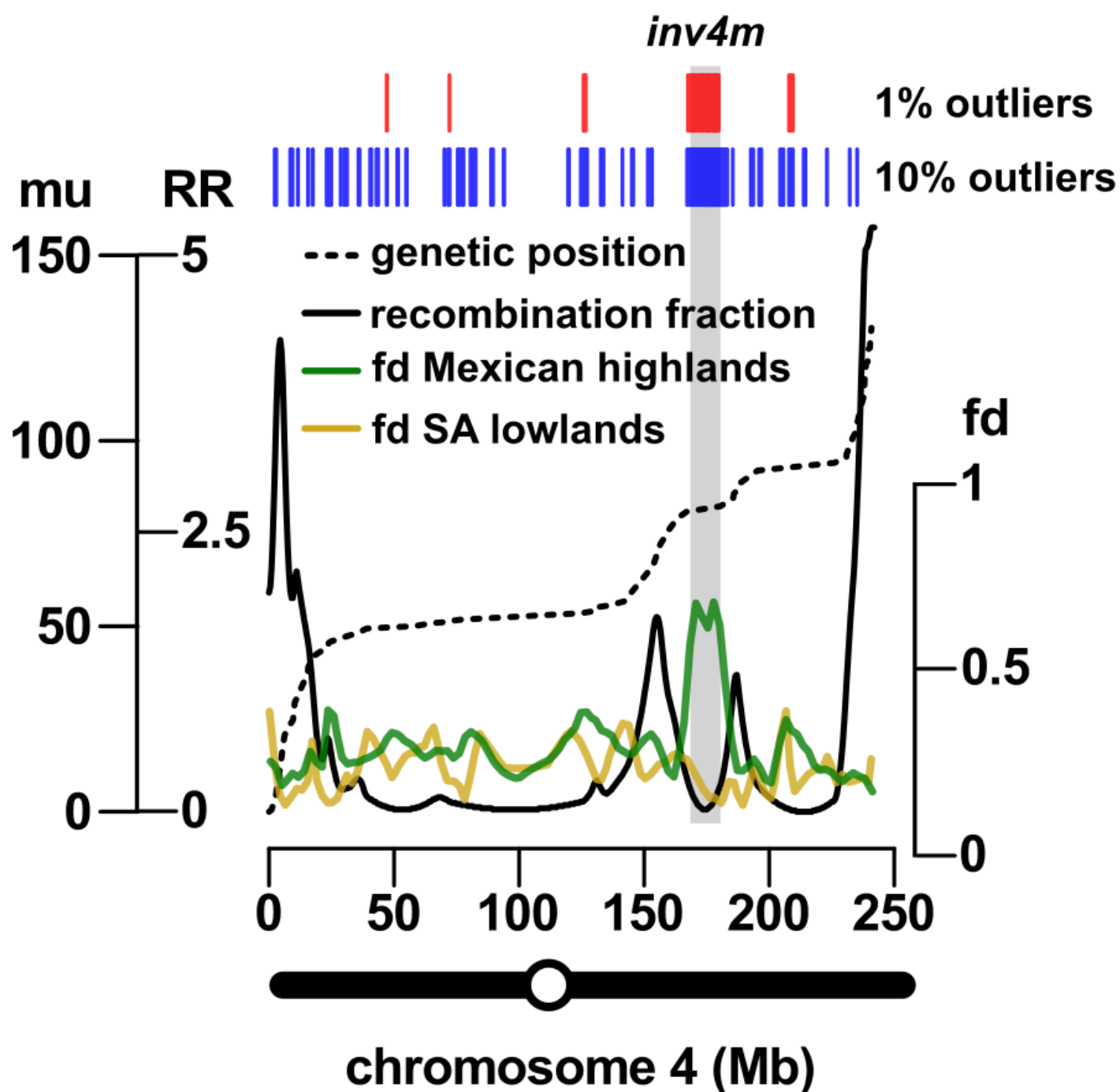
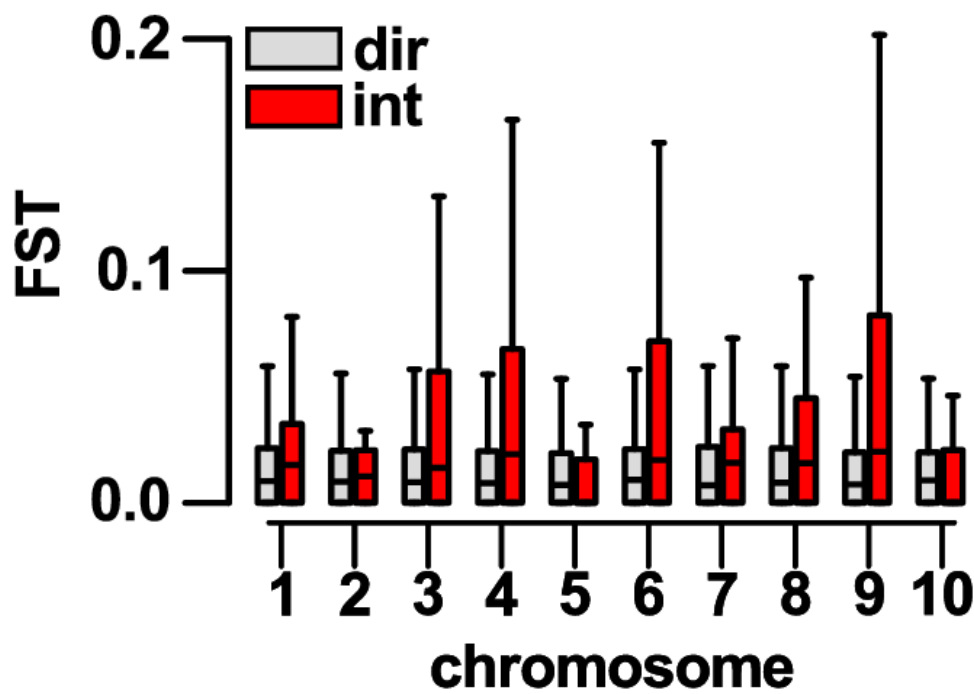


Figure 4

Genes located in introgression regions differentiate highland and lowland Mexican maize and are enriched for high effect SNPs

(A) F_{ST} between Mexican highland and lowland maize populations, as estimated by Takuno et al. (median F_{ST} of all SNPs per gene model; 21, 029 gene set), for genes located in regions of direct descent (dir) or introgression (int) events, on the ten chromosomes of maize. Boxes show 1st quartile, median and 3rd quartile. Whiskers extend to the most extreme points within 1.5x box length; outlying values beyond this range are shown as points, including a number of extreme outliers in both direct and introgression regions. (B) The number of high or moderate effect SNPs fixed in the three Mexican highland maize individuals with respect to their distribution in the other samples. Colored boxes indicate SNPs fixed for the alternate allele (blue), the reference allele (red) or segregating (magenta) in a given group. Blanks indicate that SNPs could be not called. Only sets containing greater than 25 SNPs are shown.

A



B

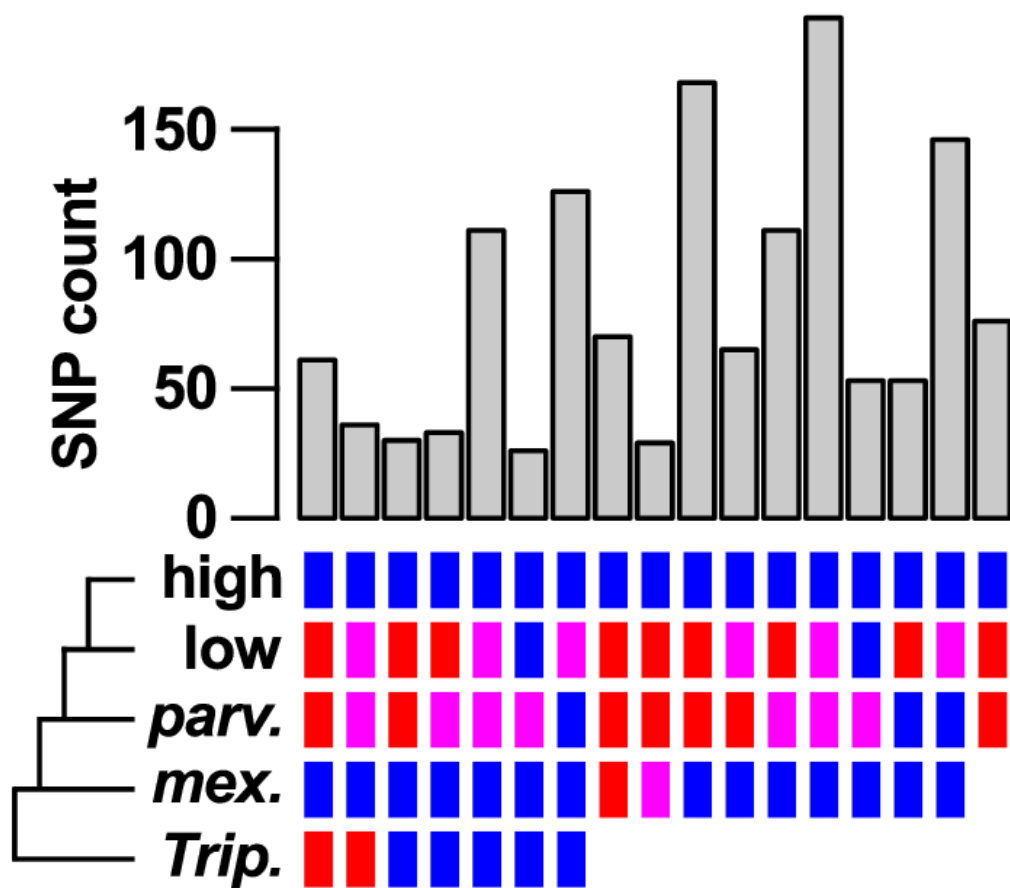


Figure 5

QTL peaks associated with morphological traits do not co-localize with large-scale introgression events.

(A) Palomero Toluqueño (PT) exhibits the extensive stem pigmentation and pubescence typical of Mexican highland maize, in contrast to (B) the lowland landrace Reventador (RV) which presents a non-pigmented, glabrous stem. Whole-genome QTL scans for the intensity of (C) stem pigmentation (pigment intensity) and (D) the degree of branching of the male inflorescence (tassel branch number). The horizontal axis shows genetic position across the ten maize chromosomes in the PT x RV map, tick-marks indicating marker position, alternating white/gray banding in the plot area indicating chromosomes. Vertical blue lines across the plot area indicate the estimated position of introgression events on the genetic map, line width proportional to event size. The vertical axis shows LOD support for the presence of a QTL. The red trace shows the output from a single-QTL interval scan. The horizontal broken black-lines indicate the 5% LOD threshold for each trait, as determined by permutation. Major peaks for pigment intensity and tassel branch number are located on chromosomes 2 and 7, respectively.

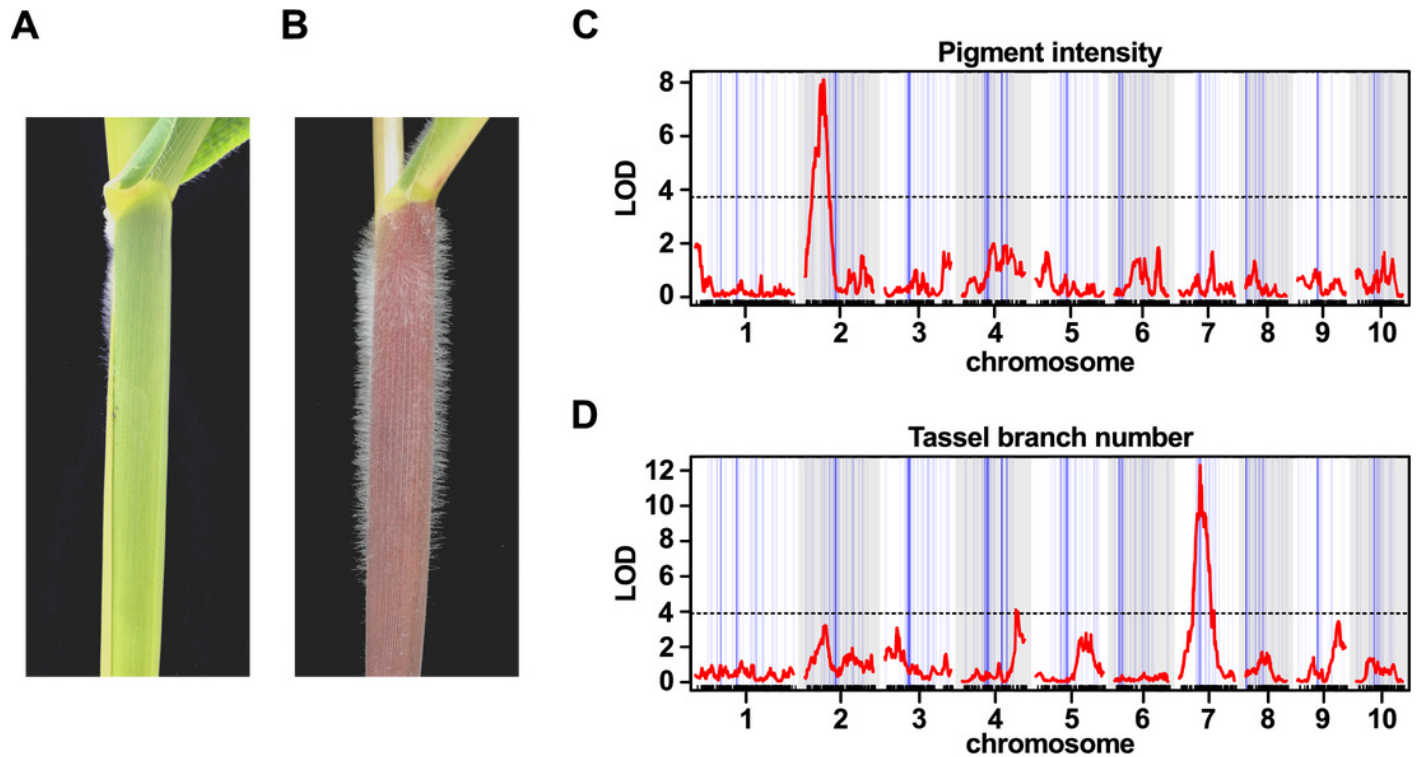


Table 1(on next page)

Introgression from mexicana to Mexican highland maize and recombination rate by chromosome

1 **Table 1. Introgression from *mexicana* to Mexican highland maize and recombination rate**
 2 **(RR) by chromosome**

Target	Size (Mb)	D	Z	p-value	Number of events ¹	Total event size (Mb)	Introgression ² (%)	Map length (cM)	RR (cM/Mb)
Chr1	301	0.066	6.0	< 0.001	84	9.3	3.1	208	0.69
Chr2	238	0.063	5.7	< 0.001	67	5.7	2.4	144	0.61
Chr3	232	0.075	5.0	< 0.001	85	2	8.8	138	0.60
Chr4	242	0.133	5.4	< 0.001	90	24.8	10.2	130	0.54
Chr5	218	0.072	5.5	< 0.001	72	8.9	4.1	142	0.65
Chr6	169	0.104	5.2	< 0.001	74	11.5	6.8	113	0.67
Chr7	177	0.055	4.5	< 0.001	51	5.1	2.9	116	0.66
Chr8	175	0.041	2.6	0.009	61	6.9	4.0	86	0.49
Chr9	157	0.028	1.7	0.09	52	4.7	3.0	100	0.64
Chr10	150	0.051	3.1	0.002	43	4.6	3.0	97	0.65
Total	2060				679	102	5.0	1275	0.62

3 1. Concatenated 10% outliers. 2. Based on physical size.

4

Table 2(on next page)

Selected classic genes located in introgression events

1 **Table 2. Selected classic genes located in introgression events**

Gene	ID	Function	Chr	Pos (Mb)	Top 10%	Top 1%
<i>Al</i>	GRMZM2G026930	Pigments	3	216	X	
<i>Acco20</i>	GRMZM2G126732	Hormones (Et)	4	178	X	X
<i>Bx8</i>	GRMZM2G085054	Defense	4	3	X	
<i>Cle24</i>	GRMZM2G123818	Morphology	4	170	X	
<i>Ct2</i>	GRMZM2G064732	Morphology	1	16	X	
<i>D8</i>	GRMZM2G144744	Hormones (GA)	1	266	X	
<i>Fea3</i>	GRMZM2G166524	Morphology	3	291	X	
<i>Fl3</i>	GRMZM2G006585	Grain development	8	52	X	
<i>Gi2</i>	GRMZM5G844173	Flowering	3	9	X	
<i>Ko1</i>	GRMZM2G059308	Hormones (GA)	9	80	X	
<i>Na1</i>	GRMZM2G449033	Hormones (Br)	3	179	X	
<i>Nat11</i>	GRMZM2G455658	Hormones (Br)	4	169	X	X
<i>O1</i>	GRMZM2G449909	Grain development	4	177	X	
<i>Orp2</i>	GRMZM2G005024	Grain development	10	84	X	
<i>Sbe1</i>	GRMZM2G088753	Grain development	5	63	X	
<i>Ss5</i>	GRMZM2G130043	Grain development	4	173	X	
<i>Su4</i>	GRMZM2G090905	Grain development	6	145	X	
<i>Pho1;2a</i>	GRMZM2G466545	Nutrition	4	172	X	X
<i>Tu1</i>	GRMZM2G370777	Morphology	4	179	X	X

2

Table 3(on next page)

QTL associated with plant morphology and flowering time

1 **Table 3. QTL associated with plant morphology and flowering time**

QTL	Chr	Peak (Mb)	95% Interval (Mb)	Var (%)	LOD	Candidate gene
qINT-2	2	19	12 - 21	23	9.2	<i>Bl</i>
qPAT-2	2	13	12 - 20	18	6.7	<i>Bl</i>
qTBN-2	2	18	3 - 196	8	4.0	
qTBN-7	7	98	37 - 118	31	13.2	<i>Ral</i>
qTL-4	4	30	10 - 155	10	3.8	
qTL-5	5	2	1 - 3	12	4.5	
qDTA-8	8	117	1 - 125	14	4.5	<i>Zcn8</i>

2